

RESEARCH

Open Access



# Mucin-microbiome signatures shape the tumor microenvironment in gastric cancer

Baptiste Oosterlinck<sup>1,2</sup>, Hannah Ceuleers<sup>1,2</sup>, Wout Arras<sup>1,2</sup>, Joris G. De Man<sup>1,2</sup>, Karen Geboes<sup>3</sup>, Heiko De Schepper<sup>1,2,4</sup>, Marc Peeters<sup>5</sup>, Sarah Lebeer<sup>6</sup>, Jurgita Skieceviciene<sup>7</sup>, Georgina L. Hold<sup>8</sup>, Juozas Kupcinskas<sup>7</sup>, Alexander Link<sup>9</sup>, Benedicte Y. De Winter<sup>1,2,4†</sup> and Annemieke Smet<sup>1,2\*†</sup>

## Abstract

**Background and aims** We aimed to identify mucin-microbiome signatures shaping the tumor microenvironment in gastric adenocarcinomas and clinical outcomes.

**Methods** We performed high-throughput profiling of the mucin phenotypes present in 108 gastric adenocarcinomas and 20 functional dyspepsia cases using validated mucin-based RT-qPCRs with subsequent immunohistochemistry validation and correlated the data with clinical outcome parameters. The gastric microbiota was assessed by 16S rRNA gene sequencing, taxonomy, and community composition determined, microbial networks analyzed, and the metagenome inferred in association with mucin phenotypes and expression.

**Results** Gastric adenocarcinomas with an intestinal mucin environment or high-level *MUC13* expression are associated with poor survival. On the contrary, gastric *MUC5AC* or *MUC6* abundance was associated with a more favorable outcome. The oral taxa *Neisseria*, *Prevotella*, and *Veillonella* had centralities in tumors with intestinal and mixed phenotypes and were associated with *MUC13* overexpression, highlighting their role as potential drivers in *MUC13* signaling in GC. Furthermore, dense bacterial networks were observed in intestinal and mixed mucin phenotype tumors whereas the lowest community complexity was shown in null mucin phenotype tumors due to higher *Helicobacter* abundance resulting in a more decreased diversity. Enrichment of oral or intestinal microbes was mucin phenotype dependent. More specifically, intestinal mucin phenotype tumors favored the establishment of pro-inflammatory oral taxa forming strong co-occurrence networks.

**Conclusions** Our results emphasize key roles for mucins in gastric cancer prognosis and shaping microbial networks in the tumor microenvironment. Specifically, the enriched oral taxa associated with aberrant *MUC13* expression can be potential biomarkers in predicting disease outcomes.

**Keywords** Stomach cancer, Mucin, Microbiota, Survival

†Benedicte Y. De Winter and Annemieke Smet contributed equally to this work.

\*Correspondence:

Annemieke Smet

annemieke.smet@uantwerpen.be

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Gastric cancer (GC) is the fifth most common cancer type and the fourth leading cause of cancer-related deaths worldwide [1]. The prognosis of GC remains poor due to the lack of symptoms in early disease, leading to a delayed diagnosis [2]. It is now widely accepted that GC is a multifactorial disease involving host genetic susceptibility and environmental factors, but the most frequent cause is infection by *Helicobacter pylori* (*H. pylori*) [3]. This class 1 human carcinogen plays a major role in the initial steps of the carcinogenic process through a hit-and-run mechanism priming the gastric mucosa for further oncogenic changes which are triggered by other microbial species [3, 4]. Studies assessing human gastric microbiota profiles have shown that dysbiosis in the stomach is a dynamic process correlating with cancer progression and that gastric adenocarcinomas are characterized by a decrease in *Helicobacter* abundance and enrichment of bacterial genera representing intestinal commensals (*Citrobacter*, *Clostridium*, *Lactobacillus*, *Achromobacter*, and *Rhodococcus*) or the oral microbiome (*Peptostreptococcus stomatis*, *Streptococcus anginosus*, *Parvimonas micra*, *Slackia exigua*, *Lactococcus*, and *Fusobacterium*) [5–8]. Furthermore, tumor microhabitats are not always as uniform as previously thought [8]. Besides changes in pH, mucin (MUC) expression, and distribution vary considerably among gastric tumor tissues with both gastric and intestinal mucins being widely expressed. Depending on the presence/absence of mucins, adenocarcinomas have been classified as having a gastric (i.e., tumors expressing only gastric MUC1, MUC5AC, and/or MUC6 mucins), intestinal (i.e., tumors expressing only intestinal MUC2, MUC3, MUC4, and/or MUC13 mucins), mixed (i.e., tumors expressing both gastric and intestinal mucins), or unclassified/null (i.e., tumors expressing nor gastric, nor intestinal mucins) mucin phenotype [9–15]. Early gastric cancers mainly exhibit a gastric mucin phenotype, whereas advanced cancers more frequently have an intestinal mucin phenotype. However, the clinical importance of mucin expression in gastric tumors is still controversial in the context of clinicopathological factors, such as disease outcome, as it remains unclear which mucin phenotype associates with a better or worse prognosis [9, 10, 12, 15–19]. Mucins are the gatekeepers of the mucus barrier covering the gastric epithelium and are expressed either as secretory or transmembrane glycoproteins [20–24]. Besides having a barrier function, they also serve as specialized niches for bacteria by acting as binding sites or metabolic substrates and are important determinants of site-specific bacterial colonization [20–24]. It has been suggested that aberrant mucin alterations due to neoplastic changes can result in the establishment of a new microbiota promoting tumor progression [7].

The differences in the abundance of arising new taxa (i.e., intestinal or oral) as previously described in GC [5, 25] may thus be assigned to the mucin phenotype (gastric, intestinal, mixed, or null mucin phenotype) of the tumor, but further investigation is required. To investigate the above hypothesis, we performed high-throughput profiling of the mucin phenotypes and bacterial communities present in 108 gastric adenocarcinoma and 20 functional dyspepsia (FD) (i.e., for comparison) cases using validated mucin-based RT-qPCRs with subsequent immunohistochemistry (IHC) validation and 16S rRNA gene sequencing integrated with clinical data to identify mucin-microbiota signatures associated with GC and clinical outcome.

## Methods

### Patients

A total of 108 GC patients undergoing gastrectomy for gastric cancer were enrolled in this study. Seventeen patients were recruited via the biobank of the Antwerp University Hospital (UZA; Belgium), 48 via the Department of Digestive Oncology of Ghent University Hospital (Belgium), and 43 via the Institute for Digestive Research of the Lithuanian University of Health Sciences (Lithuania). Twenty patients with FD undergoing gastroscopy for clinical reasons and showing no macroscopic abnormalities were included as a comparison cohort via the Department of Gastroenterology and Hepatology (UZA; Belgium). Tumor ( $n=108$ ) and adjacent non-tumor ( $n=108$ ) tissues from GC patients and biopsy tissues from FD patients were stored in RNAlater or snap frozen at  $-80^{\circ}\text{C}$  or embedded in paraffin for RNA/DNA extraction with subsequent downstream approaches or IHC analyses, respectively. The recorded data for the GC cohorts included: gender, age, tumor localization, Lauren's classification (intestinal, diffuse, or mixed histological subgroups) [26], TNM-G staging, and survival rate (Table S1). This study was approved by the Ethical Committee of the Antwerp University Hospital (EC 19/15/205 (GC) and B300201733550 (FD)) and the Kaunas Regional Ethics Committee (Protocol No—BE-2-10), and written informed consent was obtained from the patients prior to sample collection. Samples were registered and stored until analysis in the Biobank Antwerpen, Antwerp, Belgium (BE 71,030,031,000; BBMR-ERIC, Belgian no. access: 1, Last: April 10, 2021 [BIORESOURCE]).

### Mucin mRNA expression by RT-qPCR

The total RNA was extracted from tumor and adjacent non-tumor tissues and FD biopsies using the NucleoSpin RNA plus kit (Macherey–Nagel) following the manufacturer's instructions. RNA concentration was evaluated using the nanodrop ND-1000 UV–Vis

spectrophotometer (Thermo Fisher Scientific). Two hundred fifty nanograms of RNA was converted to cDNA by reverse transcription using the SensiFast cDNA synthesis kit (Bioline). Relative mucin gene expression was determined by SYBR Green RT-qPCR using validated QuantiTect primers (Qiagen, Table S2) and GoTaq qPCR master mix (Promega) using a QuantStudio 3 real-time PCR instrument (Thermo Fisher Scientific) [27]. Relative mRNA expression of mucin genes was normalized to the expression of *ACTB* and *GAPDH* housekeeping genes using qbase+ software (Biogazelle), which implements an adapted  $\Delta C_t$  method and calculates calibrated normalized relative quantities (CNRQ) for downstream analysis [28]. To define the mucin phenotypes in the tumor tissues and to clarify conflicting results previously obtained with IHC regarding the clinical importance of mucin expression, we designed a new approach based on RT-qPCR [9, 10, 12, 15–19]. This quantitative technique displays a broader dynamic range with higher sensitivity and reproducibility and is significantly more specific than IHC resulting in more reliable data [29]. More specifically, the 90% confidence interval (CI) of the relative mRNA expression of each mucin was determined for the FD cohort. The lower (LL) and upper (UL) limits of CIs were used to stratify the relative expression levels found in the tumor tissue into three distinct classes: (1) low < LL, (2) LL < mid < UL, and (3) UL < high (Table S3). Depending on the class of relative mRNA expression of mucins, adenocarcinomas were classified as having a gastric (mid or high mRNA levels of MUC5AC, MUC6, and/or MUC1 and low mRNA levels of MUC2, MUC4, and MUC13), intestinal (low mRNA levels of gastric mucins and high mRNA levels of at least one intestinal mucin), mixed (mid or high mRNA levels of at least one gastric mucin and high mRNA levels of at least one intestinal mucin), or null (low mRNA levels for all gastric and intestinal mucins) mucin phenotype.

### Immunohistochemistry

To evaluate mucin mRNA expression at the protein level, tissue segments were fixed for 24 h in 4% formaldehyde and subsequently embedded in paraffin. Five-micrometer cross-sections were deparaffinized, rehydrated, and used for immunohistochemical staining using target-specific primary antibodies and visualization with a secondary streptavidin–horseradish peroxidase antibody and 3-amino-9-ethylcarbazole (AEC) substrate to detect the expression and localization of MUC1 (AF6298, R&D systems, 1:500), MUC2 (NBP1-31231, Novus Biologicals, 1:3000), MUC4 (NBP1-52193, Novus Biologicals, 1:3000), MUC5AC (ab3649, Abcam, 1:5000), MUC6 (ab216017, Abcam, 1:50), and MUC13 (MABC209, Merck Millipore, 1:1000). The stained sections were

analyzed by light microscopy (Olympus BX43) [30]. Distinct staining in more than 10% of the gastric cells was recorded as positive immunoreactivity for the relevant mucin. Tumors containing epithelial cells expressing only gastric or intestinal-type mucins were classified as having a gastric or intestinal mucin phenotype, respectively. Those containing cells expressing both gastric and intestinal type mucins were classified as having a mixed phenotype, whereas tumors with cells expressing neither gastric nor intestinal type mucins were classified as having a null mucin phenotype. The degree of immunostaining was evaluated by two independent observers.

### Survival analysis

Kaplan–Meier curves and Cox proportional-hazards models were used for survival analysis. The Kaplan–Meier curves implementing the log-rank test were performed according to the tumor mucin phenotype and stratified mucin mRNA expression levels. The mucin phenotype and the six stratified mucin expression levels were also included in a Cox proportional-hazards model taking gender, age, tumor location, Lauren’s classification, and tumor stage into account.

### 16S rRNA gene sequencing

DNA was extracted from the gastric surgical specimens and biopsies using the DNeasy Blood and Tissue kit (Qiagen) following the manufacturer’s instructions. A negative and positive (ZymoBIOMICS Microbial Community DNA standard – cat no. D6305) control was also included and processed together with the samples. Subsequently, library preparation was performed according to the standard Illumina protocol for the V3 chemistry for paired-end sequencing (2 × 300 bp) using the universal forward 27Fmod, 5’AGRGTTGATYMTGGCTC AG and reverse 338R, 5’TGCTGCCTCCCGTAGGAGT primers targeting the V1–V2 hypervariable region of the 16S rRNA gene [31]. All samples were pooled equimolarly and sequenced using a MiSeq Illumina platform.

### Sequence filtering and annotation

Sequence quality assessment, paired-end read merging, filtering, sequence denoising, and chimeric read filtering were performed using the dada2 R-package [32]. The final sequences were aligned to the SILVA reference database (version 138) and annotated using the DECIPHER R-package [33, 34]. Any sequences that were annotated as non-bacterial were discarded. Finally, sample rarefaction curves were determined and all samples with curves not attaining a stable plateau were considered to be undersampled and thus discarded. As a final filtering step, the counts from the negative control were subtracted from all samples after which the relative

abundance was calculated. All genera with a relative abundance below 0.5% were pruned, and for the remaining genera, the original counts were used to recalculate the relative abundances [35]. Sample metadata, abundance table, taxonomy, phylogenetic tree, and reference sequence data were grouped into a Phyloseq object for downstream processing [36]. The alpha (chao1, observed richness, inverse Simpson, and Shannon) and beta diversity measures (Bray–Curtis and weighted unifrac) were calculated using the relative abundances. Differences in community composition were tested using analysis of similarities (ANOSIM) [37].

#### Differential abundance analysis

Differential abundance analyses were done using the ALDEx2 R-package [38]. This package was selected based on its performance and native support for compositional datasets based on a recent review from Nearing et al. (2022) assessing different methods for differential abundance analysis [39]. All reference sequences of the amplicon sequence variants (ASV) found were compared against sequences from the Human oral Microbiome (HOM) database (HOMD 16S rRNA RefSeq version 15.22) and the Human Intestinal 16 s rRNA gene reference database using BLAST local nucleotide alignment [40, 41]. Only matches with at least 99% identity and over 250 base pairs were retained. ASVs with a species assigned and which were not annotated as having an oral or nasal habitat in the HOM database were retained as intestinal bacterial taxa. Relative abundances of both oral and intestinal species were compared between tumor, non-tumor, and FD tissues and among tumor samples with different mucin phenotypes to determine whether oral or intestinal taxa were enriched or depleted.

#### Microbial network analysis

For calculating the co-occurrence and co-excluding microbial interactions, the co-occur R-package was used after the transformation of the count table to a presence/absence matrix [42]. The output generated is an object containing non-random associations with the probability of a lower and a higher co-occurrence for a genus pair. These probabilities were interpreted as *p* values for a negative or positive correlation, respectively [42]. Obtained networks were plotted using the ggraph R-package.

#### Functional metagenome inference

The mucosa-associated functional metagenome was predicted using the PICRUSt2 algorithm with default settings [43]. PICRUSt2 is able to predict the presence of functional genes based on a set of reference genomes and a marker gene dataset, allowing for the prediction of pathway abundances among the GC cohorts. By default,

the abundance of the enzyme commission (EC) number (= a numerical classification scheme for enzymes based on the chemical reactions they catalyze) is inferred based on the relative abundance of the genera. Subsequently, the EC numbers are transformed into MetaCyc pathways and allow for their abundance calculation. These MetaCyc pathway abundances were then used for a differential abundance analysis.

#### Data analysis

Two sample and multiple group comparisons were done using the Wilcoxon rank sum test and the Kruskal–Wallis test, respectively. Spearman's correlations between the clinical patient data, mucin mRNA expression levels, and bacterial abundance in gastric adenocarcinoma samples were calculated. A correlogram plotting the Spearman's rank correlation coefficient (*r*) between all parameter pairs was created. Differences in proportions were analyzed by Pearson's  $\chi^2$ . After differential abundance analysis with ALDEx2, a Wilcoxon rank-sum test was used to assess significant differences between conditions. *P* values below 0.05 were considered significant. All analyses were done using R version 4.2.2 in RStudio.

## Results

### Phenotypical classification of gastric adenocarcinomas based on mucin expression

In the present study, we first analyzed the tumor and adjacent non-tumor tissues of the GC patient cohorts and the biopsy tissues of the FD patients to measure the relative mRNA expression of gastric (*MUC1*, *MUC5AC*, *MUC6*) and intestinal (*MUC2*, *MUC4*, *MUC13*) mucins. Overall, mRNA expression of *MUC1*, *MUC5AC*, and *MUC6* was significantly higher in the paired adjacent non-tumor tissues compared to the tumor and FD tissues (Fig. 1A). Regarding the intestinal mucins, a significant increase in *MUC13* mRNA expression was seen in the paired tumor tissues compared to the adjacent non-tumor and FD tissues whereas no significant alterations in expression were seen for *MUC2* and *MUC4* mRNA among the different sample types (Fig. 1A). The variable expression patterns of *MUC2* (i.e., marker for intestinal metaplasia) and *MUC4* (i.e., marker for spasmolytic polypeptide expressing metaplasia (SPEM)) seen in the adjacent non-tumor tissues suggest the presence of metaplastic changes in these peritumoral sites [44, 45]. The gastric adenocarcinomas were then subdivided in gastric, intestinal, mixed, or null mucin phenotype groups based on their relative mucin mRNA expression levels. Of the 108 tumor samples, 13 (12%) were classified in gastric, 19 (17,6%) in intestinal, 17 (15,7%) in null, and 47 (43,5%) in mixed mucin phenotype groups (Fig. 1B). From 12 samples (11.1%), the mucin phenotype could not

be determined due to insufficient RNA quality (Fig. 1B). Subsequently, to validate the classification of the tumors in the different phenotype groups, a principal component analysis based on the mucin mRNA expression data was undertaken. Strikingly, *MUC2* and *MUC13* mRNA expression were the major determinants for tumors with an intestinal mucin phenotype whereas the expression of *MUC1*, *MUC5AC*, and *MUC6* mRNA were the best factors to identify tumors with a gastric mucin phenotype (Fig. 1C). Finally, mucin expression was also evaluated at the protein level by IHC further confirming our mRNA expression data (Fig. 1D).

### The intestinal mucin phenotype and aberrant *MUC13* mRNA expression correlate with worse survival

We then evaluated collinearity between the mucin mRNA expression data, age, gender, tumor stage, Lauren's classification, and survival using Spearman's correlation tests (Fig. S1). A strong positive correlation was seen between *MUC13* mRNA expression and the Lauren's intestinal phenotype whereas negative associations were noted between *MUC1* mRNA expression and survival and *MUC4* mRNA expression and age (Fig. S1). Furthermore, significant relationships among the gastric mucin mRNA expression profiles and the expression levels of the intestinal *MUC2* and *MUC4* were also identified (Fig. S1).

Subsequently, associations between mucin mRNA expression, mucin phenotypes, and the 5-year survival rate were also investigated using Kaplan–Meier and Cox proportional hazards models (Fig. 2). For the latter model, clinical patient data (age, gender, tumor stage, and location; Table S1) was also taken into account. Both methods showed an association between the intestinal mucin phenotype and a worse survival rate compared to the gastric, mixed, and null mucin phenotypes (log-rank test,  $P=0.01$ ; Wald test,  $P=0.016$ , Fig. 2A). Significant associations between individual mucin expression levels and survival rate were also identified. Specifically, Kaplan–Meier curves indicated that low mRNA

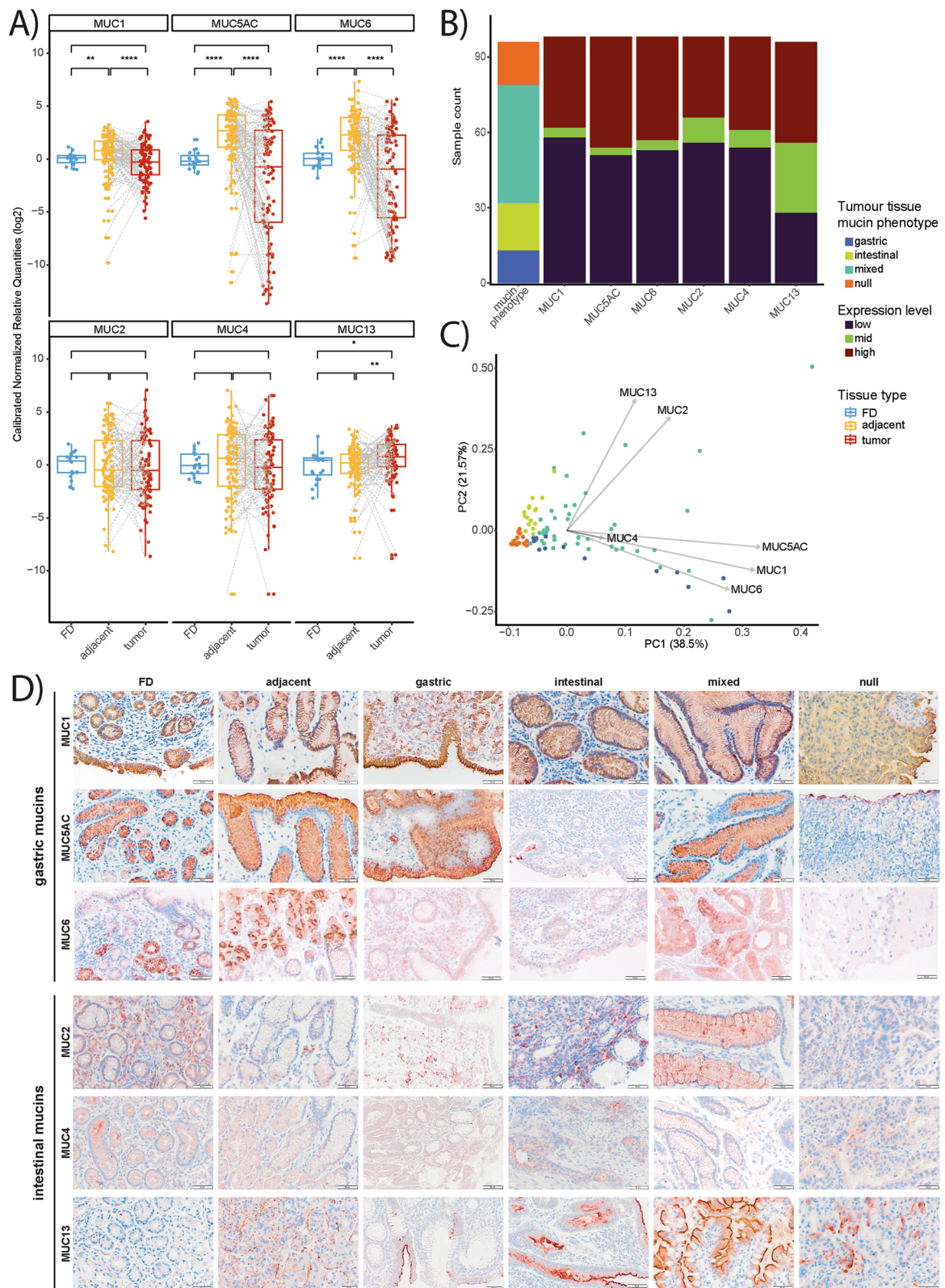
expression of *MUC5AC* and *MUC6* in gastric tumor tissue correlated with worse survival ( $P\leq 0.027$ ; Fig. 2B). A similar trend was also seen for the expression of *MUC1* mRNA, although not significant ( $P=0.069$ ; Fig. 2B). On the contrary, mid-level expression of *MUC2* mRNA ( $P=0.038$ ) associated with worse survival, whereas a trend for low- and high-level expression of *MUC13* mRNA towards an unfavorable outcome was noted ( $P=0.068$ ; Fig. 2B). However, only a significant association between high-level *MUC13* expression and worse survival was defined by the Cox-proportional hazards model ( $P=0.016$ ; Fig. 2B). In addition, the gastric cardia ( $P=0.02$ ) and stage 3 ( $P=0.001$ ) also correlated with a worse outcome (Fig. 2B).

### Microbiome dysbiosis alters between the different mucin phenotype groups

To determine dysbiosis associated with the different GC mucin phenotypes, we first evaluated the alterations in microbiome structure at the phylum level using the relative bacterial abundance per tissue type (tumor, adjacent non-tumor, and FD tissues) and per mucin phenotype (Fig. 3A). Overall, significant differences in abundance for *Patescibacteria* were found between (1) FD and adjacent non-tumor tissues ( $P=0.024$ ), (2) FD and tumor tissues ( $P=0.014$ ), and (3) FD tissues and tumors with intestinal ( $P=0.057$ ), mixed ( $P=0.068$ ), and null (0.0191) mucin phenotypes (Fig. 3A). Significant differences in abundance were also noted for *Campylobacterota* between tumors with a null mucin phenotype and gastric ( $P=0.0126$ ) or intestinal ( $P=0.0295$ ) mucin phenotype which can be assigned to the differential abundance of the *Helicobacter* genus among the phenotypes. Additionally, *Bacteroidota* differed significantly in abundance between tumors with a null mucin phenotype and those with a gastric ( $P=0.0067$ ), intestinal ( $P=0.0032$ ), or mixed ( $P=0.0249$ ) mucin phenotype (Fig. 3A). This latter phylum was also significantly altered between the FD cases and tumors with a null mucin phenotype ( $P=0.0079$ ; Fig. 3A).

(See figure on next page.)

**Fig. 1** Aberrant mucin signatures in the stomach of GC patients compared to FD patients. **A** Relative mRNA expression of gastric (*MUC1*, *MUC5AC*, and *MUC6*;  $n=100$ ) and intestinal mucins (*MUC2*, *MUC4*, and *MUC13*;  $n=100$ , 99, and 97, respectively) in gastric biopsies from FD patients ( $n=20$ ) and paired (highlighted by gray dashed lines) tumor and adjacent non-tumor tissues of 3 GC patient cohorts. Significant differences between FD, tumor, and adjacent non-tumor tissues are indicated by \* $P<0.05$ , \*\* $P<0.01$ , \*\*\* $P<0.001$ , and \*\*\*\* $P<0.0001$  (Wilcoxon rank-sum test). **B** The bar sizes represent the number of patients per mucine phenotype (null,  $n=13$ ; intestinal,  $n=19$ ; mixed,  $n=47$ ; null,  $n=17$ ) or expression level (i.e., high, mid, or low expression of *MUC1*, *MUC5AC*, *MUC6*, *MUC2*, *MUC4*, and *MUC13*). For each mucin expression level, the mean CNRQ (SD) is shown here: *MUC1*, high: 3.01 (2.05), mid: 1.07 (0.08), low: 0.46 (0.29); *MUC5AC*, high: 10.8 (9.04), mid: 1.38 (0.03), low: 0.13 (0.22); *MUC6*, high: 9.72 (10.82), mid: 1.39 (0.2), low: 0.11 (0.2); *MUC2*, high: 15.94 (25.98), mid: 1.35 (0.13), low: 0.33 (0.28); *MUC4*, high: 14.72 (22.91), mid: 1.29 (0.19), low: 0.35 (0.3); and *MUC13*, high: 5.27 (2.88), mid: 1.56 (0.33), low: 0.47 (0.3). **C** PCA plot based on mucin mRNA expression in tumor and adjacent non-tumor tissues of 3 GC patient cohorts ( $n=97-100$ ). PC1 explains 38.5% of the variation; PC2 explains 21.57% of the variation. **D** Immunohistochemistry was assessed to analyze *MUC1*, *MUC2*, *MUC5AC*, *MUC4*, *MUC6*, and *MUC13* protein expression in the different tissue types ( $n=5$  per tissue type; i.e., FD, adjacent non-tumor, and gastric adenocarcinomas with gastric, intestinal, mixed, or null mucin phenotypes). Representative images were selected. Pictures were taken at 20 × magnification and scale bars are 20 μm or 50 μm



**Fig. 1** (See legend on previous page.)

In support of the above, changes in community composition were further investigated using alpha (i.e., within samples) and beta (i.e., between samples) diversity measures at the genus level (Fig. 3B, C). Tumor samples with a null mucin phenotype showed a significantly lower alpha diversity compared to those with an intestinal mucin phenotype based on all four indexes investigated and with a mixed mucin phenotype based on the Chao1 and observed richness only (Fig. 3B). When considering the observed richness and Shannon indexes, a significant decrease in alpha diversity of tumors with a null mucin phenotype compared to those with a gastric mucin phenotype was also noted (Fig. 3B). Beta diversity was analyzed using the Bray–Curtis and weighted UniFrac phylogenetic distance metrics and visualized in Principal Coordinate Analysis (PCoA) plots (Fig. 3C). The first two axes captured 24.1% and 16.5% for the Bray–Curtis distances and was similar to the results for the weighted UniFrac distances (24.5% and 15%). Interestingly, tumor samples with an intestinal mucin phenotype were more spread over axis 2 while tumors with a gastric phenotype remained strongly clustered around the FD samples. Contrary, when testing for differences in community composition using ANOSIM, no significant differences were found (Fig. 3C).

To identify the bacterial taxa that are differentially present in GC compared to FD cases, an ALDEx2 analysis was conducted. Overall, the genera *Veillonella*, *Porphyromonas*, and *Prevotella* were found to be depleted in GC compared to the FD group whereas the abundance of *Corynebacterium*, *Fusobacterium*, *Streptococcus*, *Porphyromonas*, and *Prevotella* differed significantly between paired tumor and non-tumor tissues (Fig. S2). To further investigate the influence of aberrant mucin expression on bacterial enrichment or depletion in gastric tumors, differential abundance analysis and Spearman's correlation tests were performed assessing the association between bacterial genera and individual mucins or mucin phenotypes. The *Helicobacter* genus was found to be enriched in tumors with a null mucin phenotype compared to the other phenotypes (Fig. S3) and in tumor samples with low *MUC5AC* (Fig. 4; Fig. S5) expression. Also, *Megasphaera* was enriched in samples with low *MUC5AC* expression (Fig. 4; Fig. S5). In gastric adenocarcinomas with low *MUC1* mRNA levels, an abundance of *Porphyromonas* was observed (Fig. 4; Fig. S4). Regarding associations

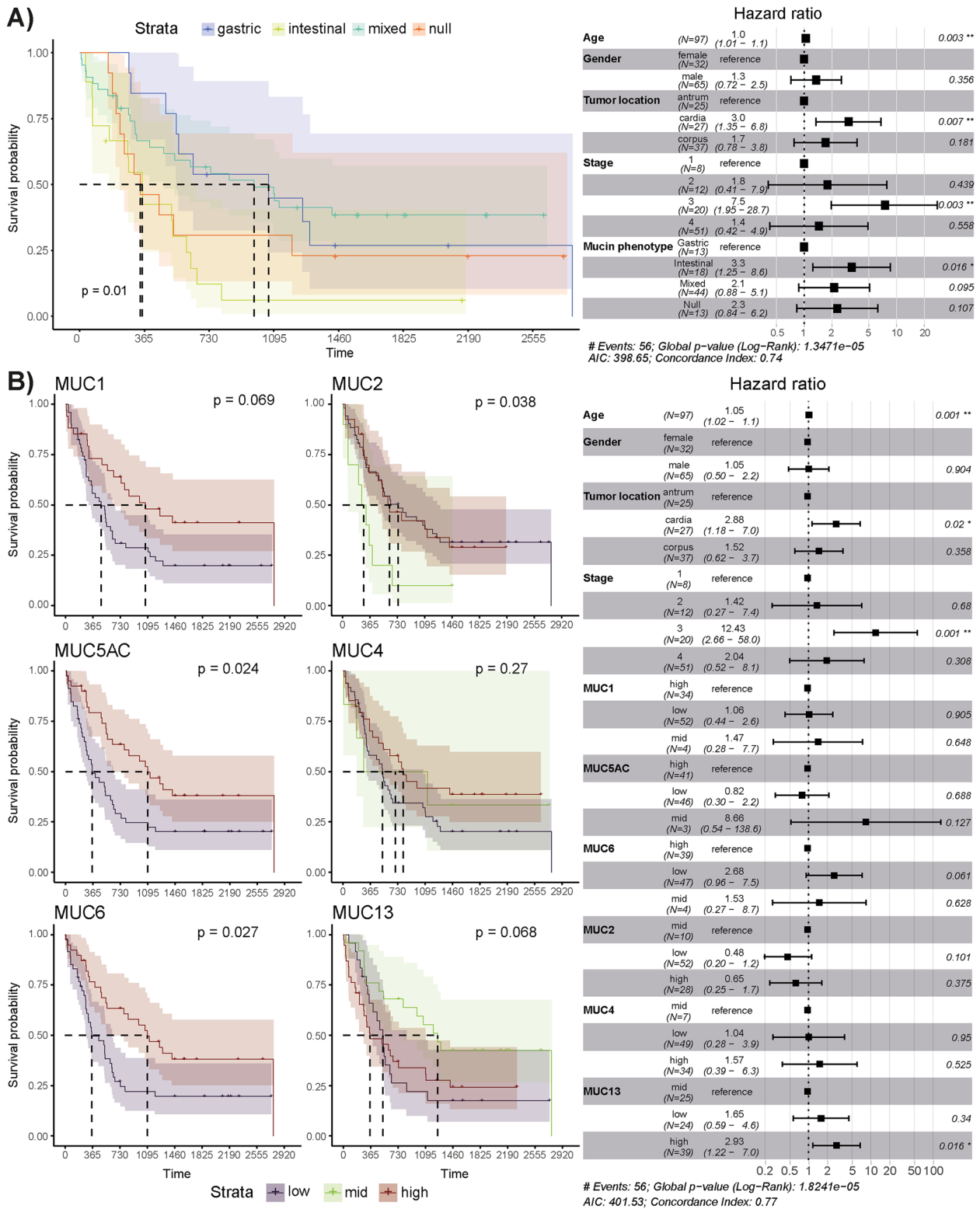
between bacterial genera and intestinal mucin expression, depletion in *Streptococcus* was identified in gastric adenocarcinomas with high *MUC2* mRNA levels (Fig. 4; Fig. S6) and *Lactobacillus* was significantly more abundant in tumors with mid-level *MUC4* mRNA expression (Fig. 4; Fig. S7). Interestingly, *Lactobacillus*, *Neisseria*, *Prevotella*, and *Veillonella* were enriched in gastric tumor samples with high *MUC13* expression (Fig. 4; Fig. S8). These latter genera also significantly correlated with *MUC13* mRNA levels as was also seen for *Helicobacter* (Fig. S1). Furthermore, *Lactobacillus* and *Neisseria* positively correlated with *MUC4* and *MUC2* mRNA expression, respectively (Fig. S1), whereas *Porphyromonas* and *Corynebacterium* significantly associated with the gastric mucin expression levels (Fig. S1).

#### Distinct community complexity in gastric adenocarcinoma with different mucin phenotypes

Co-occurrence and co-excluding interactions were analyzed using the probabilistic model of species co-occurrence to estimate positive and negative associations among bacterial genera in GC [42]. Figure 5 shows the bacterial networks identified in the different mucin phenotype groups and gastric tumors with low and high *MUC13* expression. The overall number of interactions differed significantly between the different mucin phenotype groups (Pearson's  $\chi^2$ ,  $P < 0.0001$ ), with most interactions seen in tumors with mixed and intestinal mucin phenotypes (Fig. 5B, C; Table S4). The same pattern was also seen for the number of positive and negative associations (Pearson's  $\chi^2$ ,  $P < 0.0001$ ). Furthermore, the ratio of co-excluding to co-occurring interactions differed between the different mucin phenotypes with relatively more co-occurring interactions in tumors with an intestinal, mixed, and null mucin phenotype (negative to positive ratio: 0.23, 0.30, and 0.25, respectively; Fig. 5B–D) while those with a gastric mucin phenotype had more co-excluding than co-occurring interactions (1.4; Fig. 5A). In addition, a Pearson's  $\chi^2$  test was done at genus level to test for differences in contribution to the network of different genera between gastric tumors assigned to the different mucin phenotypes. The residuals (i.e., the difference between the expected value and the observed one) were used as a measure of correlation (i.e., a high residual means a bigger contribution of the genus to a significant statistical test), and only those

(See figure on next page.)

**Fig. 2** Intestinal mucin phenotype and aberrant *MUC13* expression correlate with worse survival in GC patients. **A** Kaplan–Meier curve (left-hand side) and Cox-proportional hazards model (right-hand side) for survival analysis between patients with gastric tumors assigned to different mucin phenotypes (**A**) and showing low, mid, or high mRNA levels of a gastric (*MUC1*, *MUC5AC*, *MUC6*) or intestinal (*MUC2*, *MUC4*, *MUC13*) mucin (**B**). For the gastric mucins (*MUC5AC*, *MUC6*, and *MUC1*), the mid-level expression data was excluded due to a small number of observations ( $n \leq 4$ ). A forest plot of the Cox-proportional hazards model is shown (right). For the Kaplan–Meier curves (left), the  $P$  values were calculated using the log-rank test whereas the Wald test statistic was performed for the Cox-proportional hazards model ( $n$  for each group is shown on the forest plot)



**Fig. 2** (See legend on previous page.)



greater than two were considered. In adenocarcinomas with a gastric mucin phenotype, *Lachnoanaerobaculum*, *Gemella*, and *Reyranella* had proportionally more interactions compared to tumors with another mucin phenotype (Fig. 5A; Table S5). *Selenomonas* and *Treponema* had proportionally more influence on the bacterial community of samples with an intestinal mucin phenotype, while *Rothia* and *Prevotella* had a higher contribution in tumor samples with a null mucin phenotype (Fig. 5; Table S5). In gastric adenocarcinoma with an intestinal mucin phenotype, most interactions were seen for *Parvimonas*, *Sediminibacterium*, *Helicobacter*, *Selenomonas*, *Fusobacterium*, *Reyranella*, *Treponema*, *Leptotrichia*, *Haemophilus*, *Neisseria*, *Veillonella*, *Prevotella*, and *Streptococcus* (Fig. 5B; Table S5). Bacterial genera with the most interactions in tumor samples with a mixed mucin phenotype were *Alloprevotella*, *Fusobacterium*, *Sediminibacterium*, *Neisseria*, *Veillonella*, *Reyranella*, *Rothia*, *Streptococcus*, *Haemophilus*, *Parvimonas*, *Prevotella*, and *Helicobacter* (Fig. 5C; Table S5), whereas *Rothia*, *Prevotella*, and *Helicobacter* contributed more in samples with a null mucin phenotype (Fig. 5D; Table S5).

Finally, *Veillonella*, *Neisseria*, and *Prevotella* were found to play a role in shaping the community structure in tumors with high *MUC13* expression (Fig. 5E, F; Table S6).

### Enrichment of oral and intestinal microbes in GC depends on the mucin phenotype

To investigate whether oral and intestinal bacteria have a preference for a certain mucin phenotype in gastric tumors, we determined the overall distribution of oral and intestinal genera by profiling the sequences of all samples against the Human Oral Microbiome and the Human Intestinal 16S rRNA gene reference databases [40, 41]. The microbiome of the FD cohort showed a trend of enrichment in oral microbial taxa compared to the tumor and tumor-adjacent microbiome ( $P=0.10$ ; Fig. S9). When comparing the tumor samples with different mucin phenotypes to the FD cohort, a significant depletion in oral taxa is seen in the null mucin phenotype ( $P=0.018$ ) and a trend towards depletion is also seen in tumors with a mixed mucin phenotype ( $P=0.12$ ; Fig. S9). Additionally, when considering the tumor samples only,

a trend in oral microbial enrichment was seen in samples with an intestinal mucin phenotype compared to the null mucin phenotype ( $P=0.1$ ; Fig. S9). Inversely, for the intestinal microbial species, a significant enrichment was seen in tumor samples with a null mucin phenotype compared to those with a gastric ( $P=0.027$ ) and intestinal ( $P=0.02$ ) mucin phenotype. Of note, a similar trend was also seen between the null and mixed mucin phenotype tumors ( $P=0.064$ ; Fig. S9).

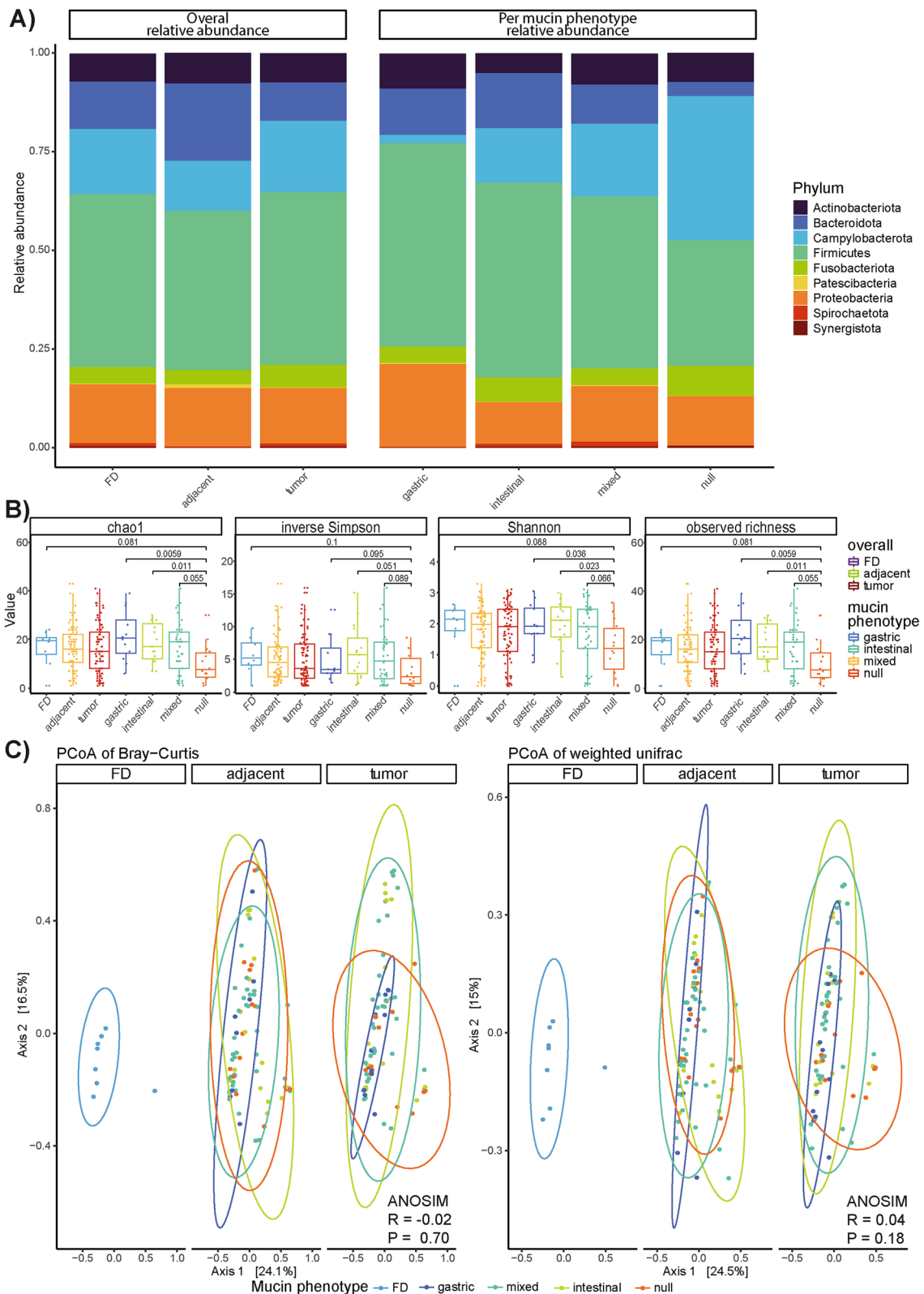
### Microbiome functional capacity in GC is defined by the mucin phenotype

The functional capacity of the mucosa-associated microbiome for each sample type (FD, tumor, and adjacent non-tumor tissues) and mucin phenotype or aberrantly expressed mucin in gastric tumors was estimated through metagenomic inference using PICRUSt2. Differences in relative pathway abundance were tested using ALDEx2. When considering the FD, tumor, and adjacent non-tumor samples, three pathways, i.e., purine, guanosine, and adenosine nucleotide degradation, were found to be differentially abundant and enriched in the GC cohort (Fig. 6A). Furthermore, five predicted KEGG pathways were found to be differentially represented between the tumors assigned to different mucin phenotypes with most of the pathways being depleted in tumors with a null mucin phenotype except for the fucose degradation pathway (Fig. 6B).

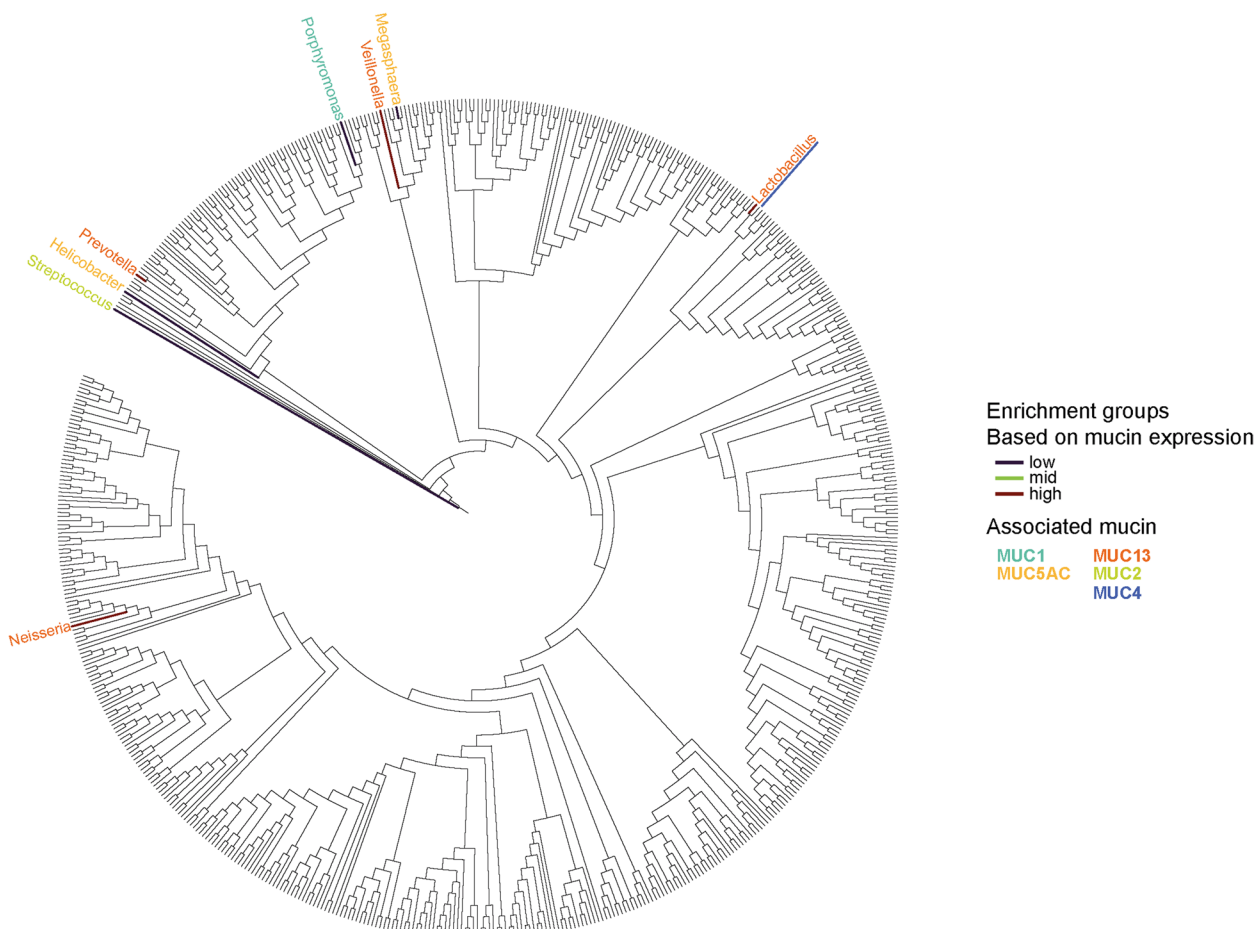
This assessment was repeated for each mucin and is summarized in Table S7. For the *MUC1* and *MUC2* mRNA expression levels, no differentially abundant pathways were found. Interestingly, in samples with low *MUC6* and *MUC5AC* expression, the purine ribonucleosides degradation pathway was depleted. In contrast, an enrichment of sugar degrading pathways was seen in samples with a high *MUC13* expression (i.e., sucrose degradation IV (sucrose phosphorylase), glycogen degradation I (bacterial), and galactose degradation I (Leloir pathway)). Additionally, heterolactic fermentation was also enriched in high *MUC13* expressing samples and pyruvate fermentation to propanoate I was depleted in the samples with a normal *MUC13* expression compared to the high and low

(See figure on next page.)

**Fig. 3** Differences in microbiota composition between FD, tumor, and adjacent non-tumor tissues and between tumors with gastric, intestinal, mixed, and null mucin phenotypes. **A** Mean relative abundance of phyla in all tissue types (FD, tumor, and adjacent non-tumor samples;  $n=8$ , 83, and 80, respectively) and in each mucin phenotype group assigned to the gastric tumor samples (gastric, intestinal, mixed, and null;  $n=12$ , 15, 41, and 14, respectively). **B** Boxplots of four common alpha-diversity indices (i.e., chao1, inverse Simpson, observed richness, and Shannon index) for FD, tumor, and adjacent non-tumor tissues ( $n=8$ , 83, and 80, respectively) as well as for tumors with gastric, intestinal, mixed, and null mucin phenotypes ( $n=12$ , 15, 41, and 14, respectively). **C** PCoA using the Bray–Curtis (left) and weighted unifrac (right) distance measures for FD, tumor, and adjacent non-tumor tissue samples ( $n=8$ , 83, and 80, respectively). The points were colored according to the mucin phenotype assigned to each tumor. The percentage of variance captured by the axes and the 95% confidence intervals per mucin phenotype (drawn as ellipses) are also shown



**Fig. 3** (See legend on previous page.)



**Fig. 4** Phylogenetic tree of GC-enriched and GC-depleted bacteria associated with aberrant mucin expression. Genera that are significantly abundant or underrepresented in gastric tumors ( $n = 82$ ) with mid, low, or high mRNA expression of a gastric (MUC1, MUC5AC, MUC6) or intestinal (MUC2, MUC4, MUC13) mucin are shown by color. The colored branches highlight the mucin mRNA expression levels (low, mid, or high) whereas the corresponding-colored genera represent the mucins they associate with. In case of overlapping mucin associations, the color of the underlining indicates the second mucin association

levels. An enrichment of the sucrose degradation IV (sucrose phosphorylase) pathway was also seen in samples with high *MUC2* expression. The abovementioned changes in the metabolic potential of the gastric cancer microbiome suggest a high reliance on mucins as a food source.

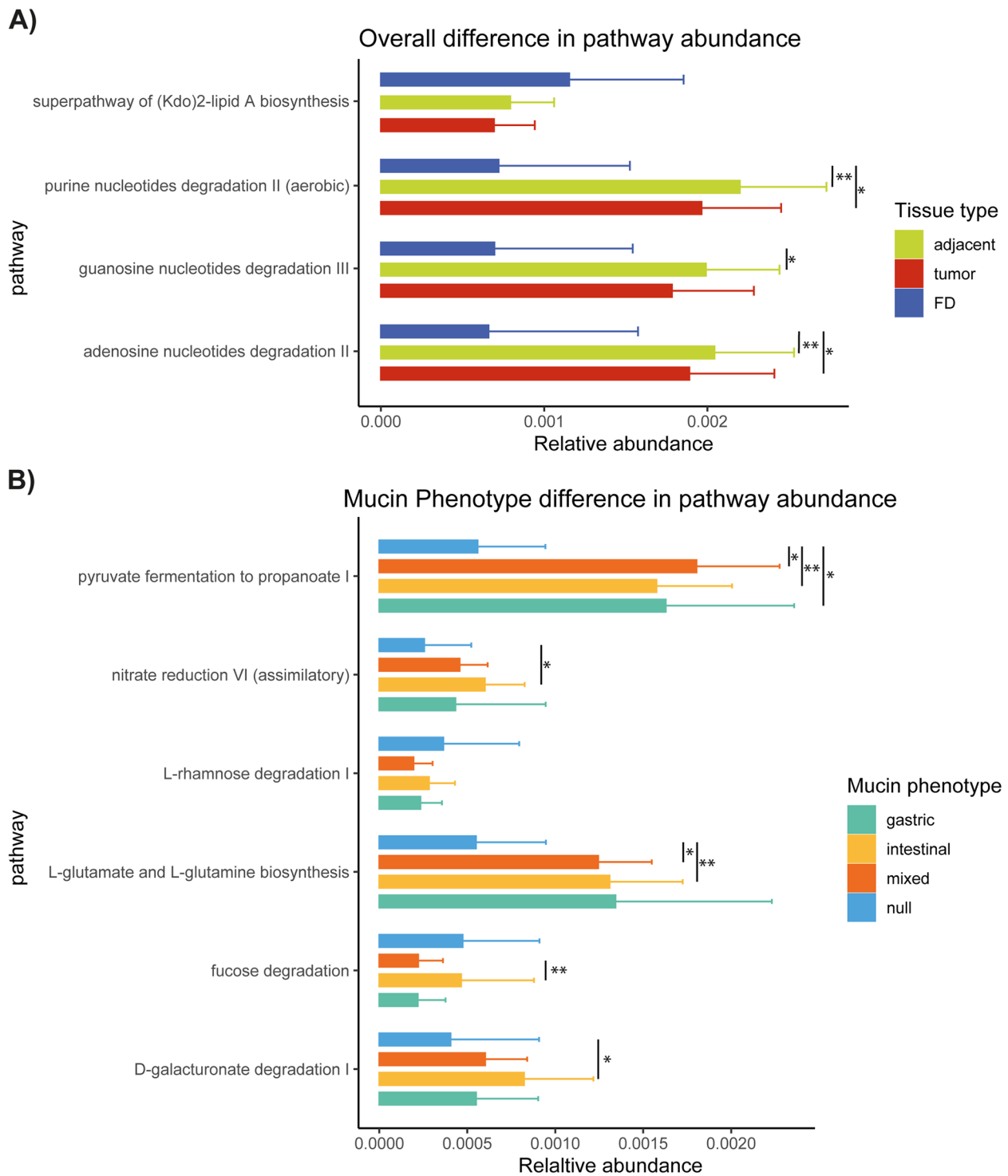
### Discussion

It is generally well-accepted that microbial dysbiosis is a dynamic process correlating with the progression to gastric cancer [7]. Previous studies always considered the whole stomach as one habitat while alterations in gastric mucosal microbiota across different stomach

(See figure on next page.)

**Fig. 5** Correlation of GC-enriched and GC-depleted genera associated with a mucin phenotype or aberrant MUC13 expression. Each sphere represents the bacterial network of gastric adenocarcinomas with **A** gastric mucin phenotype ( $n = 12$ ), **B** intestinal mucin phenotype ( $n = 15$ ), **C** mixed mucin phenotype ( $n = 41$ ), **D** null mucin phenotype ( $n = 14$ ), **E** low MUC13 mRNA expression ( $n = 26$ ), and **F** high MUC13 mRNA expression ( $n = 34$ ). Each genus found to participate in the bacterial community is represented by a point and connected through lines with interacting bacterial taxa. The red lines are positive associations between bacterial genera while the blue dotted lines are negative interactions. The point shape represents the association between a genus and a specific mucin (i.e., the genus is found to be differentially abundant with different expression levels of the mucin). The color of the point (+ genus name) represents the mucin mRNA expression level in which the genus was found to be enriched. If the genus was not found to be differentially abundant with different mucin expression levels, the shape is a full circle that is either gray (does not directly interact with other genera associated with a specific mucin) or black (the genus interacts directly with a mucin-associated bacterial taxa)





**Fig. 6** Relative abundance of metagenomic pathways is defined by the mucin phenotype present in gastric adenocarcinomas. Functional classification of the predicted metagenome content of the microbiota of FD, tumor, and adjacent non-tumor samples ( $n = 8, 83,$  and  $80,$  respectively) **(A)** and of gastric adenocarcinomas with gastric, intestinal, mixed, or null mucin phenotypes ( $n = 12, 15, 41,$  and  $14,$  respectively) **(B)** using PICRUSt2. Significance was considered for  $P < 0.05$  and indicated by  $*P < 0.05$  and  $**P < 0.01$

microhabitats (i.e., tumoral and peritumoral microhabitats) also occurs [7]. Therefore, adjacent non-tumor tissues should also be taken into account as premalignant gastric lesions (i.e., atrophy and intestinal metaplasia) can occur revealing cancer-related features such as mucin and microbial signatures with predictive potential for malignant transformation [46]. Furthermore, changes described in microbial composition due to the overrepresentation of arising new taxa remain inconsistent in gastric adenocarcinomas [7, 47]. This discrepancy may be due to tumor microenvironment heterogeneity with variable expression of gastric and intestinal mucins shaping the microbiota community and influencing disease outcome. In this study, we assigned gastric adenocarcinomas to four different phenotypes based on mucin expression, with the intestinal mucin phenotype being significantly associated with a worse survival rate. This finding was further substantiated by high-level *MUC13* expression which also correlated with an unfavorable outcome, highlighting a key role for this intestinal mucin in gastric carcinogenesis. *MUC13* overexpression has previously been described in GC and such aberrant *MUC13* signaling is known to protect colorectal cancer cells from death via NF- $\kappa$ B pathway activation thereby impacting on therapeutic efficacy and disease outcome [12, 48]. Whether the upregulation of *MUC13* has similar capacities to inhibit gastric tumor cell death requires further investigation. In addition, decreased expression levels of *MUC1*, *MUC5AC*, and *MUC6* also associated with a poorer outcome, highlighting the importance of gastric mucins in GC to increase the chances of survival [19, 49–51].

In terms of the composition of the gastric microbiota, distinct dysbiosis in gastric adenocarcinomas compared to our FD cohort and along with differences based on mucin phenotype groups was observed, with the known gastric phyla being widely present, including *Campylobacterota* [52, 53]. This latter phylum, previously part of the *Proteobacteria*, mainly consisted of *Helicobacter* genus members in both our tumor and FD samples. This shows that the FD group was colonized with *Helicobacter*, despite no gastric lesions being found, further highlighting that this genus is a core member of the normal gastric microbiota as determined previously [7]. Significant differences in *Helicobacter* abundance was noted among the different mucin phenotype groups, with the highest presence found in tumors with a null mucin phenotype [54]. This can be explained by a decrease in gastric mucin expression, a characteristic of these adenocarcinomas. More specifically, *MUC5AC* provides an important adherence site for *H. pylori*, but its absence can facilitate colonization whereas *MUC6* has antibiotic properties against *H. pylori* [55].

Reduced microbial diversity is an overall major feature in many disease states, including inflammatory bowel diseases and cancer [53, 56]. In our different patient cohorts, the lowest alpha diversity was also seen in gastric adenocarcinomas compared to FD and adjacent non-tumor tissues, and more specifically in tumors with a null mucin phenotype possibly due to a higher *Helicobacter* abundance. In contrast, no differences in beta diversity were found which could in part be due to different confounding factors influencing microbiome composition, such as age, gender, diet, and drug use [57, 58]. More specifically, microbial diversity changes throughout the human life span and is known to be associated with gender whereas diet and the usage of drugs induce temporary shifts in gut microbiota composition [59–61].

Interestingly, from our genus/species level classification, we observed that several taxa, including *Corynebacterium*, *Fusobacterium*, *Streptococcus*, *Porphyromonas*, *Veillonella*, and *Prevotella*, significantly differed in abundance between tumor and adjacent non-tumor or FD tissues. These abovementioned genera have already been linked to gastrointestinal cancers [5, 6, 62–65]. When investigating the GC microbiome in more detail, other bacterial genera were also found to be associated with aberrant mucin expression in the tumor. More specifically, *Lactobacillus* was found to be significantly more abundant in adenocarcinomas with high *MUC4* and *MUC13* expression. Members of this genus normally reside in the intestinal mucosa but are capable of colonizing and proliferating in the cancerous stomach [5, 20, 66] and most probably in the presence of an intestinal mucin environment. Furthermore, *Prevotella*, *Veillonella*, and *Neisseria* seemed also to have a higher affinity for tumors with *MUC13* overexpression which are in fact inhabitants of the oral cavity but can be opportunistic pathogens [6, 52, 53, 67]. As oral microbiota were also enriched in tumors with an intestinal mucin phenotype, a shift from predominantly gastric to intestinal mucins could thus affect the abundance of pro-inflammatory oral microbes in intestinal mucin phenotype tumors, specifically those with *MUC13* overexpression. Whether these oral bacteria play a role in *MUC13*-driven gastric carcinogenesis warrants further investigation. Importantly, we also identified the oral taxa *Porphyromonas* and *Megasphaera* to be enriched in adenocarcinoma with low *MUC1* and *MUC5AC* expression, respectively. Both genera have been associated with worse survival in GI cancers, further underlining the favorable role of gastric mucin abundance in GC outcomes [68].

Not only individual bacterial drivers, but also the microbial networks that reside within define the disease-specific microenvironment [4, 69]. As such, we observed the overall highest co-occurring and co-excluding

interactions of enriched and depleted bacteria in tumors with a mixed and intestinal mucin phenotype. Several bacterial genera were identified to play a role in shaping the microbial community. In particular, the potential role of oral pathogenic taxa in GC is highlighted by the observed centralities of *Lachnoanaerobaculum*, *Gemella*, and *Reyranella* in gastric mucin phenotype tumors; *Rothia* and *Prevotella* in null mucin phenotype tumors; and *Neisseria*, *Veillonella*, and *Prevotella* in both intestinal and mixed mucin phenotype tumors and *Fusobacterium* [70]. Although *Fusobacterium* is not differentially abundant with regard to mucin expression or phenotype, this genus is known to be an important player in gastrointestinal cancers [71–73].

After having analyzed the diversity and composition of the gastric microbiota in relation to aberrant mucin expression in GC, we finally addressed the functional features of the GC microbiota which can affect host metabolism [6]. We demonstrated an overall decrease in metabolic activity in tumors with a null mucin phenotype which could be explained by a decline in bacterial community complexity. When investigating metabolic changes between tumors with different mucin expression levels, we observed predicted functional shifts in short-chain fatty acid (SCFA) fermentation, amino acid, and sugar degradation that may reflect compositional differences in mucin expression in the tumor microenvironment. Specifically, the changes in carbohydrate digestion are predictive of bacterial production of SCFAs, which have been linked to the hyperproliferation of cells in gastrointestinal cancers [6]. Furthermore, bacterial SCFAs have also been shown to stimulate intestinal mucin expression, suggesting their importance in intestinal mucin phenotype tumors [20].

## Conclusions

Taken together, our study identified distinct mucin-microbiome signatures shaping the tumor microenvironment in gastric cancer, with an intestinal or aberrant MUC13 mucin environment associated with a poor outcome. We also showed that members of the oral pathogenic taxa, such as *Neisseria*, *Prevotella*, and *Veillonella*, are potential drivers in MUC13-mediated signaling in GC which could be useful biomarkers in predicting disease outcomes. As not all pre-malignant gastric conditions will eventually evolve into cancer, such oral taxa-MUC13 signatures in patients with pre-cancerous stadia could also help predict the potential to further evolve into cancer. Furthermore, adenocarcinomas with an intestinal mucin phenotype do favor the establishment of pro-inflammatory oral bacteria, forming strong co-occurrence networks. Ultimately, understanding these mucin-microbiome signatures in gastric carcinogenesis

may impact GC prevention and treatment strategies and adequate independent external validation in other GC cohorts is therefore recommended. Additionally, future research must also consider the inclusion of healthy patients as controls, since it currently remains unclear how microbiota composition relates to FD [74]. Nevertheless, this is the first study implicating mucins in both dysbiosis and disease outcomes in gastric cancer.

## Abbreviations

ANOSIM	Analysis of similarities
ASV	Amplicon sequence variants
CNRQ	Calibrated normalized relative quantities
CI	Confidence interval
EC	Enzyme commission
FD	Functional dyspepsia
GC	Gastric cancer
HOM	Human oral microbiome
<i>H. pylori</i>	<i>Helicobacter pylori</i>
IHC	Immunohistochemistry
LL	Lower limit
MUC	Mucin
PCoA	Principal coordinate analysis
SPEM	Spasmolytic polypeptide expressing metaplasia
SCFA	Short-chain fatty acid
UL	Upper limit

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-023-01534-w>.

**Additional file 1: Figure S1.** Associations of mucin mRNA expression with bacterial abundance and clinical data of GC patients. Correlogram of GC ( $n=108$ ) patients. Spearman's rank order correlation values ( $r$ ) are shown from blue ( $-1.0$ ) to red ( $1.0$ );  $r$  values are indicated by color.  $P$  values are indicated by black asterisks (\* $<0.05$ ; \*\* $<0.01$ ; \*\*\* $<0.001$ ). The considered parameters are age, gender ( $n=108$ ), tumor stage ( $n=101$ ), Lauren's classification ( $n=106$ ), survival (i.e. deceased or alive after 5 years follow-up;  $n=67$  and  $30$ ; respectively), MUC1 ( $n=100$ ), MUC5AC ( $n=100$ ), MUC6 ( $n=100$ ), MUC2 ( $n=100$ ), MUC4 ( $n=99$ ) and MUC13 ( $n=98$ ) mRNA expression.

**Additional file 2: Figure S2.** Relative bacterial abundance of the genera found to be differentially abundant between control, tumor adjacent and tumor tissue. The relative abundance of each bacterial genus found to be differentially abundant between control and paired (highlighted by grey dashed lines) tumor and adjacent non-tumor tissues ( $n=8$ ,  $83$  and  $80$ ; respectively) using ALDEx2 is shown.  $P$ -values found to be significant are shown on the plots and were calculated using a Wilcoxon rank sum test.

**Additional file 3: Figure S3.** Relative bacterial abundance of the genera found to be differentially abundant between tumor tissues having different mucin phenotypes. The relative abundance of each bacterial genus found to be differentially abundant between tumor tissues having different mucin phenotypes (gastric, intestinal, mixed and null;  $n=10$ ,  $15$ ,  $41$  and  $14$ ; respectively) using ALDEx2 are shown.  $P$ -values found to be significant are shown on the plots and were calculated using a Wilcoxon rank sum test.

**Additional file 4: Figure S4.** Relative bacterial abundance of the genera found to be differentially abundant between tumor tissues having a high or low MUC1 mRNA expression. The relative abundance of each bacterial genus found to be differentially abundant between tumor samples having a high or low MUC1 mRNA expression level ( $n=38$  and  $41$ ; respectively) using ALDEx2 is shown.  $P$ -values found to be significant are shown on the plots and calculated using a Wilcoxon rank sum test.

**Additional file 5: Figure S5.** Relative bacterial abundance of the genera found to be differentially abundant between tumor tissues having a high or low MUC5AC mRNA expression. The relative abundance of each bacterial genus found to be differentially abundant between tumor samples having a high or low MUC5AC mRNA expression level ( $n=38$  and  $41$ ; respectively) using ALDEx2 is shown.  $P$ -values found to be significant are shown on the plots and calculated using a Wilcoxon rank sum test.

**Additional file 6: Figure S6.** Relative bacterial abundance of the genera found to be differentially abundant between tumor tissues having a high, mid or low MUC2 mRNA expression. The relative abundance of each bacterial genus found to be differentially abundant between tumor samples having a high, mid or low MUC2 mRNA expression level ( $n=26, 9, 47$ ; respectively) using ALDEx2 is shown.  $P$ -values found to be significant are shown on the plots and calculated using a Wilcoxon rank sum test.

**Additional file 7: Figure S7.** Relative bacterial abundance of the genera found to be differentially abundant between tumor tissues having a high, mid or low MUC4 mRNA expression. The relative abundance of each bacterial genus found to be differentially abundant between tumor samples having a high, mid or low MUC4 mRNA expression level ( $n=28, 7, 47$ ; respectively) using ALDEx2 is shown.  $P$ -values found to be significant are shown on the plots and calculated using a Wilcoxon rank sum test.

**Additional file 8: Figure S8.** Relative bacterial abundance of the genera found to be differentially abundant between tumor tissues having a high, mid or low MUC13 mRNA expression. The relative abundance of each bacterial genus found to be differentially abundant between tumor samples having a high, mid or low MUC13 mRNA expression level ( $n=34, 22, 26$ ; respectively) using ALDEx2 is shown.  $P$ -values found to be significant are shown on the plots and calculated using a Wilcoxon rank sum test.

**Additional file 9: Figure S9.** Pooled relative abundance of oral and intestinal bacterial species for the FD, tumor and non-tumor adjacent samples and the tumor tissues further divided according to their mucin phenotype. The relative abundance of ASV's classified up to species level using the HOM and HIT databases were pooled per sample, according to their preferred habitat as detailed in the HOM-database, into either being part of the oral or intestinal microbiome. The relative abundances are plotted for the control, tumor and non-tumor adjacent tissues ( $n=8, 83$  and  $80$ ; respectively). For the tumor tissues, the relative abundance of the oral and intestinal species was also shown according to the respective mucin phenotype of the tumor tissue (gastric, intestinal, mixed and null;  $n=10, 15, 41$  and  $14$ ; respectively). Significant differences between control, tumor and non-tumor adjacent tissues and gastric adenocarcinomas with different mucin phenotypes are shown on the plots and calculated using a Wilcoxon rank sum test.

**Additional file 10: Table S1.** demographic information of the included patients. **Table S2.** overview of the used QuantiTect primers. **Table S3.** 90% confidence interval of the relative mucin mRNA expression levels of functional dyspepsia patients. **Table S4.** number of interactions within the bacterial communities of samples with different mucin phenotypes. **Table S5.** Analysis of interactions in bacterial communities associated with different mucin phenotypes. **Table S6.** Analysis of interactions in bacterial communities associated with different expression level of MUC13. **Table S7.** Differentially abundant metagenomic pathways.

#### Acknowledgements

We thank Lieve Vits, Angelika Jürgens, Mandy Vermont, and Ines Tuyaerts for their excellent technical assistance in the processing of the gastric samples, performance of the RT-qPCR, and IHC assays and for providing sequencing services.

#### Authors' contributions

BO and AS conceived and designed the study. AS and BDW supervised the study. BO wrote the first draft of the manuscript, analysed the raw data, interpreted the obtained results, performed the statistical analyses, and designed the graphs. HC, WA and SL provided intellectual support in the study design. KG, HDS, MP, JK, AL supervised the inclusion of the dyspepsia and gastric cancer patients and provided intellectual support regarding the clinical

patient data sets. All authors discussed the results and commented on the manuscript. The author(s) read and approved the final manuscript.

#### Funding

This work was supported by the Antwerp University Research Funds (BOF DOCPRO4 grant number 39958 and BOF-TOP grant number 35018) and the Antwerp University Valorisation Fund (IOF-SBO number 42601).

#### Availability of data and materials

The raw 16S rRNA sequencing data that support the findings of this study are available from the NCBI Sequence Read Archive (BioProject: PRJNA924141). All R markdown files used for the data analysis of the paper can be found on GitHub (BaptisteO/MucinMicrobiomeGC). All other data supporting the findings of this study are available within the article and its supplementary information files.

#### Declarations

##### Ethics approval and consent to participate

All recruitment and study procedures involving human subjects were approved by the relevant ethics committees (Ethical Committee of the Antwerp University Hospital and the Kaunas Regional Ethics Committee). All subjects provided written informed consent prior to enrolment.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare no competing interests.

##### Author details

<sup>1</sup>Laboratory of Experimental Medicine and Paediatrics, Faculty of Medicine and Health Sciences, University of Antwerp, Universiteitsplein 1, 2610 Antwerp, Wilrijk, Belgium. <sup>2</sup>Infla-Med Research Consortium of Excellence, University of Antwerp, Antwerp, Belgium. <sup>3</sup>Pathology Department, Gent University Hospital, Ghent, Belgium. <sup>4</sup>Division of Gastroenterology and Hepatology, Antwerp University Hospital, Edegem, Belgium. <sup>5</sup>Department of Oncology, Antwerp University Hospital, Edegem, Belgium. <sup>6</sup>Department of Bioscience Engineering, University of Antwerp, Antwerp, Belgium. <sup>7</sup>Department of Gastroenterology and Institute for Digestive Research, Lithuanian University of Health Sciences, Kaunas, Lithuania. <sup>8</sup>Microbiome Research Centre, St George and Sutherland Clinical School, University of New South Wales, Sydney, NSW, Australia. <sup>9</sup>Department of Gastroenterology, Hepatology and Infectious Diseases, Otto-Von-Guericke University, Magdeburg, Germany.

Received: 9 August 2022 Accepted: 22 March 2023

Published online: 21 April 2023

#### References

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71:209–49.
- Hunt RH, Camilleri M, Crowe SE, El-Omar EM, Fox JG, Kuipers EJ, et al. The stomach in health and disease. *Gut* BMJ Publishing Group. 2015;64:1650–68.
- Correa P. Human gastric carcinogenesis: a multistep and multifactorial process—First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. *Cancer Res.* 1992;52(24):6735–40.
- Lofgren JL, Whary MT, Ge Z, Muthupalani S, Taylor NS, Mobley M, et al. Lack of Commensal Flora in *Helicobacter pylori*-infected INS-GAS mice reduces gastritis and delays intraepithelial neoplasia. *Gastroenterology.* 2011;140:210–220.e4.
- Ferreira RM, Pereira-Marques J, Pinto-Ribeiro I, Costa JL, Carneiro F, Machado JC, et al. Gastric microbial community profiling reveals a dysbiotic cancer-associated microbiota. *Gut.* 2018;67:226–36.



6. Coker OO, Dai Z, Nie Y, Zhao G, Cao L, Nakatsu G, et al. Mucosal microbiome dysbiosis in gastric carcinogenesis. *Gut* BMJ Publishing Group. 2018;67:1024–32.
7. Rajilic-Stojanovic M, Figueiredo C, Smet A, Hansen R, Kupcinskas J, Rokkas T, et al. Systematic review: gastric microbiota in health and disease. *Aliment Pharmacol Ther*. 2020;51:582–602.
8. Smet A, Kupcinskas J, Link A, Hold GL, Bornschein J. The role of microbiota in gastrointestinal cancer and cancer treatment: chance or curse? *Cell Mol Gastroenterol Hepatol*. 2022;13:857–74.
9. Kim DH, Shin N, Kim GH, Song GA, Jeon T-Y, Kim D-H, et al. Mucin expression in gastric cancer: reappraisal of its clinicopathologic and prognostic significance. *Arch Pathol Lab Med*. 2013;137:1047–53.
10. Wakatsuki K, Yamada Y, Nariyo M, Ueno M, Takayama T, Tamaki H, et al. Clinicopathological and prognostic significance of mucin phenotype in gastric cancer. *J Surg Oncol*. 2008;98:124–9.
11. Sugai T, Tsukahara M, Endoh M, Shioi Y, Takebe N, Mue Y, et al. Analysis of cell cycle-related proteins in gastric intramucosal differentiated-type cancers based on mucin phenotypes: a novel hypothesis of early gastric carcinogenesis based on mucin phenotype. *BMC Gastroenterol*. 2010;10:55.
12. Shimamura T, Ito H, Shibahara J, Watanabe A, Hippo Y, Taniguchi H, et al. Overexpression of MUC13 is associated with intestinal-type gastric cancer. *Cancer Sci*. 2005;96:265–73.
13. Wang X-T, Kong F-B, Mai W, Li L, Pang L-M. MUC1 immunohistochemical expression as a prognostic factor in gastric cancer: meta-analysis. *Dis Markers*. 2016;2016:9421571.
14. Tamura Y, Higashi M, Kitamoto S, Yokoyama S, Osako M, Horinouchi M, et al. MUC4 and MUC1 expression in adenocarcinoma of the stomach correlates with vessel invasion and lymph node metastasis: an immunohistochemical study of early gastric cancer. *PLoS ONE*. 2012;7:e49251.
15. Wang R-Q, Fang D-C. Alterations of MUC1 and MUC3 expression in gastric carcinoma: relevance to patient clinicopathological features. *J Clin Pathol*. 2003;56:378–84.
16. Utsunomiya T, Yonezawa S, Sakamoto H, Kitamura H, Hokita S, Aiko T, et al. Expression of MUC1 and MUC2 mucins in gastric carcinomas: its relationship with the prognosis of the patients. *Clin Cancer Res*. 1998;4:2605–14.
17. Hwang I, Kang YN, Kim JY, Do YR, Song HS, Park KU. Prognostic significance of membrane-associated mucins 1 and 4 in gastric adenocarcinoma. *Exp Ther Med*. 2012;4:311–6 Spandidos Publications.
18. Leteurtre E, Zerimech F, Piessen G, Wacrenier A, Leroy X, Copin M-C, et al. Relationships between mucinog gastric carcinoma, MUC2 expression and survival. *World J Gastroenterol*. 2006;12:3324–31.
19. Wang J-Y, Chang C-T, Hsieh J-S, Lee L-W, Huang T-J, Chai C-Y, et al. Role of MUC1 and MUC5AC expressions as prognostic indicators in gastric carcinomas. *J Surg Oncol*. 2003;83:253–60.
20. Breugelmans T, Oosterlinck B, Arras W, Ceuleers H, De Man J, Hold GL, et al. The role of mucins in gastrointestinal barrier function during health and disease. *Lancet Gastroenterol Hepatol*. 2022;7:455–71.
21. Linden SK, Sutton P, Karlsson NG, Korolik V, McGuckin MA. Mucins in the mucosal barrier to infection. *Mucosal Immunol*. 2008;1:183–97.
22. van Putten JPM, Stribis K. Transmembrane mucins: signaling receptors at the intersection of inflammation and cancer. *J Innate Immun*. 2017;9:281–99.
23. Paone P, Cani PD. Mucus barrier, mucins and gut microbiota: the expected slimy partners? *Gut*. 2020;69:2232–43.
24. Cornick S, Tawiah A, Chadee K. Roles and regulation of the mucus barrier in the gut. *Tissue Barriers*. 2015;3:e982426.
25. Hu Y-L, Pang W, Huang Y, Zhang Y, Zhang C-J. The gastric microbiome is perturbed in advanced gastric adenocarcinoma identified through shotgun metagenomics. *Front Cell Infect Microbiol*. 2018;8:433.
26. Chen Y-C, Fang W-L, Wang R-F, Liu C-A, Yang M-H, Lo S-S, et al. Clinicopathological variation of lauren classification in gastric cancer. *Pathol Oncol Res*. 2016;22:197–202.
27. Smet A, Breugelmans T, Michiels J, Lamote K, Arras W, De Man JG, et al. A dynamic mucin mRNA signature associates with COVID-19 disease presentation and severity. *JCI Insight*. 2021;6:e151777.
28. Hellemans J, Mortier G, De Paeppe A, Speleman F, Vandesompele J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol*. 2007;8:R19.
29. Sinn H-P, Schneeweiss A, Keller M, Schlombs K, Laible M, Seitz J, et al. Comparison of immunohistochemistry with PCR for assessment of ER, PR, and Ki-67 and prediction of pathological complete response in breast cancer. *BMC Cancer*. 2017;17:124.
30. Breugelmans T, Van Spaendonck H, De Man JG, De Schepper HU, Jauregui-Amezaga A, Macken E, et al. In-depth study of transmembrane mucins in association with intestinal barrier dysfunction during the course of T cell transfer and DSS-induced colitis. *J Crohns Colitis*. 2020;14:974–94.
31. Kim S-W, Suda W, Kim S, Oshima K, Fukuda S, Ohno H, et al. Robustness of gut microbiota of healthy adults in response to probiotic intervention revealed by high-throughput pyrosequencing. *DNA Res*. 2013;20:241–53.
32. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13:581–3 Nature Publishing Group.
33. Wright ES. Using DECIPHER v2.0 to analyze big biological sequence data in R. *R J*. 2016;8:352–9.
34. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Priesse E, Quast C, et al. The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. *Nucleic Acids Res*. 2014;42:D643–8.
35. Lagkouvardos I, Fischer S, Kumar N, Clavel T. Rhea: a transparent and modular R pipeline for microbial profiling based on 16S rRNA gene amplicons. *PeerJ*. 2017;5:e2836 PeerJ Inc.
36. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLOS ONE*. 2013;8:e61217 Public Library of Science.
37. Clarke KR. Non-parametric multivariate analyses of changes in community structure. *Aust J Ecol*. 1993;18:117–43.
38. Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome*. 2014;2:15.
39. Nearing JT, Douglas GM, Hayes MG, MacDonald J, Desai DK, Allward N, et al. Microbiome differential abundance methods produce different results across 38 datasets. *Nat Commun*. 2022;13:342.
40. Chen T, Yu W-H, Izard J, Baranova OV, Lakshmanan A, Dewhirst FE. The human oral microbiome database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database*. 2010;2010:baq013.
41. Ritari J, Salojärvi J, Lahti L, de Vos WM. Improved taxonomic assignment of human intestinal 16S rRNA sequences by a dedicated reference database. *BMC Genomics*. 2015;16:1056.
42. Griffith DM, Veech JA, Marsh CJ. Cooccur: probabilistic species co-occurrence analysis in R. *J Stat Softw*. 2016;69:1–17.
43. Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, et al. PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol*. 2020;38:685–8 Nature Publishing Group.
44. Liu C, Smet A, Blaecher C, Flahou B, Ducatelle R, Linden S, et al. Gastric de novo Muc13 expression and spasmodic polypeptide-expressing metaplasia during *Helicobacter heilmannii* infection. *Infect Immun*. 2014;82:3227–39.
45. Reis CA, David L, Correa P, Carneiro F, de Bolós C, Garcia E, et al. Intestinal metaplasia of human stomach displays distinct patterns of mucin (MUC1, MUC2, MUC5AC, and MUC6) expression. *Cancer Res*. 1999;59:1003–7.
46. Kim J, Park C, Kim KH, Kim EH, Kim H, Woo JK, et al. Single-cell analysis of gastric pre-cancerous and cancer lesions reveals cell lineage diversity and intratumoral heterogeneity. *NPJ Precis Onc*. 2022;6:1–11 Nature Publishing Group.
47. Scott AJ, Alexander JL, Merrifield CA, Cunningham D, Jobin C, Brown R, et al. International Cancer Microbiome Consortium consensus statement on the role of the human microbiome in carcinogenesis. *Gut*. 2019;68:1624–32.
48. Sheng YH, He Y, Hasnain SZ, Wang R, Tong H, Clarke DT, et al. MUC13 protects colorectal cancer cells from death by activating the NF- $\kappa$ B pathway and is a potential therapeutic target. *Oncogene*. 2017;36:700–13 Nature Publishing Group.
49. Shi D, Xi X. Regulation of MUC6 methylation correlates with progression of gastric cancer. *Yonsei Med J*. 2021;62:1005–15.
50. Rico SD, Mahnken M, Büschel F, Dum D, Luebbe AM, Kluth M, et al. MUC5AC expression in various tumor types and nonneoplastic tissue: a tissue microarray study on 10 399 tissue samples. *Technol Cancer Res Treat*. 2021;20:153303382110433.
51. Kim SM, Kwon CH, Shin N, Park DY, Moon HJ, Kim GH, et al. Decreased Muc5AC expression is associated with poor prognosis in gastric cancer. *Int J Cancer*. 2014;134:14–24.
52. Dicksved J, Lindberg M, Rosenquist M, Enroth H, Jansson JK, Engstrand L. Molecular characterization of the stomach microbiota in patients with gastric cancer and in controls. *J Med Microbiol*. 2009;58:509–16.

53. Engstrand L, Lindberg M. *Helicobacter pylori* and the gastric microbiota. *Best Pract Res Clin Gastroenterol*. 2013;27:39–45.
54. Barra WF, Sarquis DP, Khayat AS, Khayat BCM, Demachki S, Anaissi AKM, et al. Gastric cancer microbiome. *Pathobiology*. 2021;88:156–69.
55. Niv Y. *Helicobacter pylori* and gastric mucin expression: a systematic review and meta-analysis. *World J Gastroenterol*. 2015;21:9430–6.
56. Ott SJ, Musfeldt M, Wenderoth DF, Hampe J, Brant O, Fölsch UR, et al. Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. *Gut*. 2004;53:685–93 BMJ Publishing Group.
57. Nardone G, Compare D, Rocco A. A microbiota-centric view of diseases of the upper gastrointestinal tract. *Lancet Gastroenterol Hepatol*. 2017;2:298–312.
58. Haro C, Rangel-Zúñiga OA, Alcalá-Díaz JF, Gómez-Delgado F, Pérez-Martínez P, Delgado-Lista J, et al. Intestinal microbiota is influenced by gender and body mass index. *PLoS ONE*. 2016;11:e0154090.
59. Leeming ER, Johnson AJ, Spector TD, Le Roy CI. Effect of diet on the gut microbiota: rethinking intervention duration. *Nutrients*. 2019;11:2862.
60. de la Cuesta-Zuluaga J, Kelley ST, Chen Y, Escobar JS, Mueller NT, Ley RE, et al. Age- and sex-dependent patterns of gut microbial diversity in human adults. *mSystems*. 2019;4:e00261-19.
61. Vich Vila A, Collij V, Sanna S, Sinha T, Imhann F, Bourgonje AR, et al. Impact of commonly used drugs on the composition and metabolic function of the gut microbiota. *Nat Commun*. 2020;11:362 Nature Publishing Group.
62. Sobhani I, Bergsten E, Couffin S, Amiot A, Nebbad B, Barau C, et al. Colorectal cancer-associated microbiota contributes to oncogenic epigenetic signatures. *Proc Natl Acad Sci U S A*. 2019;116:24285–95.
63. Dai D, Yang Y, Yu J, Dang T, Qin W, Teng L, et al. Interactions between gastric microbiota and metabolites in gastric cancer. *Cell Death Dis*. 2021;12:1104.
64. Wu S, Yu F, Ma L, Zhao Y, Zheng X, Li X, et al. Do maternal microbes shape newborn oral microbes? *Indian J Microbiol*. 2021;61:16–23.
65. Yang Y, Dai D, Jin W, Huang Y, Zhang Y, Chen Y, et al. Microbiota and metabolites alterations in proximal and distal gastric cancer patients. *J Transl Med*. 2022;20:439.
66. Li Z-P, Liu J-X, Lu L-L, Wang L-L, Xu L, Guo Z-H, et al. Overgrowth of *Lactobacillus* in gastric cancer. *World J Gastrointest Oncol*. 2021;13:1099–108.
67. Castaño-Rodríguez N, Goh K-L, Fock KM, Mitchell HM, Kaakoush NO. Dysbiosis of the microbiome in gastric carcinogenesis. *Sci Rep*. 2017;7:15957 Nature Publishing Group.
68. Yang Y, Long J, Wang C, Blot WJ, Pei Z, Shu X, et al. Prospective study of oral microbiome and gastric cancer risk among Asian, African American and European American populations. *Int J Cancer*. 2022;150:916–27.
69. Lertpiriyapong K, Whary MT, Muthupalani S, Lofgren JL, Gamazon ER, Feng Y, et al. Gastric colonisation with a restricted commensal microbiota replicates the promotion of neoplastic lesions by diverse intestinal microbiota in the *Helicobacter pylori* INS-GAS mouse model of gastric carcinogenesis. *Gut*. 2014;63:54–63.
70. Mo S, Ru H, Huang M, Cheng L, Mo X, Yan L. Oral-intestinal microbiota in colorectal cancer: inflammation and immunosuppression. *JIR Dove Press*. 2022;15:747–59.
71. Hsieh Y-Y, Tung S-Y, Pan H-Y, Yen C-W, Xu H-W, Lin Y-J, et al. Increased abundance of clostridium and fusobacterium in gastric microbiota of patients with gastric cancer in Taiwan. *Sci Rep*. 2018;8:158.
72. Hsieh Y-Y, Tung S-Y, Pan H-Y, Chang T-S, Wei K-L, Chen W-M, et al. Fusobacterium nucleatum colonization is associated with decreased survival of helicobacter pylori-positive gastric cancer patients. *World J Gastroenterol*. 2021;27:7311–23.
73. Wu J, Li Q, Fu X. Fusobacterium nucleatum contributes to the carcinogenesis of colorectal cancer by inducing inflammation and suppressing host immunity. *Transl Oncol*. 2019;12:846–51.
74. Tziatzios G, Gkolfakis P, Papanikolaou IS, Mathur R, Pimentel M, Giamarellos-Bourboulis EJ, et al. Gut microbiota dysbiosis in functional dyspepsia. *Microorganisms*. 2020;8:691 Multidisciplinary Digital Publishing Institute.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

