

81153\_Auto\_Edited-check.docx

**Name of Journal:** *World Journal of Gastroenterology*

**Manuscript NO:** 81153

**Manuscript Type:** ORIGINAL ARTICLE

*Case Control Study*

4

**Comparison of genomic and transcriptional microbiome analysis in gastric cancer patients and healthy individuals**

Nikitina D *et al.* GC microbiome DNA/RNA

## **Abstract**

### **BACKGROUND**

*Helicobacter pylori* and the stomach microbiome play a crucial role in gastric carcinogenesis, and detailed characterization of the microbiome is necessary for a better understanding of the pathophysiology of the disease. There are two common modalities for microbiome analysis: DNA (16S rRNA gene) and RNA (16S rRNA transcript) sequencing. The implications from the use of one or another sequencing approach on the characterization and comparability of the mucosal microbiome in gastric cancer (GC) are poorly studied.

### **AIM**

To characterize the microbiota of GC using 16S rRNA gene and its transcript and determine difference in the bacterial composition.

### **METHODS**

In this study, 316 DNA and RNA samples extracted from 105 individual stomach biopsies were included. The study cohort consisted of 29 healthy control individuals and 76 patients with GC. Gastric tissue biopsy samples were collected from damaged mucosa and healthy mucosa at least 5 cm from the tumor tissue. From the controls, healthy stomach mucosa biopsies were collected. From all biopsies RNA and DNA were extracted. RNA was reverse transcribed into cDNA. V1-V2 region of bacterial 16S rRNA gene from all samples were amplified and sequenced on an Illumina MiSeq platform. Bray-Curtis algorithm was used to construct sample-similarity matrices abundances of taxonomic ranks in each sample type. For significant differences between groups permutational multivariate analysis of variance and Mann-Whitney test followed by false-discovery rate test were used.

### **RESULTS**

Microbial analysis revealed that only a portion of phylotypes (18%-30%) overlapped between microbial profiles obtained from DNA and RNA samples. Detailed analysis revealed differences between GC and controls depending on the chosen modality, identifying 17 genera at the DNA level and 27 genera at the RNA level. Ten of those bacteria were found to be different from the control group at both levels. The key taxa showed congruent results in various tests used; however, differences in 7 bacteria taxa were found uniquely only at the DNA level, and 17 uniquely only at the RNA level. Furthermore, RNA sequencing was more sensitive for detecting differences in bacterial richness, as well as differences in the relative abundance of *Reyranella* and *Sediminibacterium* according to the type of GC. In each study group (control, tumor, and tumor adjacent) were found differences between DNA and RNA bacterial profiles.

## CONCLUSION

Comprehensive microbial study provides evidence for the effect of choice of sequencing modality on the microbiota profile, as well as on the identified differences between case and control.

**Key Words:** Gastric cancer; Microbiome; *Helicobacter pylori*; 16S rRNA gene; 16S rRNA transcript; 16S rDNA

Nikitina D, Lehr K, Vilchez-Vargas R, Jonaitis LV, Urba M, Kupcinskas J, Skieceviciene J, Link A. Comparison of genomic and transcriptional microbiome analysis in gastric cancer patients and healthy individuals. *World J Gastroenterol* 2022; In press

**Core Tip:** In this study, we aimed to characterize the microbiota of gastric cancer (GC) on two levels: 16S rRNA gene and its transcript. Our study showed that only a small portion of bacterial sequences overlapped using those two approaches. Moreover, our study revealed that obtained results comparing the case group with the controls depend on the chosen modality. We also showed that *Reyranella* and *Sediminibacterium* was

associated with the Lauren classification and RNA level was more sensitive to detect low abundant bacteria. This study provides novel insights into microbiome study as well as new founding related to complex GC pathogenesis.

## **INTRODUCTION**

Microbiota analyses are becoming increasingly relevant in scientific and clinical studies. Most modern microbiome studies use 16S rRNA gene sequence analysis at the DNA level, thereby enabling the identification of bacteria at all stages of their existence (active, dead, and inactive bacteria in the form of endospores) simultaneously. However, some of the more recent studies use RNA samples, which are subsequently reverse transcribed into cDNA for sequencing, giving us knowledge about the metabolic state of the microbial community<sup>[1]</sup>. RNA has a shorter half-life than DNA and turns over in cells more rapidly, providing a deeper look at bacterial activity<sup>[2]</sup>.

The stomach has long been considered an almost sterile organ due to its acidic environment and enzymatic effects<sup>[3]</sup>. Since its identification, it is known that *Helicobacter pylori* (*H. pylori*) is perfectly adapted not only to survive in the acidic environment of the stomach, but also to colonize this part of the gastrointestinal tract<sup>[4]</sup>. <sup>6</sup>*H. pylori* is the major cause of peptic ulcer disease and the most significant risk factor for <sup>3</sup>gastric cancer (GC). GC remains one of the most common cancers in the world and the fourth leading cause of cancer-related death<sup>[5]</sup>. However, only a minority of people infected with *H. pylori* develop GC, which may be linked to non-*H. pylori* microbiota-associated alterations in the stomach<sup>[6]</sup>. Studies in insulin-gastrin (INS-GAS) mice and in humans indicated the importance of other members of the stomach bacterial community in the development of gastric carcinogenesis<sup>[7-10]</sup>.

There is only one study that has compared DNA and RNA profiles of the stomach microbiota<sup>[11]</sup>. However, the profiles of the active and standing microbiota in GC have not been studied. In this study, we systematically characterized the microbiota of GC on both levels using the 16S rRNA gene (DNA level) and its transcript (RNA level). GC tumor and tumor adjacent tissue samples, as well as healthy mucosa samples from the

young control group, were used for the comparison. We obtained detailed data on bacterial composition within groups depending on study modality (DNA or RNA) and performed association analysis with clinical characteristics to question the potential impact of approach on the outcome.

## **MATERIALS AND METHODS**

### ***Study cohort***

In total, 316 DNA and RNA samples from a group of 105 individuals were included in the study (Figure 1). The study cohort consisted of 29 healthy control individuals and 76 patients with GC. Participants did not report any antibiotic intake at least a month before endoscopy. Gastric tissue biopsy samples from damaged mucosa and healthy mucosa at least 5 cm from the tumor tissue were collected from GC patients using single-bite biopsy forceps. From the controls, healthy stomach mucosa biopsies were collected. Tissue samples were placed in sterile cryotubes (Thermo Fisher Scientific, United States), snap-frozen in liquid nitrogen, and stored at -86 °C until further study. Clinical data obtained from histological examination, such as tumor size, number of lymph nodes damaged by tumor cells, presence of metastases (TNM classification), cell differentiation (grading), type of GC (Lauren classification) and stage of GC, were included in the analysis. An overview of the demographic and clinical characteristics of the study cohort is given in online Supplementary Table 1.

Study individuals were recruited at the Department of Gastroenterology at the Hospital of Lithuanian University of Health Sciences Kaunas Clinics during the years 2012-2018. This study was approved by the local ethics committee (BE-2-10), and all participants gave their written informed consent.

### ***DNA, RNA extraction, cDNA synthesis and amplicon library preparation***

Total DNA and RNA were extracted from gastric biopsy samples using an AllPrep DNA/RNA Mini kit (Qiagen, Germany) according to the manufacturer's recommendations. RNA was reverse transcribed into cDNA using the Superscript IV

First-Strand Synthesis System Purification Kit (Invitrogen, Carlsbad, CA) and random hexamer primers, following the manufacturer's instructions. Amplicon libraries were generated as described previously<sup>[12,13]</sup>. The bacterial 16S rRNA gene V1-V2 region was amplified using the 27F and 338R polymerase chain reaction primers and sequenced on a MiSeq (2 × 250 bp; Illumina, Hayward, CA).

### ***Bioinformatic and statistical analysis***

Bioinformatic processing was performed as described previously<sup>[14]</sup>. FastQ files were analyzed using the dada2 package<sup>[15]</sup>, version 1.10.1, in R. In total, 7735281 paired-end reads were received, with an average of 22953 per sample. Samples that did not reach 5000 reads were discarded from the analysis (21 samples out of initial 337). All samples were rarefied to an equal sequencing depth of 5047 reads using the phyloseq package<sup>[16]</sup>, with returning 10496 phylotypes (Supplementary Table 2). Phylotypes were annotated to a taxonomic affiliation based on the naive Bayesian classification<sup>[17]</sup> with a pseudobootstrap threshold of 80%. The relative abundances (expressed as percentages) of different microbial communities' phylogenetic ranks (from phylum to class, order, family, genus and phylotype) were used for downstream analyses.

The phylogenetic tree was built using the online tool iTOL<sup>[18]</sup>, after hierarchical clustering using the Bray-Curtis algorithm<sup>[19]</sup> at the phylotype level in Past 3<sup>[20]</sup>. Bacterial richness and Shannon diversity indices were calculated using the vegan<sup>[21]</sup> package from R. The data matrices comprising the percentage of abundances of each of the abovementioned taxa were used to construct sample-similarity matrices by the Bray-Curtis algorithm, where samples were ordinated by principal coordinate analysis (PCoA) at the phylotype level using the patchwork<sup>[22]</sup> package from R.

Differences in relative abundance of detected bacteria (at all taxonomic ranks) between study groups were evaluated by PERMANOVA and ANOSIM statistical tests, using 9999 permutations. Groups were considered significantly different if the *P*-value was < 0.05, considering an estimate effect-size *F* values for PERMANOVA and *R* values for ANOSIM tests. Calculation was made by Past 3 program. The distributions of taxa

abundance values were compared by Mann-Whitney test followed by Benjamini-Hochberg correction for multiple comparisons, named as false discovery rate value. Differences were considered significant when the corrected p value (q value) was  $< 0.05$ .

The bacterial networks were visualized using Cytoscape 3.8.0<sup>[23]</sup>, after the Spearman correlation test performed with the psych<sup>[24]</sup> package from R, with threshold of 0.2 in absolute value and  $P$ -value  $< 0.05$ . Phylotypes that accounted for at least 1% of the total number of phylotypes and at least 10% of the samples in each group were used for correlation analysis.

## **RESULTS**

### ***General cohort***

The bacterial contents of 180 biopsy samples taken from 105 individuals were characterized as described above (Supplementary Table 1). After sequencing and rarefying library size to the minimum sequencing depth, 10496 different phylotypes belonging to 23 phyla, 40 classes, 82 orders, 169 families, and 463 genera were retrieved and taxonomically annotated.

The global bacterial profiles were grouped into two clusters based on their Bray-Curtis similarities as percentages (Figure 2A). Analyzing all samples together, the main factor for clustering was bacterial heterogeneity. The first cluster consisted of samples with a more heterogeneous microbiome profile, where the most abundant bacteria accounted for less than 30%. All control samples (except T10\_2) were located in this cluster. The second cluster - where the most abundant bacteria accounted for more than 30% of GC patient samples - was shaped by the most abundant bacteria *Helicobacter*, the abundance of which reached 98%-100% in some samples (Figure 2A, Supplementary Table 2).

### ***Distinct profile of the gastric tissue microbiome at the RNA and DNA levels***

Further PERMANOVA and ANOSIM analyses showed that DNA and RNA groups of the same study individuals were significantly different in all taxonomic ranks (Figures



3A-C; Supplementary Table 3). Differences between DNA and RNA samples were noticeable even at the phylum level. Although *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, *Proteobacteria*, and *Fusobacteria* were the main bacterial phyla in all groups, *Bacteroides* and *Fusobacteria* were significantly more abundant at the DNA level. However, in the control group, *Firmicutes* and *Proteobacteria* were more abundant at the RNA level. Bacterial profile analysis indicated that only a portion of phylotypes (18%-30%) were common between bacterial profiles obtained from DNA and RNA samples (Figure 3D). PCoA supported the distinction of bacterial communities at the DNA and RNA levels, especially in the control group (Figure 3E). *Firmicutes*, *Bacteroidetes*, *Proteobacteria*, and *Fusobacteria* were major phyla determinants for sample differentiation.

More detailed analysis revealed that DNA and RNA samples differed from each other by 12, 10, and 30 phylotypes and by 18, 17, and 35 genera in the tumor, tumor adjacent, and control groups, respectively (Table 1, Supplementary Table 4). In all study groups, bacteria such as *Neisseria*, *Peptostreptococcus*, *Prevotella*, *Veillonella*, and *Oribacterium* were significantly more abundant at the DNA level, while *Staphylococcus*, *Methyloversatilis*, *Pseudomonas*, *Reyranella*, *Corynebacterium*, and *Sediminibacterium* were significantly enriched at the RNA level. Interestingly, most of these bacteria founded in the RNA samples were not observed in the DNA samples at all, or their relative abundance was low. Some changes in the relative abundance of bacteria between DNA and RNA samples were specific for the study group. For instance, in the control group, *Helicobacter*, *Gemella*, and *Streptococcus* were enriched at the RNA level, while *Actinomyces* and *Alloprevotella* were enriched at the DNA level. In the GC groups (tumor and tumor adjacent), *Fusobacterium*, *Granulicatella*, *Solobacterium*, and *Porphyromonas* were enriched at the DNA level. No bacteria were enriched at the RNA level in this group.

Nevertheless, despite the found differences between DNA and RNA, samples of the same origin tended to cluster together in each of the study groups (Figure 2). Paired samples, 46 pairs out of 64 (72%) in the tumor group and 38 pairs out of 58 (66%) in the tumor adjacent tissue group, clustered next to each other, indicating their global

similarity (Supplementary Figure 1). Paired samples from the control group were not added to this analysis due to the small number of paired samples.

### ***Revealed microbiome alterations in GC depend on the chosen sequencing modality***

The GC samples had lower bacterial richness and diversity compared to control samples (Figures 3F and 3G). While differences in diversity were found both at the DNA and RNA levels, differences in bacterial richness were found only at the RNA level. Group-average agglomerative hierarchical clustering analysis showed that it was possible to distinguish patients with GC from controls by their bacterial profile, as samples tended to cluster based on clinical status (both at the DNA and RNA levels) (Figures 2B and 2C). These results were supported by the phylogenetic analysis of global stomach bacteria, which revealed significant differences between the GC group and control groups at all taxonomic ranks (Figures 3 A-C, Supplementary Table 3).

Bacterial abundance differential analysis revealed 15 phylotypes and 17 genera that differed between the GC and control groups at the DNA level (Table 2, Supplementary Figure 2, Supplementary Table 4). Meanwhile, at the RNA level, there were twice as many differences: 40 at the phylotype level and 27 at the genus level. Half of the differences detected at the DNA level were also found at the RNA level (58% of genera and 46% of phylotypes). These bacteria include previously described bacteria, such as *Lactobacillus*, *Propionibacterium*, *Streptococcus*, and *Veillonella*, among others<sup>[25-29]</sup>.

Although fewer unique bacteria were identified only at the DNA level (8 phylotypes and 7 genera), they were more studied and more frequently discussed in the literature as being associated with various human health conditions (Supplementary Table 4). These bacteria include *Campylobacter*, *Clostridium sensu stricto*, *Prevotella*, and *Saccharibacteria*, among others. Of the listed bacteria, *Clostridium sensu stricto* was enriched, and others were decreased in GC patients. Uniquely, only at the RNA level were 33 and 17 differences at the phylotype and genus levels, respectively, found between the GC and control groups (Supplementary Table 4). Essentially, this group included such bacteria that were not established or their abundance at the DNA level

was negligible, for example, *Limnohabitans*, *Methylobacterium*, *Methyloversatilis*, *Pseudomonas*, *Reyranella*, *Rhodoluna*, *Sediminibacterium*, and *Staphylococcus*. Of the listed bacteria, only *Pseudomonas* and *Staphylococcus* were more abundant in GC samples, while all the others were more abundant in healthy individuals.

Bacterial diversity and profile comparison analysis between tumor and tumor adjacent tissues did not reveal significant differences at either the DNA or RNA level (Figures 3A-C, 3F and 3G). Moreover, assemblages of approach from each individual typically clustered together irrespective of tissue type (tumor or tumor adjacent tissue) (Figure 2A).

#### ***Bacterial networks in GC patients have fewer components and integrated connections***

Analysis of the bacterial network similarity revealed that the main network holding bacteria with the highest betweenness centrality score was different between DNA and RNA levels in all study groups (Supplementary Figure 3). In the bacterial network of the control group at the DNA level, phylotypes depending on the *Streptococcus*, *Prevotella*, and *Actinomyces* genera accounted for 68% (58 out of 85) of the total number of bacteria and formed the core network keeping bacteria, while at the RNA level, core bacteria were *Streptococcus* and *Gemella*, making up to 62% (47 out of 75) (Supplementary Figures 3A and 3B). The GC groups showed different DNA/RNA networks as well: The main network forming bacteria in the tumor adjacent tissue at the DNA level was *Prevotella*, *Gemella*, and *Granulicatella*, while the RNA network was shaped by *Streptococcus*, *Reyranella*, and *Fusobacterium* (Supplementary Figures 3C and 3D). The most critical network-forming bacteria in tumor tissue were: *Granulicatella*, *Veillonella*, and *Neisseria* at the DNA level and *Reyranella*, *Acinetobacter*, and *Prevotellaceae* at the RNA level (Supplementary Figures 3E and 3F).

Two common bacterial clusters (one at the DNA level and another at the RNA level) with strong positive correlations for tumor and tumor adjacent tissues were discovered (Supplementary Figures 3C-F), which confirms the absence of significant differences between tumor and tumor adjacent tissue microbiome profiles. At the DNA level, the

common cluster consisted of Phy6 (*Neisseria*), Phy15 (unclassified *Prevotellaceae*), Phy23 (*Neisseria perflava*), Phy29 (*Prevotella melaninogenica*), Phy87 (*Solobacterium*), and Phy98 (*Prevotella*). The common cluster at the RNA level included phylotypes such as Phy7 (*Reyranella*), Phy33 (*Sediminibacterium*), Phy46 (*Propionibacterium acnes*), Phy94 (*Methyloversatilis*), Phy107 (*Pseudomonas aeruginosa*), and Phy108 (*Sphingomonas echinoides*). Detected clusters were not found in the control group.

Generally, under the same analysis conditions, GC patients displayed a simpler bacterial network at both the DNA and RNA levels. At the DNA level, control, the tumor, and tumor adjacent groups had 85, 25, and 23 bacteria, respectively; at the RNA level, they had 75, 21, and 18, respectively. Moreover, analysis of bacterial interactions in controls had not only positive but also negative correlations, while GC analysis showed mostly positive correlations.

#### ***GC microbiota alterations and clinical parameters***

At the DNA level, according to clinical parameters, statistically significant differences were found only in the decrease in bacterial richness between smaller tumors (T1-T2) and extended tumors (T4). The RNA level turned out to be more sensitive and allowed us to detect richness differences between grade II and grade III (Figures 4A and 4B). Moreover, at the RNA level, the relative abundance of the Phy7 (*Reyranella*) and Phy33 (*Sediminibacterium*) phylotypes was lower in the diffuse type of GC than in the intestinal type (Figures 4C and 4D). No differences were found between subgroups at the DNA level.

#### ***The effect of *H. pylori* infection on stomach microbiota***

PCoA showed *Helicobacter* <sup>1</sup> to be the major determinant for differentiating samples based on their bacterial composition in the stomach (Supplementary Figure 4). Overall, *H. pylori* was detected in 115 and 117 DNA and RNA samples, respectively. In the control group, *H. pylori* was lower than that in the GC groups at both the DNA and RNA levels (Figure 5A). The tumor adjacent sample group showed the highest number of samples

with high *H. pylori* abundance (Figure 5B). Both in tumor and tumor adjacent groups the mean abundance of *H. pylori* was increased at the RNA level, although no significant differences between DNA and RNA samples were found.

High *H. pylori* relative abundance (> 15%) led to an increase in the relative abundance of *Proteobacteria* and a decrease in other major bacterial phyla, such as *Firmicutes*, *Bacteroidetes*, and *Fusobacteria* (Figure 5C). In tumor tissues analyzed at the DNA and RNA level, *Helicobacter* was only one genus which changed significantly between samples with high and low *H. pylori* abundance (Figure 5D, Supplementary Table 5). On the other hand, in tumor adjacent tissues, more bacteria were found, the number of which changed together with *Helicobacter*. At the DNA level, as the relative abundance of *H. pylori* increased, the abundance of *Porphyromonas* and *Prevotella* significantly decreased. In line with our previous results, more significant differences were found at the RNA level: *Staphylococcus* significantly increased and seven bacteria (*Campylobacter*, *Fusobacterium*, *Prevotella*, *Pseudomonas*, *Reyranella*, *Sediminibacterium*, *Streptococcus*) decreased (Figure 5C, Supplementary Table 5). *Porphyromonas* tended to decrease in tumor adjacent RNA samples as well, although it did not reach a statistically significant level (Supplementary Table 5).

## **DISCUSSION**

Despite growing interest in the study of microbiota, there is still limited agreement on the most appropriate standard for such studies, especially using 16S rRNA sequencing. Here, we performed systematic analysis of bacterial communities at both the 16S rRNA gene and 16S rRNA transcript levels. To estimate the impact of the different approaches, we used the GC model and considered not only healthy gastric tissues but also GC tumor and adjacent tissues.

The analysis of the study results showed that there were significant differences in the relative abundance of the gastric tissue microbiome between 16S rRNA gene transcript and 16S rRNA gene levels in all study groups (control, tumor, and tumor adjacent). This

is the first GC study indicating that active and standing gastric microbiomes are distinct even at the largest taxonomic levels.

Differences in bacterial communities at the DNA and RNA levels could be explained by several possibilities. Using DNA as a research material summed up all bacteria, both biologically active passive in the form of endospores, and DNA sequences of already destroyed and dead bacteria<sup>[2,30]</sup>. The presence and number of ribosomes in bacteria reflects their metabolic activity; thus, the analysis at the RNA level shows the metabolic activity of live and active bacteria in the community<sup>[31-33]</sup>. For instance, previously GC-associated bacteria such as *Prevotella*, *Veillonella*, and *Neisseria* in our study were present in high abundance in all analyzed groups at the DNA level but were greatly reduced at the RNA level. In contrast, *Pseudomonas*, *Reyranella*, and *Staphylococcus* were present in higher abundance at the RNA level in all groups. However, it is erroneous to assume that only active bacteria can influence host responses. Many studies have shown that inactivated bacteria or parts of their cells can also influence inflammatory processes or other responses in host tissues. For example, Rabie *et al*<sup>[34]</sup> showed that thermally inactivated *Salmonella*, *Staphylococcus*, *Escherichia*, and *Pseudomonas* strains with unchangeable surface proteins cause colon and breast cancer cell proliferation. In Suprewicz *et al*<sup>[35]</sup>'s study, heat-inactivated *Enterococcus faecalis*, *Actinomyces odontolyticus*, and *Propionibacterium acnes* caused cell proliferation changes in lung, breast, and ovarian carcinoma. Postbiotics work based on the same principle. To avoid possible bacterial infection during therapy, instead of active bacteria, their metabolites, which are involved in anti-inflammatory and anticancer mechanisms, are used<sup>[36]</sup>.

It is also cannot be excluded that the shift in bacterial abundance between DNA and RNA levels might stem from varying numbers of copies of the 16S rRNA gene<sup>[37]</sup> or target sequence quantity inequality<sup>[38]</sup>. Bacterial rRNAs (16S rRNA, 23S rRNA and 5S rRNA) are typically organized into one operon, and their transcription occurs together, with the number of such operons varying from 1 to 15<sup>[38]</sup>. In the case of active bacteria, an increase in 16S rRNA gene copies proportionally increases the pool of 16S rRNA

transcripts. However, in the case of inactive bacteria, a larger number of 16S gene copies enables the detection of some bacteria, which could not be detected at the RNA level.

The amount of target sequences using the 16S rRNA gene and its transcript are not the same. Of all types of RNA molecules present in the cell, the most common (80%-90%) are included in the ribosome structure rRNAs<sup>[39]</sup>. 16S rRNAs make up one-third of the total rRNAs. On the other hand, when analyzing the microbiota using the 16S rRNA gene, only one gene is amplified out of the total number of genes, which in different bacteria varies from 1500 to 7000<sup>[37]</sup>. Thus, the initial larger amount of the bacterial target sequence at the RNA level makes it possible to increase the depth of sequencing and detect more rare bacteria that would be lost during DNA-level analysis. In addition, a shift toward DNA or RNA levels can also be caused by ingestion of bacterial parts from the higher parts of the digestive tract.

Our analysis revealed that the profile of differences found between GC and control tissue depended on the chosen modality: At the DNA level, 17 bacterial genera were detected, and at the RNA level, 27 bacterial genera were detected. Ten of those bacteria (*Actinomyces*, *Alloprevotella*, *Atopobium*, *Granulicatella*, *Lactobacillus*, *Megasphaera*, *Propionibacterium*, *Rothia*, *Streptococcus*, *Veillonella*) were found to be different from the control group at both levels of sequencing; seven bacterial taxa (*Campylobacter*, *Clostridium sensu stricto*, *Leptotrichia*, *Oribacterium*, *Prevotella*, *Saccharibacteria* genera *incertae sedis*, *Stomatobaculum*) were found uniquely only at the DNA level; and 17 (*Anaerococcus*, *Corynebacterium*, *Eubacterium*, *Flavobacterium*, *Gemella*, *Legionella*, *Limnochabitans*, *Massilia*, *Methylobacterium*, *Methyloversatilis*, *Parvimonas*, *Pseudomonas*, *Reyranella*, *Rhodoluna*, *Sediminibacterium*, *Solobacterium*, *Staphylococcus*) were found uniquely only at the RNA level. These results confirm the importance of unifying the procedures for studying the microbiota.

Although our study focused on differences in methodology, it did reveal several important findings for the GC study as well. Fourteen bacteria genera were identified to be decreased in patients with GC. Eleven of these bacteria (*Actinomyces*, *Atopobium*, *Propionibacterium*, *Streptococcus*, *Granulicatella*, *Veillonella*, *Rothia*, *Parvimonas*, *Gemella*,



*Prevotella*, *Leptotrichia*) were previously established in the stomach of healthy people in the absence of gastrointestinal diseases<sup>[40]</sup>. Most of them are common members of the upper gastrointestinal tract and have strong enzymatic activities. Our study also found four bacteria genera, which were significantly increased in GC patients' stomach biopsy: *Lactobacillus*, *Clostridium sensu stricto*, *Staphylococcus*, and *Pseudomonas*.

*Lactobacillus* is commonly used as a probiotic; however, it has been verified in multiple studies to be enriched in GC<sup>[41]</sup>. *Lactobacillus* strains, as well as *Clostridium* and *Staphylococcus*, can reduce nitrate to nitrite<sup>[42,43]</sup>. During the nitrate-reducing process, many N-nitroso compounds are formed that inhibit cell apoptosis and promote mutagenesis and protooncogene expression<sup>[44-47]</sup>. *Clostridium* is part of the normal gastrointestinal tract; however, in several previous studies, as in ours, an increase in the number of *Clostridium sensu stricto* was found<sup>[48-50]</sup>. Interestingly, Lertpiriyapong *et al*<sup>[9]</sup> showed earlier onset and faster progression of GC in INS-GAS mice with restricted microbiota (including *Clostridium*, *Lactobacillus*, and *Bacteroides*), highlighting a possible role of these bacteria in GC. Additionally, several studies have detected increased levels of *Staphylococcus* in patients with upper gastrointestinal diseases<sup>[51-53]</sup>. One of the reasons for this may be that strains of *Staphylococcus* have the enzyme urease, and are able to catalyze the hydrolysis of urea to carbon dioxide and ammonia, which can neutralize gastric hydrochloric acid, thus promoting bacterial existence. Although we found an increased number of *Pseudomonas*, this bacterial infection affects people with weakened immune systems (including patients with cancer), and thus, it is more likely that this finding is the result of already developed pathological processes.

We did not detect significant bacterial abundance, richness, or diversity alterations at either the DNA or RNA level between tumor and tumor adjacent tissues. This result is consistent with two previous studies<sup>[54,55]</sup> but contradicts recent GC studies where significant differences between tumor-affected and nearby healthy tissues were found<sup>[27,56]</sup>. Moreover, we found the same clusters of bacterial networks in tumor and tumor adjacent tissues at both the DNA and RNA levels. These results may suggest that with the onset and development of carcinogenic processes, local changes in stomach



tissues lead not only to a change in the bacterial composition but are also precise uniformity between cancer-affected and still healthy tissues (at least within a radius of 5 cm from the tumor area).

Studying GC samples at the RNA level, we managed to identify microbiome associations with clinical data. Analysis revealed two phylotypes (Phy7 and Phy33) related to *Reyranella* and *Sediminibacterium*, respectively. The relative number of those phylotypes gradually decreased from healthy to GC patients through intestinal growth type (considered as less aggressive cell growing type) to GC patients with diffuse growth type of cancer cells with worse outcome prognosis. To our knowledge, this is the first mention of these bacteria associated with the GC cell growth type. *Reyranella* is part of *Proteobacteria* and has previously been associated with the main chemokine expression, which is involved in T-cell attraction during cancerogenesis<sup>[57,58]</sup>. In another study, it was shown that there are significantly lower amounts of circulating natural killer and Treg cells in patients with diffuse/mixed-type GC compared to intestinal-type GC<sup>[59]</sup>. Taken together, these results suggest that *Reyranella* may be involved in the decrease in T-cell number and thus stimulation of cell growth of diffuse-type GC. *Sediminibacterium* was reported to be associated with GC, but there is no knowledge about the possible role of this bacteria in the pathophysiological processes<sup>[60,61]</sup>. Therefore, more detailed research on the effects of *Reyranella* and *Sediminibacterium* on GC cells is needed to be able to use these bacterial phylotypes as potential biomarkers.

*H. pylori* is the most common bacterial infection worldwide, as well as the main risk factor for GC<sup>[40]</sup>. It has been shown that during the transition from *H. pylori*-induced inflammation to the growth and development of carcinogenic cells, *H. pylori* is no longer detected in the affected areas in such large abundance<sup>[62]</sup>. Our results, showing that more *H. pylori* were found in tumor adjacent tissue than in tumor tissue, both at the DNA and RNA levels, confirm this. According to some previous reports, infection with *H. pylori* promotes the proliferation of non-*Helicobacter* bacteria from *Proteobacteria*, *Spirochetes*, and *Acidobacteria* and limits the spread of bacteria such as *Actinobacteria*, *Bacteroidetes* and *Firmicutes*<sup>[63,64]</sup>. Although most of the bacteria we found with altered

numbers in GC were not associated with *H. pylori*, changes in the number of bacteria, such as *Granulicatella*, *Lactobacillus*, *Rothia*, *Pseudomonas*, *Gemella*, *Prevotella*, *Leptotrichia*, *Clostridium sensu stricto*, and *Fusobacterium*, were associated with high *H. pylori* abundance.

The question regarding the causality in the gastric microbiome is still partially unanswered. On the one hand, alterations in gastric microbiota have a causal role in the progression of carcinogenesis (e.g., *H. pylori*). On the other hand, the role of other bacteria is less understood. However, there are new studies that strongly suggest the impact of the gastric microbiome on inflammation and carcinogenesis. For instance, a recent study by Kwon *et al*<sup>[65]</sup> showed that intestinal metaplasia or GC patient gastric microbiome transplantation contributes to changes in the phenotype of premalignant lesions. In this regard, a detailed understanding of the output of different sequencing technologies and comparability between RNA/DNA-based analyses is critical.

Since systematic analysis to assess the differences with respect to GC has not been performed before, we would like to point out some limitations of this work. While the primary focus of the work was related to technical differences, thus food preferences, sex, and aging can be potential contributing factors that have not been thoroughly considered in this study. Overall, the focus was on providing a truly confirmed healthy cohort for the most precise comparison to strengthen the differences. Nevertheless, PERMANOVA and Mann-Whitney analyses performed in each of the study groups (tumor DNA, tumor RNA, tumor adjacent DNA, tumor adjacent RNA, control DNA, control RNA) did not reveal significant differences between the sexes and age (divided by median) (Supplementary Tables 6 and 7). Furthermore, due to the sample size, we did not consider the newly proposed TCGA classification for subsequent analysis nor to assess the impact of 4 subtypes on bacterial composition in tumors.

## **CONCLUSION**

In conclusion, our study provides evidence that the tumor microbiome of GC patients has a distinct pattern compared to healthy controls, while the difference analyzed from

adjacent tissue was rather low. Despite some overlap between the data obtained from the 16S rRNA transcript and 16S rRNA gene, our results showed the critical importance of the chosen study material on the resulting bacterial profile. Thus, researchers comparing their results with previous studies might take into consideration which initial material was used, either the 16S rRNA gene or 16S rRNA transcript. Our results showed that the RNA level was more sensitive for detecting low abundance bacteria and allowed us to detect differences according to GC clinical data.

## **ARTICLE HIGHLIGHTS**

### ***Research background***

There is currently no gold standard for analyzing the microbiome in 16S rRNA studies. Two common modalities are: Sequencing of DNA (16S rRNA gene) and sequencing of RNA (16S rRNA transcript). <sup>5</sup> Gastric cancer (GC) remains one of the most common cancers in the world and microbiome takes important place in its carcinogenesis.

### ***Research motivation***

Microbiota studies are becoming more relevant and widespread. Comparison of different approaches for microbiome studying is necessary for correct interpretation of other studies results, as well as for a deeper understanding of bacterial composition.

### ***Research objectives***

To investigate how the choice of sequencing modality affects the bacterial profile of differences between case and controls as well as to characterize the microbiota of GC tissues using 16S rRNA gene and its transcript.

### ***Research methods***

The study included healthy tissues from the control group, as well as tumor and tumor adjacent tissues from GC patients. From all biopsies RNA and DNA were extracted. 16S rRNA V1-V2 region was sequenced for all samples. <sup>2</sup> For significant differences between

groups permutational multivariate analysis of variance and Mann-Whitney test followed by false-discovery rate test were used.

### ***Research results***

Only a small portion of bacterial sequences overlapped on DNA and RNA levels in all groups. Differences between GC and control groups also only partially overlapped on DNA and RNA levels. RNA sequencing was more sensitive for detecting differences in bacterial richness, low abundance bacteria, and changes in the relative abundance of *Reyranella* and *Sediminibacterium* according to the type of GC. In each study group differences between DNA and RNA bacterial profiles were identified.

### ***Research conclusions***

Chosen study material (16S rRNA transcript or 16S rRNA gene) greatly affects detectable microbiome profile as well as the differences between cases and controls.

### ***Research perspectives***

This study provides microbiome analysis applying two different methodologies using GC gastric tissues as example and could serve as a reference for future research.

## ORIGINALITY REPORT

# 3%

SIMILARITY INDEX

## PRIMARY SOURCES

- 1

Ramiro Vilchez-Vargas, Jurgita Skieceviciene, Konrad Lehr, Greta Varkalaite et al. "Gut microbial similarity in twins is driven by shared environment and aging", eBioMedicine, 2022  
Crossref

57 words — 1%
- 2

[www.nxtbook.com](http://www.nxtbook.com)  
Internet

36 words — 1%
- 3

[www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)  
Internet

19 words — < 1%
- 4

[figshare.com](http://figshare.com)  
Internet

14 words — < 1%
- 5

Jun Wei, Yimin Wang, Bo Xie, Jiachi Ma, Yaguo Wang. "Cortactin and HER2 as potential markers for dural-targeted therapy in advanced gastric cancer", Clinical and Experimental Medicine, 2021  
Crossref

13 words — < 1%
- 6

[eprints.nottingham.ac.uk](http://eprints.nottingham.ac.uk)  
Internet

12 words — < 1%

EXCLUDE QUOTES ON

EXCLUDE BIBLIOGRAPHY ON

EXCLUDE SOURCES < 12 WORDS

EXCLUDE MATCHES < 12 WORDS

