

Sequencing analysis of *Helicobacter pylori* infection in gastric mucosa and its progression to gastric mucosa-associated tissue (MALT) lymphoma

Authors

Seung-Joo Yoo^{1,2#}, Youngil Koh^{1,2,3,4#}, Ja Min Byun^{1,3}, Junshik Hong^{1,2,3,4}, Dong-Yeop Shin^{1,2,3,4}, Jaeyoung Chun^{5,6*}, Hyunsoo Chung^{7,8*}, and Sung-Soo Yoon^{1,2,3,4*},

#Co-first authors

*Co-corresponding authors

Affiliations

¹Department of Internal Medicine, Seoul National University Hospital, Seoul, Korea

²Cancer Research Institute, Seoul National University College of Medicine, Seoul, Korea

³Biomedical Research Institute, Seoul National University Hospital, Seoul, Korea

⁴Cancer for Medical Innovation, Seoul National University Hospital, Seoul, Korea

⁵Department of Internal Medicine, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Korea

⁶Division of Gastroenterology, Department of Internal Medicine, Yonsei University College of Medicine, Seoul, Korea

⁷Department of Medical Device Development, Seoul National University College of Medicine, Seoul, Korea

⁸Department of Internal Medicine and Liver Research Institute, Seoul National University Hospital, Seoul, Korea

Corresponding authors

Jaeyoung Chun, MD, PhD

Division of Gastroenterology, Department of Internal Medicine, Yonsei University College of Medicine

20, Eonju-ro 63-gil, Gangnam-gu, Seoul, Republic of Korea, Seoul 06229, Korea

Tel. +82-2-2019-3310

Fax +82-2-3463-3882

E-mail: chunjmd@yuhs.ac

Hyunsoo Chung, MD, PhD

Department of Internal Medicine, Seoul National University College of Medicine

101, Daehak-ro, Jongno-gu, Seoul 03080, Republic of Korea

Tel +82-2-2072-4844

Fax +82-2-762-9662

E-mail: h.chung@snu.ac.kr

Sung-Soo Yoon, MD, PhD

Department of Internal Medicine, Seoul National University Hospital

101, Daehak-ro, Jongno-gu, Seoul 03080, Republic of Korea

Tel +82-2-2072-3079

E-mail ssysmc@snu.ac.kr

Abstract

Introduction: The pathogenesis of gastric mucosa-associated tissue (MALT) lymphoma is associated with *Helicobacter pylori* infection. Although treatment strategies and responsiveness according to the stage of gastric MALT lymphoma have been widely reported, a detailed study of the biological carcinogenic process is still required.

Method: Paired, fresh tumor-adjacent normal and gastric mucosal tissue samples from 13 patients with gastric MALT lymphoma were prospectively collected. Whole exome sequencing (WES) and whole-transcriptome sequencing (WTS) data were generated. The analysis of mutations, gene fusion, gene expression, and the microbiome was stratified by *H. pylori* infection and disease status.

Results: Somatic mutations in *TRAF3* and *TNFAIP3* were identified in *H. pylori*-negative gastric MALT lymphoma. Fusions involving *BIRC3-MALT1* (2 samples) and *TBL1XR1-TP63* (1 sample with *H. pylori* infection) were detected. Stepwise comparative analysis of RNA expression revealed upregulation of immune response, inflammatory responses, and the NF- κ B signaling pathway in *H. pylori*-positive MALT lymphoma cases. Pathways associated with pathogens were upregulated in *H. pylori*-negative MALT lymphoma cases, suggesting that infections other than *H. pylori* may affect lymphomagenesis. Microbiome analysis revealed that *genus_Rothia* was negatively correlated with alpha diversity.

Conclusion: A stepwise approach using diverse stages of WTS data revealed detailed pathogenic mechanisms of gastric MALT lymphoma. Chronic inflammation following infection contributes to gastric MALT lymphomagenesis in both *H. pylori* positive and negative cases.

Keywords: *Helicobacter pylori*, MALToma, sequencing

Background

Gastric mucosa-associated lymphoid tissue (MALT) lymphoma is a low-grade non-Hodgkin lymphoma that originates from the post-germinal center B cells[1]. Gastric MALT lymphoma is the most common primary gastric lymphoma[2]. *Helicobacter pylori* (*H. pylori*) is associated with gastric MALT lymphoma; approximately 90% of patients with gastric MALT lymphoma are infected with *H. pylori* and 60–92% of cases respond to eradication therapy[3, 4].

In patients infected with *H. pylori*, the stomach is exposed to a low-pH environment and chronic stimulation, leading to the development of chronic gastritis and the acquisition of MALT[1, 5]. The development of MALT lymphoma is induced by genetic alterations and/or proliferation of monoclonal B cells following stimulation of *H. pylori*[5]. Translocations of t(1;2)(p22;p12)/*IGK-BCL10* and genetic alterations in *TNFAIP3*, *CARD11*, *among other things*, that indicate an abnormal activation of the NF-κB signaling pathway, play a fundamental role in driving gastric MALT lymphoma in both *H. pylori*-positive and *H. pylori*-negative cases[6].

In the early stages, MALT lymphoma depends on *H. pylori*, thus it is expected to regress with eradication therapy. However, in the advanced stages of the disease, MALT lymphoma transforms into an *H. pylori*-independent cancer, exhibiting a poor response to eradication therapy[7, 8]. The treatment strategies and responsiveness according to the stage of *H. pylori*-positive gastric MALT lymphoma have been reported[9-11]. However, detailed biological studies of this carcinogenic process are required. Although some studies have reported the results of microarray analysis of gastric MALT lymphoma, RNA-seq analysis using diverse stages of gastric mucosa samples would be helpful in elucidating the carcinogenesis of gastric MALT lymphoma by *H. pylori*[12-14].

In addition, in cases of *H. pylori*-negative gastric MALT lymphoma, infection with other microbiota, autoimmune diseases, and genetic alterations are believed to contribute to

carcinogenesis[15]. The most common genetic alteration is the translocation of t(11;18)(q21;q21)/*BIRC3-MALT1*, which is also associated with poor response to eradication therapy in *H. pylori*-negative cases[8, 16]. Interestingly, some patients without translocations respond to eradication even though they were not infected with *H. pylori*[17]. Considering the pathophysiology of *H. pylori*-positive cases and their phenotypic similarity to *H. pylori*-negative cases, a microbiome other than *H. pylori* may contribute to the development of *H. pylori*-negative gastric MALT lymphoma. Similarly, the differences in the microbiome between *H. pylori*-negative gastric MALT lymphoma and control patients have also been reported[18]. In summary, these findings highlight the importance of analyzing the microbiota of *H. pylori*-negative gastric MALT lymphomas.

Therefore, the current study aimed to elucidate the pathways that contribute to the carcinogenesis of MALT lymphoma following microbial infection. Additionally, we aimed to identify the microbiome responsible for initiating carcinogenesis in *H. pylori*-negative cases. For this purpose, the genetic changes and RNA expression patterns of gastric MALT lymphoma, adjacent normal tissue, and follow-up gastric mucosal tissue were compared by prospectively collecting fresh samples from patients with newly diagnosed gastric MALT lymphoma.

Methods

Patients and samples

A total of thirteen patients diagnosed with gastric MALT lymphoma were included in this study. Paired fresh samples of tumors and adjacent normal gastric mucosal tissues were collected using endoscopy before treatment. Additionally, follow-up gastric mucosal tissue and saliva samples were collected. To identify whether patients were infected with *H. pylori*, we conducted an Immunoglobulin G test, *Campylobacter*-like organism test, Urea Breath Test, and pathological tests on patients and samples. This study was approved by the IRB (Institutional

Review Board) of Seoul National University Hospital (No. H-1704-070-845).

Sequencing data generation

The DNA collected from the sample was used for whole exome sequencing (WES) and whole transcriptome sequencing (WTS). Library construction for WES data generation was performed using the SureSelect XT Human All Exon Kit V5 (Agilent, CA, USA), followed by sequencing on an Illumina HiSeq2500 platform with a 2×100 read length. To generate the WTS data, libraries were prepared using the TruSeq Stranded Total RNA Kit with Ribo-Zero Globin (Illumina, San Diego, CA). Paired-end sequencing of these libraries was conducted using the Illumina HiSeq2500 (Illumina, San Diego, CA) in accordance with the manufacturer's instructions. Of the 13 cases, only 10 sets of WES and WTS data were generated; three patients provided only WTS data.

Somatic mutation calling & filtering

To identify somatic mutations in cancer samples, whole-exome sequencing of cancer and saliva DNA samples was used. The quality of the sequenced reads was initially assessed using Fastqc (v.0.11.9)[19]. Low-quality reads were removed using Trimmomatic (version 1.39)[20]. The filtered sequencing reads were aligned to the human reference genome (GRCh37/hg19) using the Burrows–Wheeler Aligner[21]. Following alignment, somatic SNVs (Single Nucleotide Variations)/indels were identified using the Genome Analysis Tool Kit (GATK, v.4.2.0.0; <https://software.broadinstitute.org/gatk/>)[22]. The somatic mutations were identified using Mutect2 of GATK and the biological information was annotated using AnnoVar and VEP[23, 24]. Using these annotation tools, information from the 1000 Genomes Project, gnomAD, ClinVar, and Catalogue of Somatic Mutations in Cancers (COSMIC, version 96)[25–

28] was added. To filter out sequencing artifacts, variants with a depth lower than 50 and variants with allelic fractions between 7% and 40% were excluded. Additionally, variants with strand bias between 0.25 and 4 were included. We removed variants reported in origins other than hematopoietic lymphoid tissue and stomach. In the gnomAD, variants with an allele frequency of 0.1 or more in the East Asian region were excluded.

Subgroups of samples

The samples were divided into subgroups based on the identification of *the presence of H. pylori*. As mentioned previously, multiple tests were conducted on each sample. To judge the presence of *H. pylori*, an IgG test was prioritized, and other tests supported this decision.

Differentially expressed genes & gene set enrichment analysis

The RNA expression levels were estimated using whole-transcriptome sequencing data. The sequenced reads were aligned to the human reference genome (GRCh37/hg19) using STAR (Spliced Transcript Alignment in Reference v.3.6)[29]. RNA expression normalization and read count calculations were performed using RSEM (RNA-seq by Expectation Maximization, v.1.3.2)[30]. Using the estimated RNA expression levels, differentially expressed genes (DEG) were analyzed across samples in multiple comparison groups and gene set enrichment analysis (GSEA) was performed using R software (version 4. 2. 2) and the R package, DESeq2 (v.1.38.3)[31, 32].

Detection of genetic translocation

The Genetic translocations in the tumor samples were identified using three detection tools:

Fusioncatcher, STAR-Fusion, and Arriba[33-35]. Raw FASTQ files were used as input files for Fusioncatcher and STAR-Fusion, whereas Arriba was fed with BAM files generated using STAR-Aligner processing[29]. To remove possible false-positive variants, only variants detected by all three tools were retained. Furthermore, fusions with limited supporting evidence reads were filtered out and assessed as false positives using an integrative genomics viewer[36].

Detection & analysis of the mucosal microbiome

To identify the microbiome from the WTS data, a microbiome detection protocol was followed using Kraken software[37, 38]. Before classifying the microbiome, it was isolated from the transcriptome data by excluding reads mapped to the human reference genome (GRCh37/hg19). Subsequently, the reads were classified by mapping them to a reference sequence of the microbiota. During the process, the minimum hit groups were set to three, as recommended in the protocol, and a confidence score of 0.10 to filter out the reads with insufficient supporting evidence[37].

Results

The clinical characteristics of the 13 patients are shown in Table 1. Based on *H. pylori* status, location, and timing of the samples, the samples were categorized into six subgroups. As a result, 37 WTS data points from 13 patients with gastric MALT lymphoma were classified into *H. pylori*-positive cancer, *H. pylori*-positive normal, *H. pylori*-negative cancer, *H. pylori*-negative normal, *H. pylori*-positive follow-up, and *H. pylori*-negative follow-up groups (Table 1).

Genetic translocation

To identify recurrent or novel gene fusions in gastric MALT lymphoma, genetic translocations from the WTS data were detected using Fusioncatcher, STAR-Fusion, and Arriba. We identified two genetic translocations: t(11;18)(q21;q21)/*BIRC3-MALT1* and inv(3)(q26q28)/*TBLIXR1-TP63* (Figure 1B). *BIRC3-MALT1* gene fusions were identified in *H. pylori*-negative samples, consistent with previous reports on recurrent cases of *H. pylori*-negative gastric MALT lymphoma patients[39]. Additionally, an inversion between *TBLIXR1-TP63* was found in *H. pylori*-positive samples. The *TBLIXR1-TP63* gene fusion has been reported in a study of transcriptome data of diffuse large B-cell lymphoma cases[40].

Somatic mutation

From the 10 WES data, somatic mutations were identified and false-positive mutations were filtered out. Twenty gene mutations were identified in five samples (two *H. pylori*-positive and three *H. pylori*-negative samples). *TNFAIP3*, *TRAF3*, and *KLHL6*, which were mutated in *H. pylori*-negative gastric MALT lymphoma cases, were also mutated in *H. pylori*-negative samples (case 06). No somatic mutations were found in the cases (case 04 and case 07) with genetic translocations, t(11;18)(q21;q21)/*BIRC3-MALT1* (Figure 1A). *MUC4*, *TRIM*, *ALPG*, and other genes found in other cancer types were detected in both *H. pylori*-negative and *H. pylori*-positive cancer samples.

Gene expression analysis

To analyze the differentially expressed genes among the sample subgroups, we conducted multiple statistical tests using R statistical software. Our objective was to understand the transcriptional differences between the following samples at various stages of gastric MALT lymphoma carcinogenesis: 1) *H. pylori*-positive normal samples versus *H. pylori*-negative

normal samples; 2) *H. pylori*-positive cancer samples versus *H. pylori*-positive normal samples; 3) *H. pylori*-positive cancer samples versus *H. pylori*-negative cancer samples; and 4) *H. pylori*-negative cancer samples & *H. pylori*-negative normal samples.

1) *H. pylori*-positive normal samples & *H. pylori*-negative normal samples

The differences in gene expression between *H. pylori*-positive and *H. pylori*-negative normal samples were investigated to identify biological changes in the gastric mucosa when infected with *H. pylori*. Eleven down-regulated and two upregulated genes were statistically significant in the *H. pylori*-positive group. Of these, *MUC4*, *CXCL5*, *RNU4-2*, *NLRP7* and *IGKV2-24*, associated with regulating pH in the gastric environment and the proton pump, were down-regulated (Figure 2B; Table 2). From the results of the GSEA using Gene Ontology (GO) terms, the response to bacteria, immune response-regulating signaling pathway, inflammatory response, T cell activation, and B cell activation were upregulated in *H. pylori*-positive normal samples. TNF- α signaling and interferon gamma response were upregulated in *H. pylori*-positive normal samples, which was similar to the study that reported increased levels of cytokine, interferon- γ , and tumor necrosis factor- α in the *H. pylori*-infected human stomach [41]. Furthermore, upregulated B cell-mediated immunity and B cell receptor signaling pathways were found, which could develop into the oncogenic pathway of gastric MALT lymphoma.

2) Comparison of *H. pylori*-positive cancer & *H. pylori*-positive normal samples

When comparing *H. pylori*-positive cancer and *H. pylori*-positive normal samples, we anticipated discovering oncogenic dysregulation of genes and pathways, such as the NF- κ B and the B cell signaling pathways, as previously reported[42, 43].

From the DEG analysis, 263 upregulated and 128 down-regulated genes were observed. Of the

significantly dysregulated genes in *H. pylori*-positive cancer samples, *SELL*, *HLA-DMB*, *CXCR4* and *EIF3CL* were upregulated, whereas *NKX6-3*, *GAST*, and *MEIS1* were downregulated (Table 2). Following GSEA, upregulated immune and inflammatory responses in *H. pylori*-positive cancer samples were observed. Upregulation of the NF-κB signaling pathway, B cell proliferation, B cell signaling pathway via JAK-STAT, and hematopoietic cell lineage in cancer samples were found, all of which are considered as oncogenic pathways. These results are consistent with those of a previous study showing that dysregulation of oncogenic pathways following chronic inflammation induced by *H. pylori* plays a significant role in the development of gastric MALT lymphoma[2, 5, 6].

3) *H. pylori*-positive & *H. pylori*-negative cancer samples

The analysis of DEGs revealed 22 upregulated and 12 down-regulated genes between *H. pylori*-positive and *H. pylori*-negative cancer samples. In that set, *CXCL5*, *CXCL8* and *CXCL2* were up-regulated (Figure 2A). Gene set enrichment analysis indicated that immune response, inflammatory response, and pathogen infection were upregulated in the *H. pylori*-positive cancer group; this supports inflammation-related carcinogenesis in *H. pylori*-infected cases. The absence of a difference in the carcinogenic pathways, such as the NF-κB signaling pathway, suggests common carcinogenesis at the molecular level regardless of *H. pylori* infection status. Moreover, the causes of *H. pylori*-negative gastric MALT lymphoma may include gene fusion events, infections with other microbiota, and autoimmune diseases. Of the *H. pylori*-negative samples, two exhibited the t(11;18)(q21;q21)/*BIRC3-MALT1* gene fusion. None of the samples showed characteristics of an autoimmune disease.

4) *H. pylori*-negative cancer & *H. pylori*-negative normal samples

To investigate the potential drivers of *H. pylori*-negative gastric MALT lymphoma, DEG and GSEA were performed on *H. pylori*-negative cancer and *H. pylori*-negative normal samples. A

total of 766 genes were differentially expressed between the groups. The upregulated genes in cancer tissues included *CNR1*, *RUBCNL*, *C3*, *AIM2*, *CD37* and *CCL19*, whereas *C11orf86*, *EDN3*, *CHGB*, and *SST* were down-regulated in cancer samples (Table 2). GSEA revealed that the GO terms (B cell receptor signaling pathway, B cell proliferation, and immune responses) and the KEGG (NF-κB and JAK-STAT signaling pathways) were upregulated in cancer samples (Figure 2C). Additionally, the activation of pathways related to *Staphylococcus aureus* infection, tuberculosis, and toxoplasmosis in KEGG terms was detected, suggesting that some of the *H. pylori*-negative samples may be related to infections other than *H. pylori*.

Microbiome analysis

Using kraken2, a k-mer-based taxonomic classification, sequencing reads not aligned to the human reference genome, including follow-up samples, were classified. A total of 37 WTS data points were used to detect microbiomes. Initially, the relative abundance of the microbiome at the genus level was analyzed. Some cases that were *H. pylori*-positive had no *g_Helicobacter*, whereas some *H. pylori*-negative cases had *g_Helicobacter*. Across all samples, various microbiomes were detected, but certain genera consistently exhibited high abundance: *g_Helicobacter*, *g_Rothia*, *g_Veillonella*, and *g_Porphyromonas* (Figure 3A).

As previously reported, patients infected with *H. pylori* have a lower alpha diversity than *H. pylori*-negative cases[44]. The alpha diversities of the samples were investigated using the Shannon and Simpson indices, and no difference between *H. pylori*-positive and *H. pylori*-negative samples was found ($p = 0.4$). Therefore, a linear regression analysis was conducted to determine whether *H. pylori* was associated with alpha diversity, and to assess the impact of other microbiomes on alpha diversity. The quantified counts of the sequenced microbiome reads were used as independent variables and alpha diversity was used as the response variable.

In the univariate linear regression, among the four genera (*g_Helicobacter*, *g_Rothia*, *g_Veillonella*, and *g_Porphyromonas*) only *g_Helicobacter* was correlated with alpha diversity (Shannon: beta= 0.017826, p = 1.64×10^{-7} / Simpson: beta= -0.004026, p = 6.56×10^{-8}).

As alpha diversity decreased with an increase in *g_Helicobacter* reads, a multivariate linear regression with the four genera was conducted to identify linearly-related and independent variables. Both *g_Rothia* and *g_Helicobacter* were negatively correlated with alpha diversity (Figure 3B–C). In general, low alpha diversity is associated with unfavorable gut microbiome conditions related to chronic illnesses. Considering the correlation between *g_Helicobacter* and *g_Rothia*, and alpha diversity, we suggest that *g_Rothia* may be a potential factor contributing to *H. pylori*-negative gastric MALT lymphoma.

Discussion

In this study, we prospectively collected samples and generated next-generation sequencing (NGS) data from patients with gastric MALT lymphoma and divided them into two groups for further analysis: *H. pylori*-positive and *H. pylori*-negative. WES and WTS data were produced and utilized to analyze genetic alteration, DEGs, GSEA, and microbiome detection.

Using fusion gene analysis, the *BIRC3–MALT1* gene fusion in *H. pylori*-negative cases was identified, consistent with previous findings[15-17]. Additionally, we discovered an inversion of *TBL1XR1–TP63* in *H. pylori*-positive cases. The *TBL1XR1–TP63* gene fusion has been reported as a recurrent somatic gene fusion in high-grade lymphomas, such as diffuse large B cell lymphoma (DLBCL) and is known to drive tumor survival through *EZH2* and upregulate *MYC*, *EZH2*, and *EED*[40, 45]. Indeed, the sample with the *TBL1XR1–TP63* gene fusion in the current study showed high expression of *MYC*, *EZH2*, and *EED* (supplemental Figure 1). In

addition, patients with the *TBL1XR1-TP63* gene fusion showed a poor response to *H. pylori* eradication therapy. We speculate that the *TBL1XR1-TP63* fusion is one of the drivers of the high-grade transformation of gastric MALT lymphoma. Biological studies suggest B-cell lymphoma with *TBL1XR1-TP63* gene fusion might be susceptible to EZH1/2 inhibition[45].

Using RNA sequencing data from the fresh gastric mucosa of various clinical statuses, the biological phenomena associated with gastric MALT lymphomagenesis were successfully revealed. A multi-step, comparative analysis was conducted to detail the mechanism of gastric MALT lymphoma caused by *H. pylori*. First, by comparing *H. pylori*-positive and negative normal groups, the biological changes that occurred in the gastric mucosa upon infection with *H. pylori* were identified. It was expected that immune and inflammatory responses would be upregulated in the *H. pylori*-positive group due to the infection with *H. pylori*, which was indeed the case. The TNF- α signaling pathway was upregulated in the *H. pylori*-positive group, consistent with the studies that *H. pylori* induces TNF- α in the gastric mucosa[46, 47]. When comparing the *H. pylori*-positive cancer and *H. pylori*-positive normal groups, upregulation of oncogenic pathways, such as the NF- κ B and B cell signaling pathways, was observed in the cancer group. Furthermore, in *H. pylori*-infected patients, the cancer group exhibited upregulation of chronic inflammation, immune response, and other pathways associated with the bacterial response in comparison with the normal group. The elevated expression of *CXCL5* and *CXCL8*, which had been studied as being upregulated by the inflammatory response after *H. pylori* infection, is also observed in the cancer group infected with *H. pylori*[48]. This suggests that inflammation triggered by *H. pylori* continues to occur in the cancer group. The upregulation of the immune response and other factors in *H. pylori*-positive compared to *H. pylori*-negative gastric MALT lymphoma further support this hypothesis. In summary, our stepwise approach confirmed that chronic inflammation and immune responses play critical roles in *H. pylori* lymphomagenesis. These findings suggest that for the treatment of gastric

MALT lymphoma with *H. pylori* infection, immunotherapy agents may yield benefits.

It is also worth mentioning the pathophysiology of *H. pylori*-negative gastric MALT lymphomas using the stepwise approach. In the comparison of cancer and normal samples within the *H. pylori*-negative cases, aberrant NF-κB and B cell signaling pathways were dominant in lymphoma samples. This concurs with the studies that identified the upregulated NF-κB signaling pathway induced by some recurrent translocations [t(1;14)(p22;q32), t(14;18)(q32;q21), and t(11;18)(q21;q21)] in MALT lymphoma[49, 50]. Furthermore, upregulated biological pathways associated with pathogen infection were identified in the *H. pylori*-negative cancer compared to the *H. pylori*-negative normal group. These findings deepen our understanding of the molecular pathogenesis of *H. pylori*-negative gastric MALT lymphomas.

Based on the dysregulated pathways potentially influenced by specific microbiomes, we explored whether microbiomes other than *H. pylori* could be linked to gastric MALT lymphoma. There was a time when it was believed that no microorganisms could survive in the stomach; however, the discovery of *H. pylori* confirmed that microorganisms can thrive in the gastric environment. With advancements in NGS and analytical technologies, it has become evident that, in addition to *H. pylori*, various other microorganisms can interact within the gastric milieu. Studies have been conducted in Korean cohorts to analyze the stomach microbiome[51, 52]. Apart from the *g_Helicobacter*, diverse microbiota have been identified. In the current study, reads that were not mapped to the human reference genome in the WTS data were reclassified as part of the microbiome analysis, facilitating the detection of the microbiome present in each sample. Relative abundance and species diversity were investigated. The *g_Rothia* generally exhibited high abundance at the genus level, but several other genera also displayed noteworthy abundance. As observed in previous studies, *H. pylori*-

infected patients exhibited lower alpha diversity at the genus level, as confirmed by linear regression. Moreover, multivariate linear regression suggested an association between *g_Rothia* and low alpha diversity, particularly in conjunction with *g_Helicobacter*. Low alpha diversity is considered indicative of an unhealthy microbial environment and has been studied in relation to chronic diseases. Although a direct association between low diversity and gastric MALT lymphoma has not been firmly established, given the confirmed association with *g_Helicobacter*, it is plausible that the *g_Rothia* plays a significant role in the gastric environment and may contribute to gastric MALT lymphoma.

List of abbreviations

DEG differentially expressed gene

GO Gene Ontology

GSEA gene set enrichment analysis

KEGG Kyoto Encyclopedia of Genes and Genomes

MALT gastric mucosa-associated tissue

NGS next-generation sequencing

WES whole exome sequencing

WTS whole-transcriptome sequencing

Declarations

Not applicable

Ethics approval and consent to participate

All research was approved by the Institutional Review Board (IRB) of Seoul National University Hospital (IRB No. H-1704-070-845). All participants provided written informed consent to participate in this study.

Consent for publication

Not applicable

Availability of data and materials

Sequence data that support the findings of this study have been deposited in the Sequence Read Archive with the primary accession code PRJNA1088908.

Competing interests

The authors declare no competing interests.

Funding

This study was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science and ICT, South Korea(grand no.2021R1A4A2001553)

Authors' contributions

Acknowledgements

We thank the Global Science experimental Data hub Center (GSDC) and Korea Research Environment Open NETwork (KREONET) service for data computing and network provided by the Korea Institute of Science and Technology Information (KISTI).

Authors' information (optional)

Seung-Joo Yoo and Youngil Koh contributed equally to this work

References

1. Du M-Q, Isaccson PG: **Gastric MALT lymphoma: from aetiology to treatment.** *The lancet oncology* 2002, **3**(2):97-104.
2. Isaacson PG, Du MQ: **Gastrointestinal lymphoma: where morphology meets molecular biology.** *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland* 2005, **205**(2):255-274.
3. Zullo A, Hassan C, Ridola L, Repici A, Manta R, Andriani A: **Gastric MALT lymphoma: old and new insights.** *Annals of Gastroenterology* 2014, **27**(1):27.
4. Bayerdörffer E, Rudolph B, Neubauer A, Thiede C, Lehn N, Eidt S, Stolte M, Group MLS: **Regression of primary gastric lymphoma of mucosa-associated lymphoid tissue type after cure of Helicobacter pylori infection.** *The Lancet* 1995, **345**(8965):1591-1594.
5. Sagaert X, Van Cutsem E, De Hertogh G, Geboes K, Tousseyn T: **Gastric MALT lymphoma: a model of chronic inflammation-induced tumor development.** *Nature reviews Gastroenterology & hepatology* 2010, **7**(6):336-346.
6. Troppan K, Wenzl K, Neumeister P, Deutsch A: **Molecular pathogenesis of MALT lymphoma.** *Gastroenterology research and practice* 2015, **2015**.
7. Stolte M, Bayerdörffer E, Morgner A, Alpen B, Wündisch T, Thiede C, Neubauer A: **Helicobacter and gastric MALT lymphoma.** *Gut* 2002, **50**(suppl 3):iii19-iii24.
8. Liu H, Ruskon-Fourmestraux A, Lavergne-Slove A, Ye H, Molina T, Bouhnik Y, Hamoudi RA, Diss TC, Dogan A, Megraud F: **Resistance of t (11; 18) positive gastric mucosa-associated lymphoid tissue lymphoma to Helicobacter pylori eradication therapy.** *The Lancet* 2001, **357**(9249):39-40.
9. Filip PV, Cuciureanu D, Diaconu LS, Vladareanu AM, Pop CS: **MALT lymphoma: epidemiology, clinical diagnosis and treatment.** *Journal of medicine and life* 2018, **11**(3):187.
10. Juárez-Salcedo LM, Sokol L, Chavez JC, Dalia S: **Primary gastric lymphoma, epidemiology, clinical diagnosis, and treatment.** *Cancer Control* 2018, **25**(1):1073274818778256.
11. Thieblemont C, Zucca E: **Clinical aspects and therapy of gastrointestinal MALT lymphoma.** *Best Practice & Research Clinical Haematology* 2017, **30**(1-2):109-117.
12. Zhang J, Wei J, Wang Z, Feng Y, Wei Z, Hou X, Xu J, He Y, Yang D: **Transcriptome hallmarks in Helicobacter pylori infection influence gastric cancer and MALT lymphoma.** *Epigenomics* 2020, **12**(8):661-671.
13. Sagaert X, Van Cutsem E, De Hertogh G, Van Loo P, De Wolf-Peeters C, Geboes K, Tousseyn T: **Microarray ANALYSIS of t (11; 18)(Q21; Q21)-positive and-negative gastrointestinal MALT lymphomas.** *Journal of Clinical Oncology* 2010, **28**(15_suppl):e14610-e14610.
14. Quang Huynh M, Wacker H-H, Wündisch T, Sohlbach K, Daniel Kim T, Krause M, Stabla K, Roth P, Fischbach W, Stolte M: **Expression profiling reveals specific gene expression signatures in gastric MALT lymphomas.** *Leukemia & lymphoma* 2008, **49**(5):974-983.

15. Asano N, Iijima K, Koike T, Imatani A, Shimosegawa T: **Helicobacter pylori-negative gastric mucosa-associated lymphoid tissue lymphomas: a review.** *World journal of gastroenterology: WJG* 2015, **21**(26):8014.
16. Ye H, Liu H, Raderer M, Chott A, Ruskone-Fourmestraux A, Wotherspoon A, Dyer MJ, Chuang S-S, Dogan A, Isaacson PG: **High incidence of t (11; 18)(q21; q21) in Helicobacter pylori-negative gastric MALT lymphoma.** *Blood, The Journal of the American Society of Hematology* 2003, **101**(7):2547-2550.
17. Takigawa H, Yuge R, Masaki S, Otani R, Kadota H, Naito T, Hayashi R, Urabe Y, Oka S, Tanaka S: **Involvement of non-Helicobacter pylori helicobacter infections in Helicobacter pylori-negative gastric MALT lymphoma pathogenesis and efficacy of eradication therapy.** *Gastric Cancer* 2021, **24**:937-945.
18. Tanaka T, Matsuno Y, Torisu T, Shibata H, Hirano A, Umeno J, Kawasaki K, Fujioka S, Fuyuno Y, Moriyama T: **Gastric microbiota in patients with Helicobacter pylori-negative gastric MALT lymphoma.** *Medicine* 2021, **100**(38).
19. Andrews S: **FastQC: a quality control tool for high throughput sequence data.** In.: Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom; 2010.
20. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data.** *Bioinformatics* 2014, **30**(15):2114-2120.
21. Li H: **Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.** *arXiv preprint arXiv:13033997* 2013.
22. McKenna N, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M: **The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.** *Genome research* 2010, **20**(9):1297-1303.
23. Wang K, Li M, Hakonarson H: **ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.** *Nucleic acids research* 2010, **38**(16):e164-e164.
24. Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, Bhai J: **Ensembl 2021.** *Nucleic acids research* 2021, **49**(D1):D884-D891.
25. Consortium GP, Auton A, Brooks L, Durbin R, Garrison E, Kang H: **A global reference for human genetic variation.** *Nature* 2015, **526**(7571):68-74.
26. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W: **ClinVar: improving access to variant interpretations and supporting evidence.** *Nucleic acids research* 2018, **46**(D1):D1062-D1067.
27. Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, Alföldi J, Watts NA, Vittal C, Gauthier LD *et al.* **A genome-wide mutational constraint map quantified from variation in 76,156 human genomes.** *bioRxiv* 2022:2022.2003.2020.485034.
28. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, Creatore C, Dawson E: **COSMIC: the catalogue of somatic mutations in cancer.** *Nucleic acids research* 2019, **47**(D1):D941-D947.
29. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras

TR: **STAR: ultrafast universal RNA-seq aligner.** *Bioinformatics* 2013, **29**(1):15-21.

30. Li B, Dewey CN: **RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.** *BMC bioinformatics* 2011, **12**:1-16.

31. R Core Team R: **R: A language and environment for statistical computing.** 2013.

32. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.** *Genome biology* 2014, **15**(12):1-21.

33. Nicorici D, Şatalan M, Edgren H, Kangaspeska S, Murumägi A, Kallioniemi O, Virtanen S, Kilkku O: **FusionCatcher-a tool for finding somatic fusion genes in paired-end RNA-sequencing data.** *biorxiv* 2014:011650.

34. Haas BJ, Dobin A, Stransky N, Li B, Yang X, Tickle T, Bankapur A, Ganote C, Doak TG, Pochet N: **STAR-Fusion: fast and accurate fusion transcript detection from RNA-Seq.** *BioRxiv* 2017:120295.

35. Uhrig S, Ellermann J, Walther T, Burkhardt P, Fröhlich M, Hutter B, Toprak UH, Neumann O, Stenzinger A, Scholl C: **Accurate and efficient detection of gene fusions from RNA sequencing data.** *Genome research* 2021, **31**(3):448-460.

36. Thorvaldsdóttir H, Robinson JT, Mesirov JP: **Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration.** *Briefings in bioinformatics* 2013, **14**(2):178-192.

37. Lu J, Rincon N, Wood DE, Breitwieser FP, Pockrandt C, Langmead B, Salzberg SL, Steinegger M: **Metagenome analysis using the Kraken software suite.** *Nature protocols* 2022, **17**(12):2815-2839.

38. Wood DE, Lu J, Langmead B: **Improved metagenomic analysis with Kraken 2.** *Genome biology* 2019, **20**:1-13.

39. Hyeon J, Lee B, Shin S-H, Yoo HY, Kim SJ, Kim WS, Park W-Y, Ko Y-H: **Targeted deep sequencing of gastric marginal zone lymphoma identified alterations of TRAF3 and TNFAIP3 that were mutually exclusive for MALT1 rearrangement.** *Modern Pathology* 2018, **31**(9):1418-1428.

40. Scott DW, Mungall KL, Ben-Neriah S, Rogic S, Morin RD, Slack GW, Tan KL, Chan FC, Lim RS, Connors JM: **TBL1XR1/TP63: a novel recurrent gene fusion in B-cell non-Hodgkin lymphoma.** *Blood, The Journal of the American Society of Hematology* 2012, **119**(21):4949-4952.

41. Yamamoto T, Kita M, Ohno T, Iwakura Y, Sekikawa K, Imanishi J: **Role of tumor necrosis factor-alpha and interferon-gamma in Helicobacter pylori infection.** *Microbiology and immunology* 2004, **48**(9):647-654.

42. Du M-Q: **MALT lymphoma: A paradigm of NF-κB dysregulation.** In: *Seminars in cancer biology*: 2016: Elsevier; 2016: 49-60.

43. Du M-Q: **MALT lymphoma: genetic abnormalities, immunological stimulation and molecular mechanism.** *Best Practice & Research Clinical Haematology* 2017, **30**(1-2):13-23.

44. Sugihartono T, Fauzia KA, Miftahussurur M, Waskito LA, Rejeki PS, I'tishom R, Alfaray RI,

Doohan D, Amalia R, Savitri CMA: **Analysis of gastric microbiota and Helicobacter pylori infection in gastroesophageal reflux disease.** *Gut Pathogens* 2022, **14**(1):38.

45. Wu G, Yoshida N, Liu J, Zhang X, Heavican-Foral TB, Liu H, Nelson GM, Yang L, Chen R, Jones MK: **TP63 fusions drive enhancer rewiring, lymphomagenesis, and dependence on EZH2.** *Blood* 2022, **140**(Supplement 1):1696-1697.

46. Crabtree J, Shallcross T, Heatley R, Wyatt J: **Mucosal tumour necrosis factor alpha and interleukin-6 in patients with Helicobacter pylori associated gastritis.** *Gut* 1991, **32**(12):1473-1477.

47. Moss S, Legon S, Davies J, Calam J: **Cytokine gene expression in Helicobacter pylori associated antral gastritis.** *Gut* 1994, **35**(11):1567-1570.

48. Tran CT, Garcia M, Garnier M, Buruoa C, Bodet C: **Inflammatory signaling pathways induced by Helicobacter pylori in primary human gastric epithelial cells.** *Innate immunity* 2017, **23**(2):165-174.

49. Uren AG, O'Rourke K, Aravind LA, Pisabarro MT, Seshagiri S, Koonin EV, Dixit VM: **Identification of paracaspases and metacaspases: two ancient families of caspase-like proteins, one of which plays a key role in MALT lymphoma.** *Molecular cell* 2000, **6**(4):961-967.

50. Lucas PC, Yonezumi M, Inohara N, McAllister-Lucas LM, Abazeed ME, Chen FF, Yamaoka S, Seto M, Núñez G: **Bcl10 and MALT1, independent targets of chromosomal translocation in malt lymphoma, cooperate in a novel NF-κB signaling pathway.** *Journal of Biological Chemistry* 2001, **276**(22):19012-19019.

51. Gunathilake M, Lee J, Choi IJ, Kim Y-I, Yoon J, Sul WJ, Kim JF, Kim J: **Alterations in gastric microbial communities are associated with risk of gastric cancer in a Korean population: A case-control study.** *Cancers* 2020, **12**(9):2619.

52. Gunathilake MN, Lee J, Choi IJ, Kim Y-I, Ahn Y, Park C, Kim J: **Association between the relative abundance of gastric microbiota and the risk of gastric cancer: a case-control study.** *Scientific reports* 2019, **9**(1):13589.

Tables

No	Age	Gender	Location	<i>H. pylori</i>
1	69	M	MBPW	positive
2	37	M	Angle	Positive
3	37	F	AntrumGC	Negative
4	57	M	AntrumGC	Negative
5	71	F	AntrumAW	Positive
6	68	F	MBPW	Negative
7	67	F	AntrumGC	Negative
8	47	F	MBGC	Positive
9	71	F	LBPW	Positive
10	53	F	LBLC	Positive
11	52	F	AntrumGC	Positive
12	60	M	MBGC	Positive
13	57	F	LBAW	Positive

Table 1. Clinicopathologic features of gastric MALT lymphoma. MB: mid body, LB: lower body, PW: post wall, GC: great curvature, AW: ant wall, LC: less curvature

GENE	log2FoldChange	padj
<i>H. pylori</i> -positive normal vs <i>H. pylori</i> -negative normal		
<i>MUC4</i>	4.768	1.900E-17
<i>CXCL5</i>	6.539	2.089E-09
<i>RNU4-2</i>	-3.316	2.179E-05
<i>NLRP7</i>	4.434	3.245E-04
<i>IGKV2-24</i>	3.369	6.311E-04
<i>H. pylori</i> -positive cancer vs <i>H. pylori</i> -positive normal		
<i>SELL</i>	2.241	1.035E-06
<i>HLA-DMB</i>	1.241	2.269E-06
<i>NKX6-3</i>	-4.902	5.988E-06
<i>IRF8</i>	1.189	3.295E-05
<i>TMEM178B</i>	1.453	3.476E-05
<i>EIF3CL</i>	7.380	3.869E-05
<i>CXCR4</i>	1.577	3.878E-05
<i>H. pylori</i> -positive cancer vs <i>H. pylori</i> -negative cancer		
<i>CXCL5</i>	7.085	1.843E-08
<i>CXCL8</i>	6.188	1.425E-04
<i>FCGR2C</i>	-2.382	1.425E-04
<i>ABCA12</i>	4.532	2.719E-04
<i>FCGR3B</i>	5.330	6.975E-04
<i>FFAR2</i>	4.352	2.534E-03
<i>CXCL2</i>	2.633	2.631E-03
<i>H. pylori</i> -negative cancer vs <i>H. pylori</i> -negative normal		
<i>CNR1</i>	4.750	2.195E-06
<i>RUBCNL</i>	4.350	6.220E-06
<i>C3</i>	4.167	6.220E-06
<i>AIM2</i>	4.766	6.220E-06
<i>CD37</i>	4.089	6.431E-06
<i>CCL19</i>	4.952	6.430E-06

Table 2. A table displays the results of DEG analysis from multiple comparisons. The genes in this table were selected by log2FoldChange and padj values and the full set of genes which were statistically significant are listed in the supplemental Table S1-

Figures

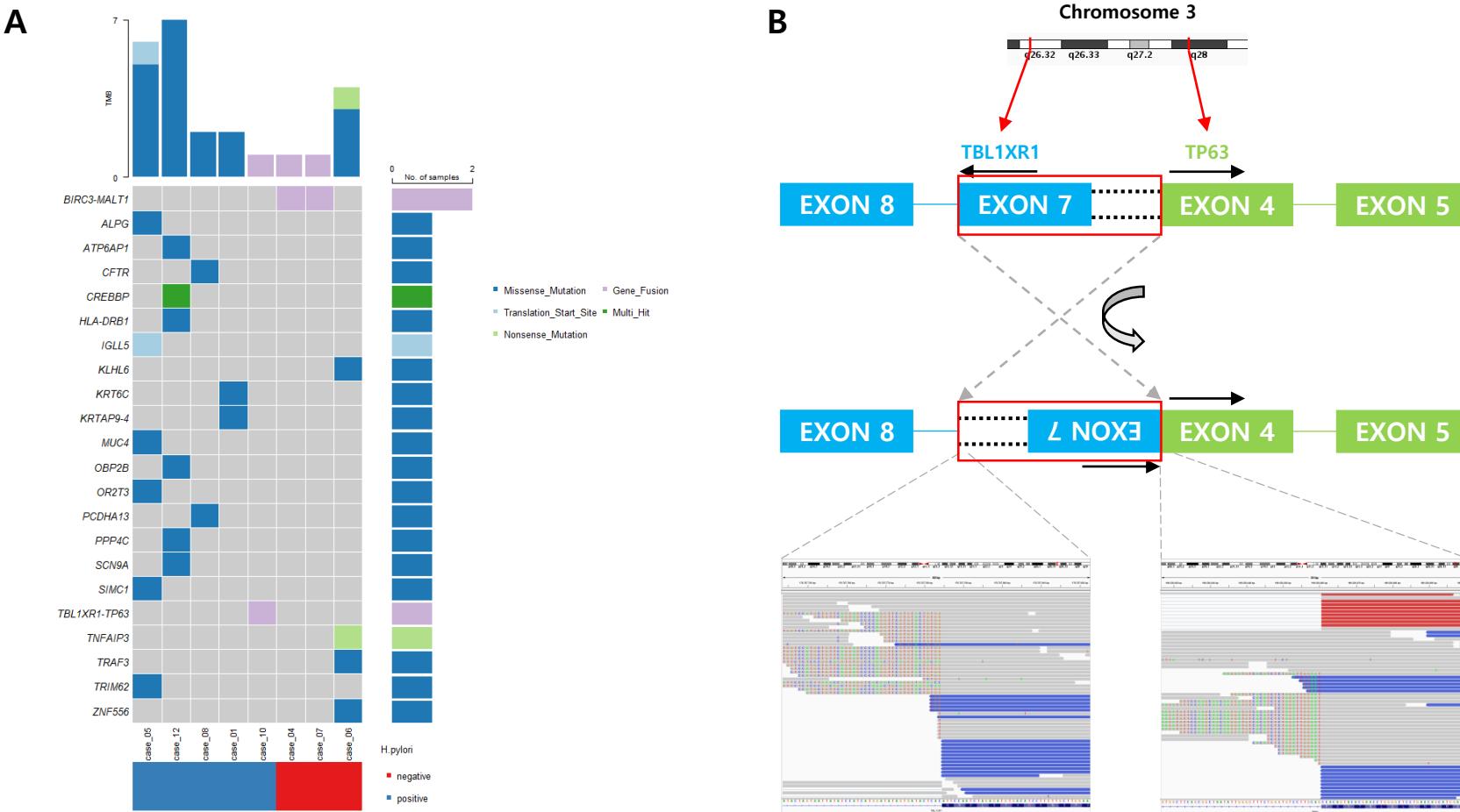


Figure 1 A-B. **A.** An oncoplot represents the somatic mutations from WES data including genetic translocations from WTS data. **B.** A schematic diagram represents the inversion of *TBL1XR1* and *TP63*. Inversion of *TBL1XR1* and *TP63*. A schematic diagram shows detailed breakpoints of the genetic translocations. The screenshot of IGV shows the evidence sequencing reads from the cancer sample.

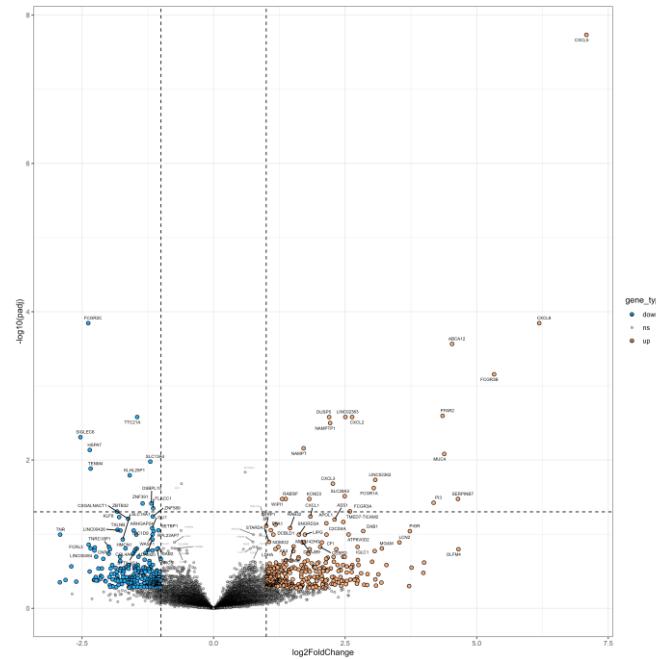
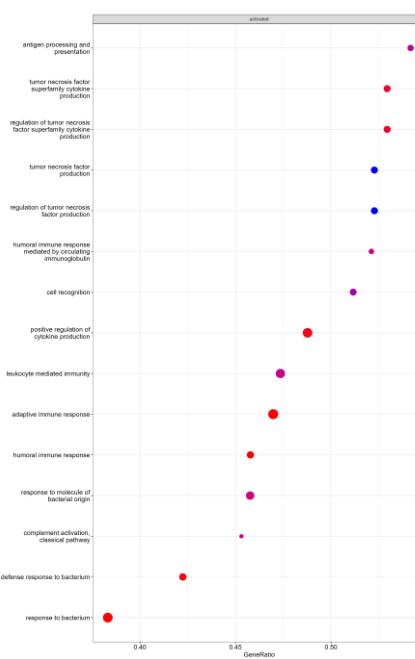
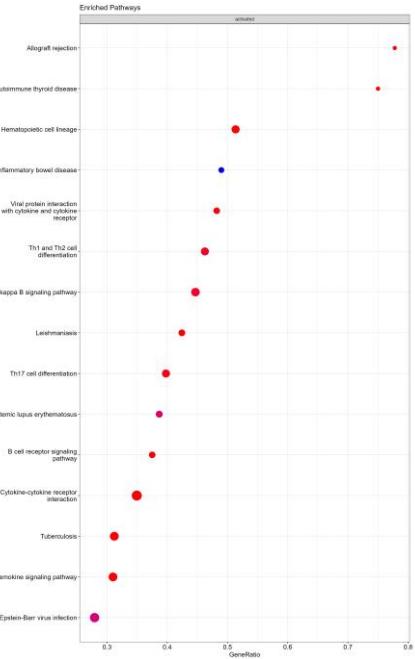
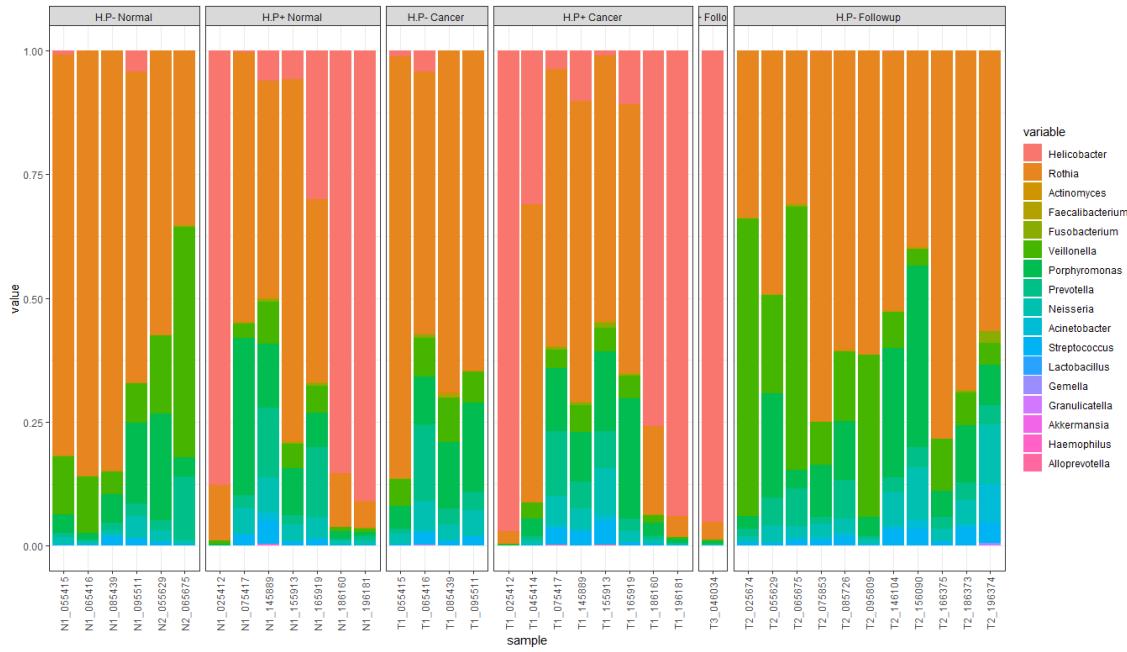
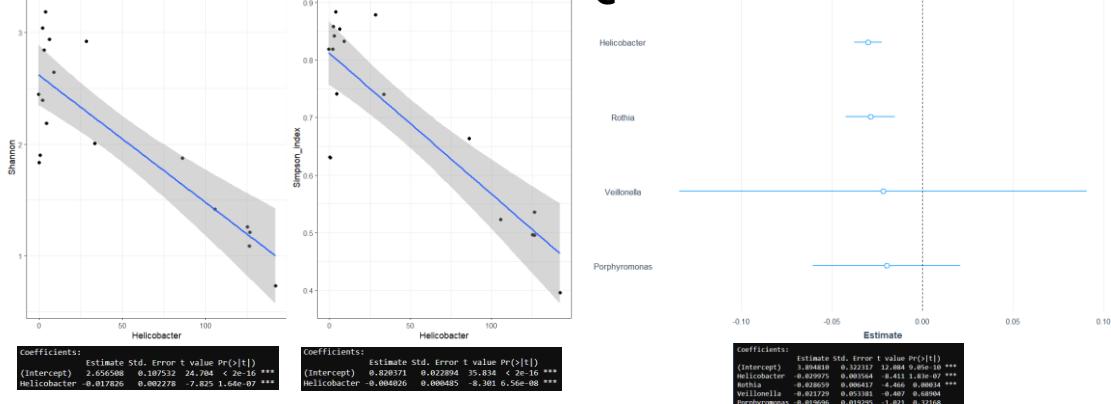
A**B****C**

Figure 2 A-C. **A.** A volcano plot shows the results of DEG analysis between *H. pylori*-positive cancer samples and *H. pylori*-negative cancer samples. Genes that have the following range of fold changes are marked with blue and red colors and the dashed line shows a cut-off of adjusted p-value and fold changes. **B-C.** The two plots show the results of GSEA from " *H. pylori*-positive normal samples & *H. pylori*-negative normal samples" and " *H. pylori*-negative cancer samples & *H. pylori*-negative normal samples."

A



B



C

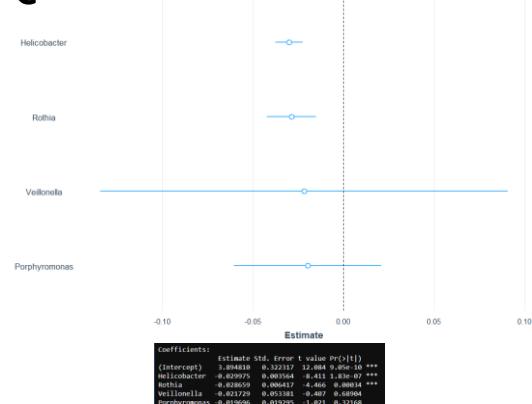


Figure 3. A-C : A. A stacked bar shows a relative abundance of the microbiome from all WTS data. The designation of each group means the presence of *H. pylori* and tissue type. B. Each dot represents an index value matching the quantified read count of *H. pylori*. C. The dots and lines represent the estimate and error term from multivariate linear regression.