Am J Cancer Res 2021;11(2):561-575
www.ajcr.us /ISSN:2156-6976/ajcr0123246

Original Article

Microbiota alteration at different stages in gastric lesion progression: a population-based study in Linqu, China

Gaohaer Kadeerhan[1], Markus Gerhard[2,3,4], Juan-Juan Gao[1], Raquel Mejías-Luque[2,3,4], Lian Zhang[1], Michael Vieth[2,5], Jun-Ling Ma[1], Monther Bajbouj[2,6], Stepan Suchanek[2,7], Wei-Dong Liu[8], Kurt Ulm[2,9], Michael Quante[2,6], Zhe-Xuan Li[1,2], Tong Zhou[1], Roland Schmid[2,6], Meinhard Classen[2,10], Wen-Qing Li[1,2], Yang Zhang[1,2], Wei-Cheng You[1,2], Kai-Feng Pan[1,2]

[1]Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Cancer Epidemiology, Peking University Cancer Hospital & Institute, Beijing, China; [2]PYLOTUM Key Joint Laboratory for Upper GI Cancer, Technische Universität München, Munich, Germany, Peking University Cancer Hospital & Institute, Beijing, China; [3]Institute of Medical Microbiology, Immunology and Hygiene, Technische Universität München, Munich, Germany; [4]German Center for Infection Research, Partner Site Munich, Munich, Germany; [5]Institute of Pathology, Klinikum Bayreuth, Bayreuth, Germany; [6]II. Medizinische Klinik, Klinikum Rechts der Isar, Technische Universität München, Munich, Germany; [7]Department of Medicine, 1st Faculty of Medicine, Military University Hospital, Charles University, Prague, Czech Republic; [8]Linqu Public Health Bureau, Linqu, Shandong, China; [9]Institute of Medical Informatics, Statistics and Epidemiology, Technische Universität München, Munich, Germany; [10]International Digestive Cancer Alliance, Germany

Received September 27, 2020; Accepted December 16, 2020; Epub February 1, 2021; Published February 15, 2021

Abstract: In addition to Helicobacter pylori (H.pylori), gastric microbiota may be involved in carcinogenesis process. However, the longitudinal study to assess changes in the gastric microbiota associated with the development of gastric carcinogenesis is still limited. The aim of this study is to explore dynamic microbial alterations in gastric cancer (GC) development based on a 4-year endoscopic follow-up cohort in Linqu County, China. Microbial alterations were investigated by deep sequencing of the microbial 16S ribosomal RNA gene in 179 subjects with various gastric lesions, and validated in paired gastric biopsies prospectively collected before and after lesion progression and in non-progression controls. Significant differences were found in microbial diversity and community structure across various gastric lesions, with 62 candidate differential taxa between at least two lesion groups. Further validations identified Helicobacter, Bacillus, Capnocytophaga and Prevotella to be associated with lesion progression-to-dysplasia (DYS)/GC (all $P < 0.05$), especially for subjects progressing from intestinal metaplasia (IM) to DYS/GC. The combination of the four genera in a microbial dysbiosis index showed a significant difference after lesion progression-to-DYS/GC compared to controls ($P = 0.027$). The panel including the four genera identified subjects after progression-to-DYS/GC with an area under the receiver-operating curve (AUC) of 0.941. Predictive significance was found before lesion progression-to-DYS/GC with an AUC = 0.776 and an even better AUC (0.927) for subjects progressing from IM to DYS/GC. Microbiota may play different roles at different stages in gastric carcinogenesis. A panel of bacterial genera associated with gastric lesions may help to assess gastric microbial dysbiosis and show potential predictive values for lesion progression. Our findings provide new clues for the microbial mechanism of H.pylori-associated carcinogenesis.

Keywords: Gastric microbiota, lesion progression, Helicobacter pylori, dysbiosis

Introduction

Gastric cancer (GC) is one of the most common malignancy in the world and in China [1, 2]. Deeper understanding of the etiological factors for gastric carcinogenesis is urgently needed for GC control and prevention.

Helicobacter pylori (*H.pylori*) is one of the most important drivers of the multi-stage process leading to GC development [3]. Accelerated neoplasia progression by co-colonization of intestinal bacteria with *H.pylori* in insulin-gastrin mice suggests potential contributions of non-*H.pylori* bacteria to GC development [4, 5].

Retrospective studies have demonstrated *H.pylori*-associated microbial dysbiosis and altered bacterial interactions in precancerous gastric lesions and GC [6, 7]. However, the complex influence factors on gastric microbiota, such as host genetic background, dietary habit and history of antibiotic use, require a validation of previous results in a prospective study.

Our previous intervention study found recovery of gastric microbial dysbiosis and significant alterations of *H.pylori*-interactive bacteria (*Prevotella*, *Neisseria*, *Fusobacterium*, etc) in paired gastric biopsies before and six months after eradication [8]. Although the associations between these candidate *H.pylori*-interactive bacteria and precancerous lesions were preliminarily validated, the causal and temporal relationships between gastric microbiota and the natural evolution of precancerous lesions to GC still need solid evidence from a long-term follow-up cohort.

In the present study, gastric microbial profiling was compared in various gastric lesions by deep sequencing of the 16S ribosomal RNA (16S rRNA) gene. Differential taxa selected comprehensively according to the gastric lesion and *H.pylori* infection status were further validated in a 4-year endoscopic follow-up cohort with paired biopsies before and after lesion progression. This unique prospective self-control design helps us to better understand microbial alterations during GC evolution and to explore predictive microbial markers for gastric lesion progression.

Material and methods

Patient and public involvement

Linqu County in Shandong province, China, is a high-risk area with one of the highest GC mortality rates in the world. From 2012 to 2016, the National Upper Gastrointestinal Cancer Early Detection Project conducted endoscopic examinations in about 1500 Linqu County residents (aged 40-69 years) annually. About 70% of the project participants were selected using cluster randomization by village for initial screening, and 30% were invited from the previous screening participants (especially those subjects with advanced gastric lesions) for follow-up examination.

Within the framework of this project, 332 volunteers were recruited from 10 villages in December 2016 for initial screening. A total of 193 subjects were enrolled for completing standard upper endoscopic examination and providing extra fresh gastric biopsy samples for microbiota analysis. Among them, 16S rRNA sequencing results were successfully obtained from 158 subjects, including 35 showing normal/superficial gastritis (SG), 52 presenting chronic atrophic gastritis (CAG), 67 with intestinal metaplasia (IM), 2 with Dysplasia (DYS) and 2 with GC (**Figure 1**).

For the 4-year endoscopic follow-up participants, 31 cases were enrolled as progression subjects with higher gastric lesion grades in follow-up endoscopic examinations compared to initial screening and paired fresh gastric biopsies. For each progression subject, one control was randomly selected from individuals who did not show lesion progression from initial to follow-up time point. Controls were matched by sex, age, and calendar year of paired fresh biopsy collection to progression subjects. Sequencing results were successfully obtained in 26 initial (diagnosed as 8 normal/SGs, 4 CAGs, 11 IMs and 3 DYSs) and 28 follow-up biopsies (7 progressed to IM, 15 to DYS and 6 to GC) from 31 progression subjects, and 29 initial (13 normal/SGs, 10 CAGs, 6 IMs) and 26 follow-up biopsies (18 normal/SGs, 3 CAGs, and 5 IMs) from 31 non-progression controls. To investigate microbiota in various gastric lesions, we added the 21 follow-up DYS/GC subjects after lesion progression with completed sequencing results to 158 initial screening subjects for the small case number of initial 2 DYSs and 2 GCs (**Figure 1**).

All subjects provided general information about age, sex, cigarette and alcohol consumption habits, and written informed consent. This study was approved by the Institutional Review Boards of Peking University Cancer Hospital and Institute.

Upper endoscopic examination and histopathology

Upper endoscopic examinations were conducted by two experienced gastroenterologists using video endoscopes (Olympus). The gastric mucosa was examined and biopsies were col-

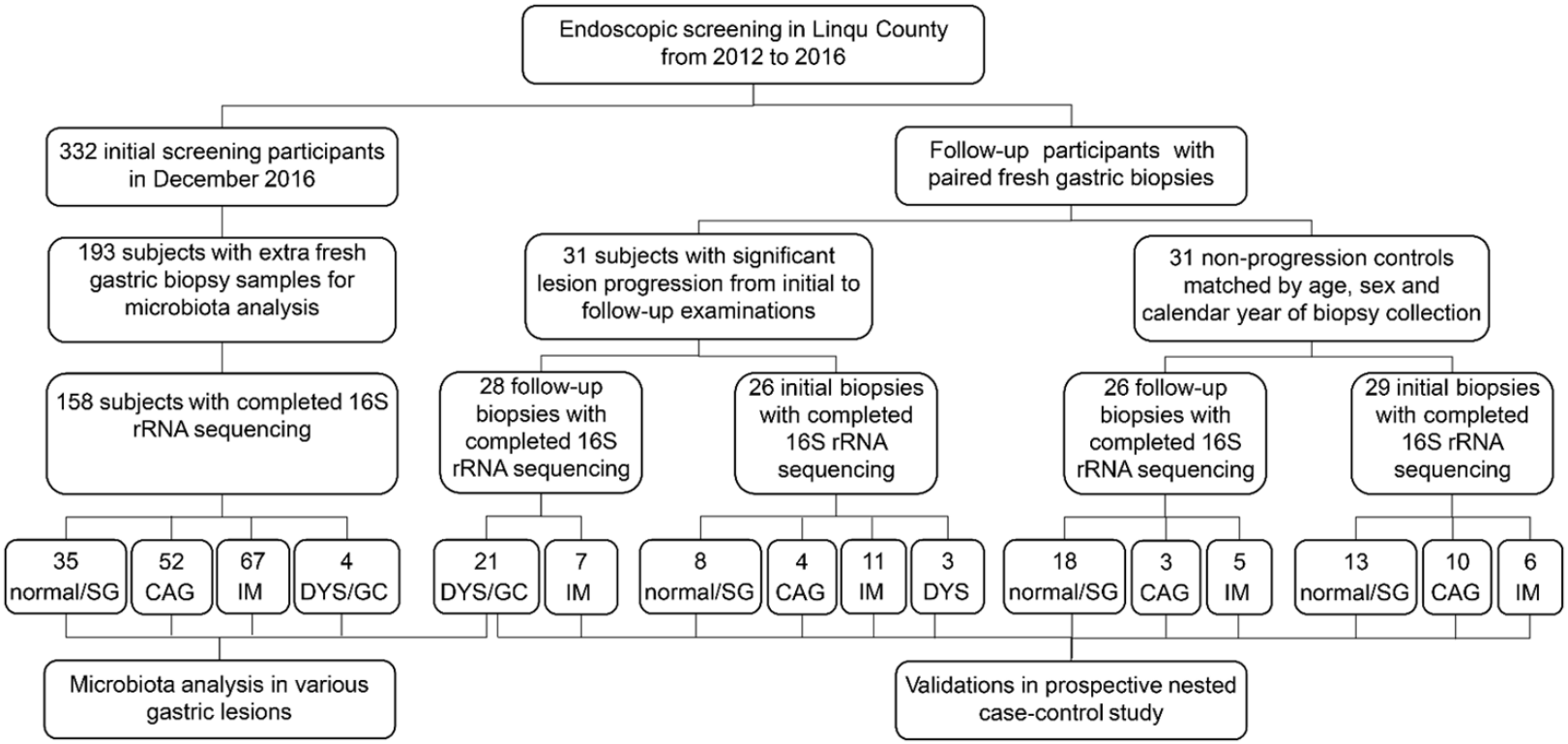

Figure 1. Flow diagram of study design and participant involved. CAG, chronic atrophic gastritis; DYS, dysplasia; GC, gastric cancer; IM, intestinal metaplasia; SG, superficial gastritis.

lected from the antrum or suspicious appearing lesions for pathological diagnosis. An extra fresh biopsy was taken from the lesser curvature of the antrum and frozen immediately in liquid nitrogen for microbiota analysis. The gastric mucosa specimens were reviewed blindly by two pathologists according to the Chinese Association of Gastric Cancer [9] and the Updated Sydney System [10]. Each biopsy was graded as normal, SG, CAG, IM, DYS and GC based on the most severe histology.

DNA extraction and 16S rRNA gene sequencing

DNA extraction was performed using the QIAamp DNA Mini Kit according to the manufacturer's instructions. The hypervariable region V3-V4 of microbial 16S rRNA gene was amplified using universal primers (341F, 5'-CCTACGGGNBGCASCAG-3'; 805R, 5'-GACTACNVGGGTATCTAAT CC-3'). The PCR products were purified using QIAquick Gel Extraction Kit (Qiagen). The resulting amplicon library was sequenced on an Illumina Hiseq 2500 PE250 platform.

Sequencing data analysis and H.pylori infection status determination

The 16S rRNA gene sequence raw reads were processed using the IMNGS (www.imngs.org) platform [11], a UPARSE based analysis pipeline [12]. Pairing, quality filtering and OTU clustering at 97% similarity with a relative abundance ≥ 0.1% in at least one sample were performed by USEARCH 8.0 [13]. Taxonomic classification was assigned by RDP classifier version 2.11 training set 15 [14].

To determine the *H.pylori* infection status, the 16S rRNA gene sequences were analyzed by QIIME software package and the UPARSE pipeline. The sequences were annotated to species level using the Greengenes database. Samples with *H.pylori* relative abundance < 1% were defined as *H.pylori*-negative, while samples with *H.pylori* relative abundance > 1% were defined as *H.pylori*-positive, as previously described [15].

Statistical analysis

Microbial diversity indexes were profiled using Rhea [16] based on R software. Comparisons of richness and Shannon indexes were performed by unconditional logistic regression adjusting for age, sex, smoking and alcohol consumption status among various gastric lesions. The generalized Unifrac distance was used for microbial community structure comparison, and non-metric multi-dimensional scaling (NMDS) plots were generated for visualization. *p* values were calculated by the PERMANOVA test and adjusted for multiple comparisons by the false discovery rate (FDR) [17]. The corresponding *q*-values < 0.05 were considered statistically significant.

Candidate differential taxa across gastric lesions were preliminarily selected for subsequent validation by unconditional logistic regression adjusting for age, sex, smoking and alcohol consumption status with *q*-values < 0.10 after multiple testing adjustment. Mann-Whitney U test was used for the comparisons of the candidate genera between progression and non-progression subjects. Multivariate logistic regression adjusted for age and sex was performed to compare the specific genera between progression-to-DYS/GC subjects and non-progression controls.

Functional capabilities of mucosal-associated microbiota was predicted using Tax4Fun [18] based on SILVA SSU rRNA database [19] and Kyoto Encyclopedia of Genes and Genomes (KEGG) [20]. The comparisons of functional compositions and metabolic pathways among subjects with different gastric lesions were performed by multivariate logistic regression adjusted for age, sex, smoking and alcohol consumption status, with a significance threshold of *q*-values < 0.05 after multiple testing adjustment by FDR. Further validation of candidate differential pathways between progression and non-progression IM subjects was performed by Mann-Whitney U test with the significance threshold of *p*-values < 0.05.

Results

Characteristics of study subjects

Among the 193 initial endoscopic screening subjects enrolled from the National Upper Gastrointestinal Cancer Early Detection Project in 2016 Dec, 16S rRNA sequencing results were obtained from 158 subjects, including 35 showing normal/SG, 52 presenting CAG, 67

with IM, 2 with DYS and 2 with GC. In the endoscopic follow-up participants from 2012 to 2016, 31 lesion progression subjects and 31 matched non-progression controls were enrolled with the paired fresh gastric biopsies from initial and follow-up examinations. Sequencing results were successfully obtained in 26 initial (diagnosed as 8 normal/SGs, 4 CAGs, 11 IMs and 3 DYSs) and 28 follow-up biopsies (7 progressed to IM, 15 to DYS and 6 to GC) from 31 progression subjects, and 29 initial (13 normal/SGs, 10 CAGs, 6 IMs) and 26 follow-up biopsies (18 normal/SGs, 3 CAGs, and 5 IMs) from 31 non-progression controls. To investigate microbiota in various gastric lesions, we added the 21 follow-up DYS/GC subjects after lesion progression with completed sequencing results to 158 initial screening subjects for the small case number of initial 2 DYSs and 2 GCs (**Figure 1**).

The general characteristics of the 179 subjects (158 initial screening subjects and 21 follow-up DYS/GC subjects) showing various gastric lesions are presented in Supplementary Table 1. Compared to normal/SG group, subjects in IM and DYS/GC groups were older and showed higher frequencies of males and cigarette smokers (all $P < 0.05$). Alcohol consumption was higher in CAG, IM and DYS/GC groups compared to normal/SG (all $P < 0.05$). The presence of *H.pylori* infection was increased significantly from normal/SG (74.3%) to CAG (92.3%) and IM (94.0%, both $P < 0.001$), while decreased in DYS/GC (60.0%, $P = 0.241$).

Associations between gastric microbial diversity and gastric lesions

Microbial alpha diversity analysis revealed that the richness and Shannon indexes were significantly decreased from normal/SG to CAG and IM (all $P < 0.001$). The indexes in DYS/GC were higher than those in CAG and IM (all $P < 0.001$), while similar to those in normal/SG (**Figure 2A**, **2B**).

Microbial community structure comparison found significant differences when comparing CAG, IM and DYS/GC with normal/SG (all $q = 0.002$) and when comparing DYS/GC and IM groups ($q = 0.038$), while no significant differences between IM and CAG ($q = 0.630$), or DYS/GC and CAG ($q = 0.536$) were detected (**Figure 2C-H**).

Differential taxa among various gastric lesions

To screen candidate differential bacteria across gastric lesions for further validation, the taxa with relative abundance median > 0.1% in at least one gastric lesion group were compared. A total of 62 candidate taxa were preliminarily selected with $q < 0.10$ after multiple-testing FDR correction in the comparisons of any two lesion groups (Supplementary Table 2). Among them, the relative abundances of 5 taxa were significantly higher in CAG and IM compared to normal/SG, but lower in DYS/GC compared to IM and CAG (all $q < 0.05$), which were all *H.pylori* related including *Proteobacteria* (phylum), *Epsilonproteobacteria* (class), *Campylobacterales* (order), *Helicobacteraceae* (family) and *Helicobacter* (genus).

In the other 57 candidate taxa, the abundances of *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, *Fusobacteria* at phylum level were lower in CAG or IM with normal/SG as reference (all $q < 0.10$). Although no significant difference was found between DYS/GC and normal/SG groups, the abundances of *Actinobacteria*, *Bacteroidetes*, *Firmicutes* were found to be marginally higher in DYS/GC compared to IM or CAG (all $q < 0.10$). In addition, 17 non-*Helicobacter* genera were found differentially distributed in the comparisons of any two lesion groups (all $q < 0.10$, Supplementary Table 2).

Because of the different distribution of *H.pylori* in various lesion groups, we further analyzed the associations between the 17 non-*Helicobacter* candidate genera and gastric lesions stratified by *H.pylori* status. In *H.pylori* positive subjects, 14 genera were found in lower abundances in CAG or IM compared to normal/SG, and 9 genera were found in higher amounts in DYS/GC compared to IM (all $P < 0.05$, **Table 1**). In *H.pylori* negative subjects, no significant difference was found in various groups (all $P > 0.05$, Supplementary Table 3).

Prospective validation of the candidate genera associated with gastric lesions

Helicobacter and the 17 non-*Helicobacter* genera associated with gastric lesions were preliminarily validated in progression and non-progression subjects based on our 4-year endoscopic follow-up cohort. *Bacillus* was found to be more abundant in the initial biopsies before

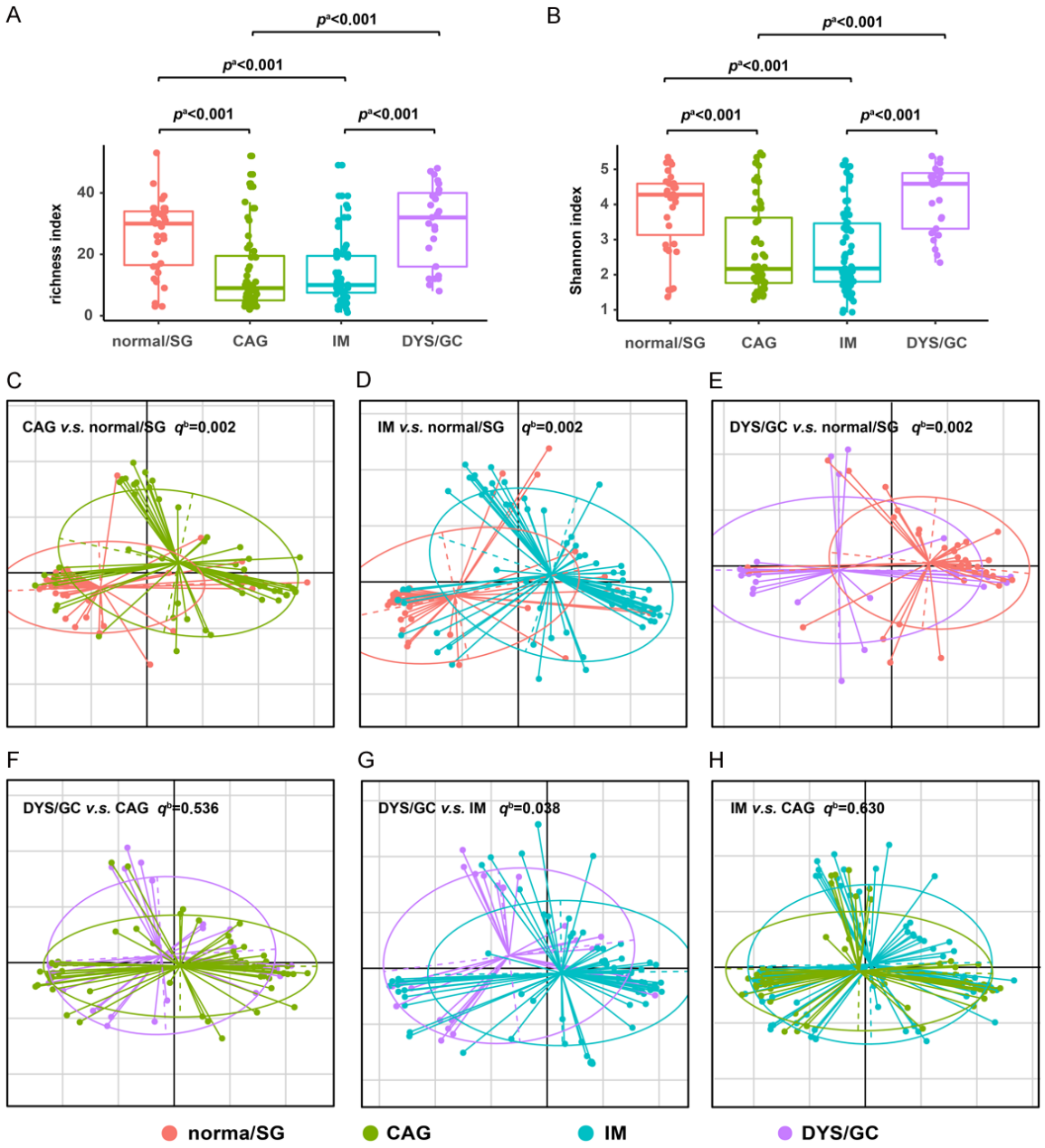

Figure 2. Microbial diversity and community structure in various gastric lesions. Boxplots presenting (A) richness index and (B) Shannon index decreased from normal/SG to CAG and IM, and increased in DYS/GC compared to CAG and IM. Microbial community structure comparisons in various gastric lesions showing significant differences (C) between normal/SG and CAG, (D) between normal/SG and IM, (E) between normal/SG and DYS/GC, and (G) between DYS/GC and IM, respectively. While no significant difference was found in microbial community structure (F) between DYS/GC and CAG, and (H) between IM and CAG. [a]Unconditional logistic regression adjusted for age, sex, smoking and alcohol consumption status. [b]PERMANOVA test. CAG, chronic atrophic gastritis; DYS, dysplasia; GC, gastric cancer; IM, intestinal metaplasia; SG, superficial gastritis.

progression compared to non-progression controls, P = 0.040. When comparing follow-up biopsies, *Helicobacter* abundance was lower and 9 non-*Helicobacter* genera were enriched in the progression compared to non-progression group (all P < 0.05, **Table 2**).

The 10 differential genera were compared between 21 progression-to-DYS/GC subjects and non-progression controls. The remarkable decline of *Helicobacter* and enrichment of *Bacillus*, *Capnocytophaga*, and *Prevotella* were observed after progression-to-DYS/GC com-

Table 1. Significantly altered non-*Helicobacter* genera in *H.pylori* positive subjects with various gastric lesions

Taxa	normal/SG[a]	CAG[a]	IM[a]	DYS/GC[a]	CAG *v.s.* normal/SG		IM *v.s.* normal/SG		DYS/GC *v.s.* normal/SG		DYS/GC *v.s.* CAG		DYS/GC *v.s.* IM	
	n = 26	n = 48	n = 63	n = 15	OR[b]	*p* value[b]	OR[b]	*p* value[b]	OR[b]	*p* value[b]	OR[b]	*p* value[b]	OR[b]	*p* value[b]
g__Acinetobacter	0.16%	0.01%	0.01%	0.24%	1.05	0.584	0.66	0.269	1.75	0.422	1.05	0.656	4.79	0.007
g__Actinomyces	0.14%	0.03%	0.04%	0.12%	0.07	0.015	0.04	0.005	0.21	0.252	10.82	0.099	19.90	0.032
g__Bacillus	0.02%	0.00%	0.00%	0.06%	0.01	0.043	0.01	0.020	0.74	0.802	> 999.99	0.020	505.92	0.032
g__Campylobacter	0.13%	0.01%	0.01%	0.08%	0.03	0.015	< 0.001	0.001	0.99	0.996	14.10	0.094	825.20	0.024
g__Capnocytophaga	0.13%	0.02%	0.02%	0.08%	0.21	0.022	0.17	0.011	0.36	0.266	1.15	0.890	4.47	0.177
g__Fusobacterium	1.50%	0.17%	0.20%	0.34%	0.61	0.003	0.49	0.001	0.52	0.099	0.99	0.961	1.19	0.579
g__Granulicatella	0.42%	0.05%	0.09%	0.15%	0.26	0.013	0.52	0.110	0.58	0.583	1.51	0.578	1.28	0.627
g__Neisseria	4.32%	0.37%	0.40%	1.52%	0.88	0.009	0.81	0.002	0.86	0.119	1.01	0.822	1.22	0.038
g__Peptostreptococcus	0.08%	0.02%	0.01%	0.05%	0.03	0.006	0.01	0.006	0.22	0.328	8.97	0.169	39.44	0.029
g__Porphyromonas	1.06%	0.12%	0.15%	0.40%	0.58	0.007	0.46	0.004	0.55	0.149	0.96	0.884	1.70	0.187
g__Prevotella	3.13%	0.45%	0.40%	0.87%	0.84	0.025	0.72	0.001	0.72	0.109	1.01	0.959	1.25	0.114
g__Pseudomonas	0.13%	0.01%	0.01%	0.32%	0.08	0.008	0.08	0.007	2.27	0.305	26.51	0.037	239.67	0.002
g__Ralstonia	0.01%	0.00%	0.00%	0.23%	0.25	0.108	0.83	0.725	6.96	0.093	11.24	0.026	6.00	0.007
g__Rothia	0.69%	0.07%	0.14%	0.30%	0.42	0.006	0.43	0.002	0.69	0.260	1.89	0.067	1.81	0.079
g__Sphingomonas	0.18%	0.01%	0.01%	0.48%	0.91	0.779	0.65	0.373	3.60	0.089	2.22	0.030	7.96	0.003
g__Streptococcus	2.74%	0.48%	0.86%	0.72%	0.75	0.003	0.91	0.047	0.94	0.533	1.21	0.045	1.07	0.244
g__Veillonella	0.42%	0.13%	0.24%	0.26%	0.50	0.017	0.54	0.020	0.84	0.651	1.45	0.320	1.75	0.107

[a]Relative abundance median of non-*Helicobacter* genera; [b]Unconditional logistic regression adjusted for age, sex, smoking and alcohol consumption status. CAG, chronic atrophic gastritis; DYS, dysplasia; GC, gastric cancer; IM, intestinal metaplasia; OR, odds ratio; SG, superficial gastritis.

Table 2. The validation of gastric lesion associated genera in progression and non-progression subjects

	Initial biopsies			Follow-up biopsies		
	Progression n = 26	Non-progression n = 29	*p* value[a]	Progression n = 28	Non-progression n = 26	*p* value[a]
g__Acinetobacter	0.44%	0.64%	0.601	0.41%	0.19%	0.188
g__Actinomyces	0.27%	0.29%	0.866	0.40%	0.16%	0.039
g__Bacillus	0.04%	0.01%	0.040	0.12%	0.00%	< 0.001
g__Campylobacter	0.12%	0.14%	0.649	0.19%	0.07%	0.023
g__Capnocytophaga	0.23%	0.21%	0.625	0.39%	0.13%	0.018
g__Fusobacterium	0.72%	0.64%	0.590	1.37%	0.43%	0.059
g__Granulicatella	0.22%	0.26%	0.866	0.56%	0.21%	0.045
g__Helicobacter	1.77%	3.82%	0.686	1.04%	50.31%	0.033
g__Neisseria	1.47%	2.31%	0.625	4.92%	1.54%	0.036
g__Peptostreptococcus	0.14%	0.09%	0.273	0.28%	0.11%	0.062
g__Porphyromonas	0.99%	0.53%	0.711	0.86%	0.54%	0.046
g__Prevotella	2.27%	1.91%	0.866	3.97%	1.36%	0.002
g__Pseudomonas	0.21%	0.35%	0.337	0.32%	0.41%	0.755
g__Ralstonia	0.43%	0.86%	0.074	0.76%	0.46%	0.324
g__Rothia	0.56%	0.72%	0.774	0.91%	0.31%	0.087
g__Sphingomonas	0.61%	0.85%	0.381	0.79%	0.55%	0.640
g__Streptococcus	1.56%	2.27%	0.893	2.55%	1.20%	0.062
g__Veillonella	0.76%	0.56%	0.438	1.38%	0.37%	0.019

[a]Mann-Whitney U test.

Table 3. The validation of gastric lesion associated genera in subjects who progressed to DYS/GC and non-progression controls

	Initial biopsies			Follow-up biopsies		
	Progression-to-DYS/GC n = 20	Non-progression n = 29	*p* value[a]	Progression-to-DYS/GC n = 21	Non-progression n = 26	*p* value[a]
g__Actinomyces	0.17%	0.29%	0.523	0.30%	0.16%	0.500
g__Bacillus	0.12%	0.01%	0.091	0.19%	0.00%	0.005
g__Campylobacter	0.07%	0.14%	0.349	0.14%	0.07%	0.082
g__Capnocytophaga	0.07%	0.21%	0.163	0.30%	0.13%	0.042
g__Granulicatella	0.20%	0.26%	0.802	0.25%	0.21%	0.193
g__Helicobacter	3.77%	3.82%	0.472	1.92%	50.31%	0.045
g__Neisseria	0.89%	2.31%	0.303	4.84%	1.54%	0.206
g__Porphyromonas	0.40%	0.53%	0.399	0.77%	0.54%	0.304
g__Prevotella	0.81%	1.91%	0.423	2.86%	1.36%	0.033
g__Veillonella	0.37%	0.56%	0.914	0.99%	0.37%	0.250

[a]Unconditional logistic regression adjusted for age and sex. DYS, dysplasia; GC, gastric cancer.

pared to non-progression controls (all $P < 0.05$, **Table 3**). In initial biopsies, no significant difference of genera was found before progression-to-DYS/GC compared to the controls, all $P > 0.05$.

Because of the different microbiota alterations in early (normal/SG to CAG or IM) and late (IM to DYS/GC) stages according to our cross-sectional comparisons, the four significant genera associated with advanced lesion progression

were validated in 11 pairs of biopsies before and after progression from IM to DYS/GC and 5 pairs of biopsies from non-progression IM controls. The abundances of *Bacillus*, *Capnocytophaga*, *Prevotella* were increased, while *Helicobacter* was decreased significantly in follow-up biopsies after the progression compared to the controls (all $P < 0.05$). Similar alteration trends were also found in initial biopsies before the progression compared to the controls, although marginal significance could only be found for *Bacillus* ($P = 0.069$, Supplementary Table 4).

Associations of microbial dysbiosis with advanced gastric lesion progression

We calculated Microbial Dysbiosis Index (MDI) with *Helicobacter*, *Bacillus*, *Capnocytophaga* and *Prevotella* according to the following formula: MDI = log (total abundance of genera increased after lesion progression/total abundance of genera decreased after lesion progression). MDI median was higher in DYS/GC compared to all of the other lesions (normal/SG/CAG/IM), $P < 0.001$ (**Figure 3A**). Although the MDI medians in the initial biopsies showed no difference between progression-to-DYS/GC and non-progression subjects, the increasing and decreasing trends of MDIs in two groups may lead to higher MDI in follow-up biopsies of progression-to-DYS/GC subjects compared to controls, $P = 0.027$ (**Figure 3B**). The same tendency of MDI can also be observed when comparing the progression-to-DYS/GC with non-progression IM subjects ($P = 0.009$, **Figure 3C**).

Discrimination of gastric lesion prognosis by specific genera

By combining the four significant genera (*Bacillus*, *Capnocytophaga*, *Helicobacer*, *Prevotella*) with age and sex, receiver operating characteristics (ROC) curve analysis showed outstanding performance in distinguishing follow-up subjects after lesion progression from non-progression controls, with an area under the curve (AUC) of 0.927 (**Figure 4A**). A similar AUC (0.941) was found to distinguish subjects after progression-to-DYS/GC from controls, all $P < 0.001$ (**Figure 4B**).

To investigate the predictive significance for lesion progression, ROC analysis was performed in the initial biopsies of progression and non-progression subjects. The panel including age, sex and the four genera did not differentiate between progression and non-progression subjects with an AUC of 0.639, $P = 0.077$ (**Figure 4C**). In contrast, when restricting to the progression-to-DYS/GC subjects or on the subset of progression subjects from IM to DYS/GC, the AUCs improved to 0.776 ($P = 0.001$) and 0.927 ($P = 0.008$), respectively (**Figure 4D**, **4E**).

Alterations of predicted microbiota functional capacity in gastric lesion progression

Microbial functional capacity prediction preliminarily found 47 up-regulated and 91 down-regulated metabolic pathways in CAG and IM subjects compared to normal/SG (all $q < 0.001$). When we compared DYS/GC with IM, 151 significantly up-regulated and 46 down-regulated pathways were detected (all $q < 0.001$, Supplementary Table 5).

From the 151 up-regulated pathways in DYS/GC, 96 candidates ($q < 0.001$, fold change > 5) were validated in paired biopsies from progression-to-DYS/GC (n = 11) and non-progression (n = 5) IM subjects. The most significant (all $P = 0.006$) pathways identified in follow-up biopsies after progression compared to non-progression controls included “protein digestion and absorption”, “lipoic acid metabolism”, “biosynthesis of type II polyketide products”, “biosynthesis of 12-, 14- and 16-membered macrolides”, “steroid biosynthesis”, “sesquiterpenoid and triterpenoid biosynthesis”, “serotonergic synapse”, “steroid degradation”, “adipocytokine signaling pathway”, “PPAR signaling pathway” and “DDT degradation” (Supplementary Table 6).

The exploration of nitrite metabolic related functional orthologues revealed over-representation of 24 proteins in DYS/GC compared to IM (all $q < 0.001$, Supplementary Table 7). Prospective validation confirmed increases of 12 proteins after progression compared to non-progression IM subjects ($P < 0.05$) including nitrite reductase, nitric oxide dioxygenase, nitrate/nitrite response regulator NarL, nitrite reductase (NAD(P)H) subunits, nitric oxide-sensitive transcriptional repressor, nitrate/nitrite transporter, etc. Similarly, increasing nitric oxide dioxygenase was also observed in the intra-individual comparison between follow-up and initial biopsies of progression IM subjects, ($P = 0.010$, Supplementary Table 8).

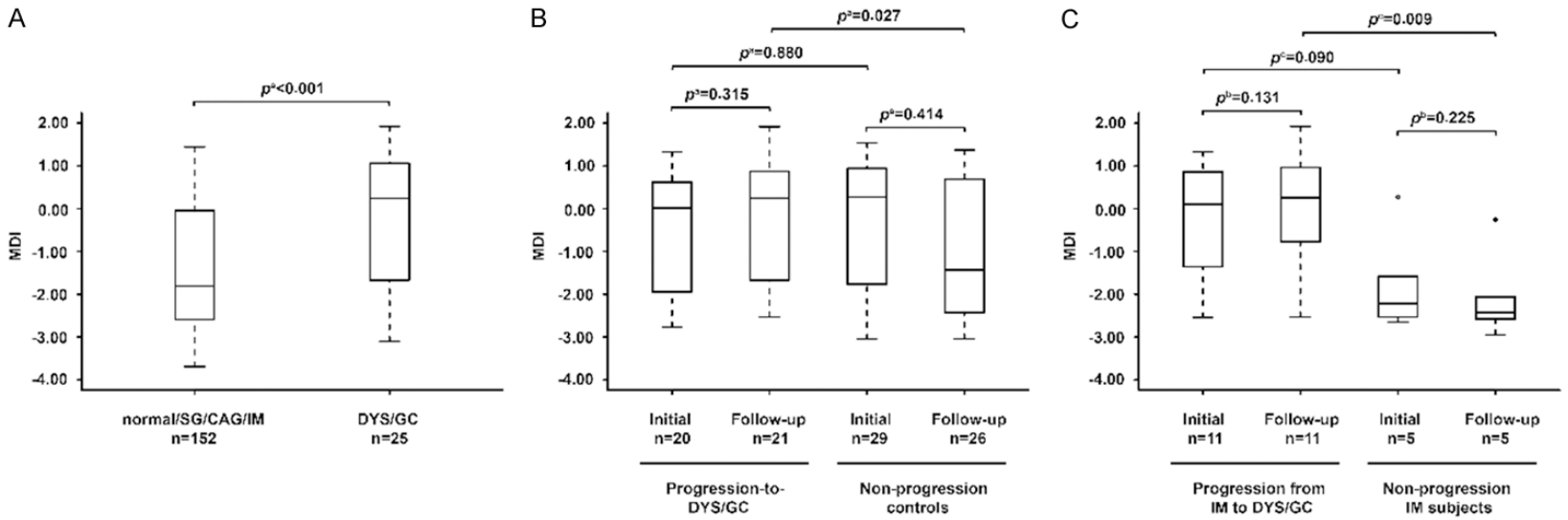

Figure 3. Microbial dysbiosis is associated with gastric lesion progression. Box plot showing (A) increased MDI in advanced lesions (DYS/GC) compared to all the other gastric lesions (normal/SG/CAG/IM); (B) increased MDI in follow-up biopsies of progression-to-DYS/GC subjects compared to non-progression controls; (C) increased MDI in follow-up biopsies of progression-to-DYS/GC compared to non-progression IM subjects. [a]Unconditional logistic regression adjusted for age, sex, smoking and alcohol consumption status. [b]Wilcoxon signed rank tests. [c]Mann-Whitney U test. CAG, chronic atrophic gastritis; DYS, dysplasia; GC, gastric cancer; IM, intestinal metaplasia; MDI, microbial dysbiosis index; SG, superficial gastritis.

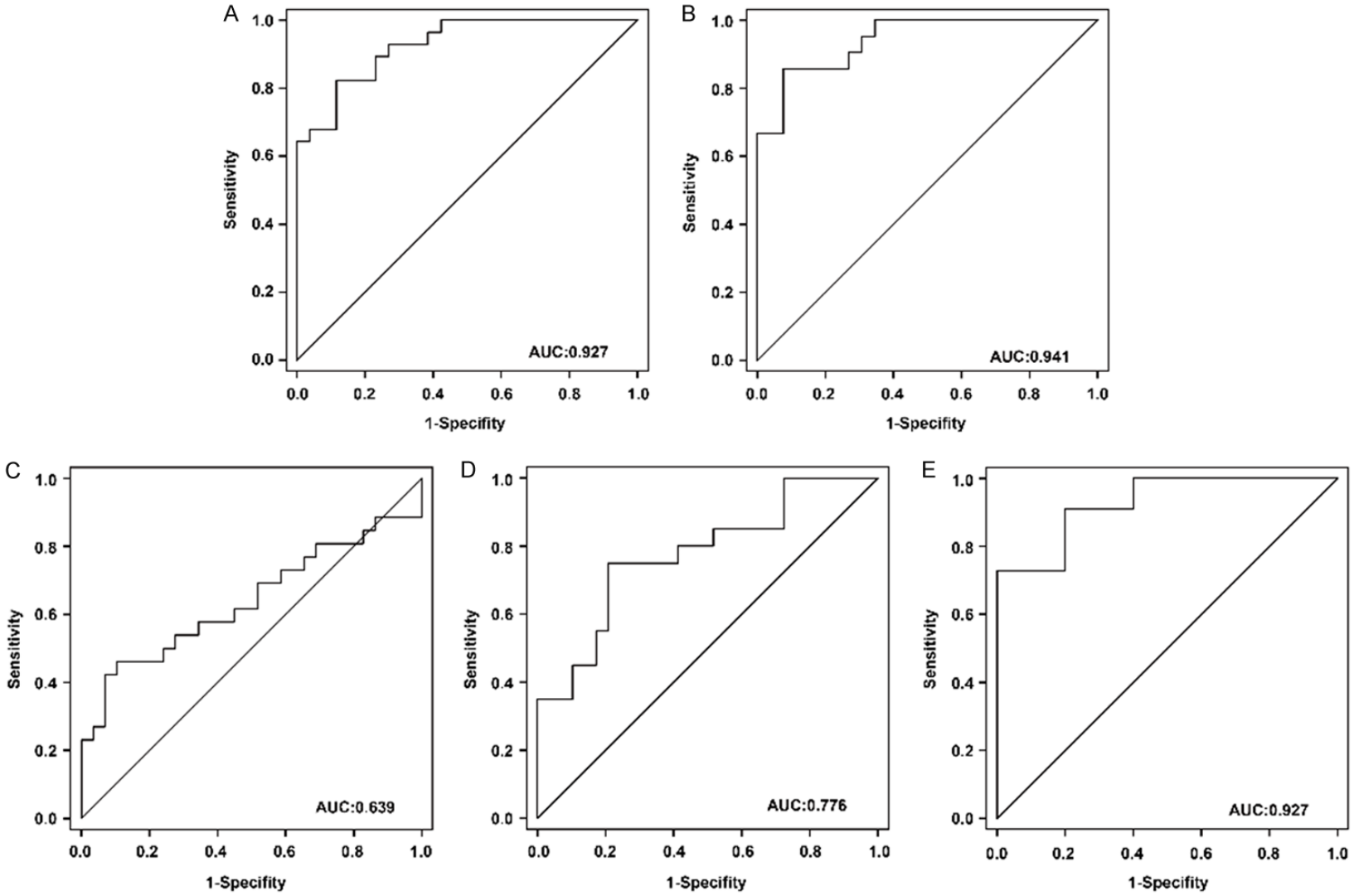

Figure 4. The discrimination of lesion progression by a panel with specific genera. Outstanding discriminatory performance was found by receiver operating characteristic (ROC) curve analysis using the panel of *Bacillus*, *Capnocytophaga*, *Prevotella*, *Helicobacter*, age and sex for (A) follow-up biopsies after lesion progression with an area under the curve (AUC) of 0.927 ($P < 0.001$) and for (B) follow-up biopsies after lesion progression-to-DYS/GC with an AUC of 0.941 ($P < 0.001$). Predictive significance was also found by ROC curve analysis using a panel including *Bacillus*, *Capnocytophaga*, *Prevotella*, *Helicobacter*, age and sex for (C) initial biopsies before lesion progression with an AUC of 0.639 ($P = 0.077$), for (D) initial biopsies before lesion progression-to-DYS/GC with an AUC of 0.776 ($P = 0.001$) and for (E) initial IM biopsies before lesion progression with an AUC of 0.927 ($P = 0.008$). AUC, area under the receiver-operating curve; DYS, dysplasia; GC, gastric cancer; IM, intestinal metaplasia; ROC, receiver operating characterist.

Discussion

Our population-based long-term follow-up study in a high-risk area of GC suggests that microbiota may act differently at different stages of GC development. This includes the initial driving of *H.pylori* and suppression of gastric commensals at the early stage, and depletion of *H.pylori* with enrichment of other genera at the late stage. A panel of four genera associated with gastric lesions may characterize microbial dysbiosis and help to discriminate and even predict the lesion progression.

Studies have revealed that low gastric microbial diversity induced by *H.pylori* infection may be associated with precancerous lesions and GC [6, 7, 21]. Our previous study found that successful eradication can restore microbial diversity to similar levels as that observed in *H.pylori* negative subjects [8]. Our results confirm a significantly lower microbial diversity and different microbial community structure with the greater abundances of *H.pylori* related taxa in CAG or IM compared to normal/SG. We also found an interesting restoration of gastric microbial diversity with decreased *H.pylori* related taxa from IM to DYS/GC, although the microbial community structure still showed a remarkable difference between DYS/GC and IM or normal/SG. Our results suggest that *H.pylori* may disturb gastric microbiota and initiate gastric lesion progression from an early stage, which may be further altered in later stages.

In the early stage of gastric lesions, *H.pylori* is dominant in the stomach due to its adaptability to acidic pH [22, 23]. The subsequent persistence of inflammation, and the loss of acid-secreting parietal cells after *H.pylori* infection, make the environment more suitable for colonization of other bacteria and contribute to lesion progression [24, 25]. In addition to *H.pylori* related taxa, our study further identified 57 other differential taxa including lower abundances of *Actinobacteria*, *Bacteroidetes*, *Firmicutes* and *Fusobacteria* in CAG or IM compared to normal/SG, and greater abundances of *Actinobacteria*, *Bacteroidetes* and *Firmicutes* in DYS/GC compared to IM. The occurrence of differential taxa supports different microbial mechanisms in early and late gastric carcinogenesis stages.

We found that 17 non-*Helicobacter* genera were significantly associated with gastric lesions only in *H.pylori* positive subjects, instead of in negative subjects. The 17 lesion associated genera also include the five strong co-excluding interactive genera of *Helicobacter* in advanced gastric lesions [8] in our previous intervention study. These consistent results suggest possible interactions between *Helicobacter* and the non-*Helicobacter* genera in gastric lesion development. Although non-*H.pylori* bacteria were reported to be associated with persistent inflammation and atrophy/IM in the stomach in a 1 year follow-up study after *H.pylori* eradication [26], our long-term prospective study further identified that the decrease of *Helicobacter* and concomitant increase of *Bacillus*, *Capnocytophaga*, *Prevotella* may be associated with lesion progression-to-DYS/GC. Our results confirm the hypothesis that the replacement of *H.pylori* by other bacteria may favor late-stage progression.

Although the functions of *H.pylori* in GC development have been well studied, the roles of other non-*H.pylori* bacteria have only recently started attracting attention [25]. *Prevotella* and *Capnocytophaga* are commensal in the oral cavity and associate with several cancers, including oral squamous cell carcinoma, lung cancer and GC [27]. They may act as opportunistic pathogens by producing inflammatory mediators, inducing chronic inflammation, and facilitating cell proliferation and oncogene activation [28, 29]. *Bacillus* genus, which was reported to be enriched in GC [30, 31], has been considered transient intestinal microbiota and can secrete a wide range of compounds with systemic effects on the host [32]. Further studies are needed to investigate whether the newly found non-*Helicobacter* genera can serve as independent risk factors for GC progression.

The MDI integrating *Helicobacter* and the three non-*Helicobacter* genera showed higher degree of microbial dysbiosis in DYS/GC compared to the benign conditions, which is in line with a previous retrospective study [7]. Furthermore, our long-term follow-up study allowed the prospective monitoring of microbial dysbiosis in gastric carcinogenesis. The increasing and decreasing trends of dysbiosis from initial status in progression and non-progression subjects cause significantly higher MDI after progression-to-DYS/GC (especially for initial IM subjects) compared to controls. These dynamic changes of microbial dysbiosis may help us to

better understand the role of microbiota in lesion progression.

Tentative explorations have been conducted in some retrospective studies using selected bacteria to detect GC [6, 7]. Our panel of age, sex and four specific genera can easily distinguish the advanced gastric lesions (especially DYS/GC) after long-term progression from controls. This panel also shows potential predictive values in the initial biopsies before the progression-to-DYS/GC with an AUC of 0.776. An even better predictive effect (AUC of 0.927) was achieved to predict the progression-to-DYS/GC in IM subjects, although the discrimination and prediction effects still need future studies in larger cohorts with longer follow-up.

Microbiota may be associated with energy metabolism, nutrients absorption and pathogens defense [33]. Microbial functional capacity prediction during carcinogenesis, especially at the late stage, can help us to better understand the possible mechanisms. The increases of protein and adipose metabolism pathway, PPAR signaling pathway, nitrite reductase and nitric oxide dioxygenase, validated both in the cross-sectional and prospective comparisons from IM to DYS, suggest an important role of microbial metabolic regulation at the critical stage of malignant transformation.

The strengths of our study lie in complementary design by combining a cross-sectional study with a 4-year follow-up study with paired biopsies before and after gastric lesion progression (especially to DYS/GC). The microbial alterations in different lesions can be proven dynamically during the natural evolution of GC, enabling the verification of their temporal relationships. The initial biopsies collected before lesion progression provide us with a unique opportunity to evaluate the predictive value of the microbial panel for the risk of gastric lesion progression, especially for DYS/GC lesions. However, our study has some limitations, including a small sample size of paired biopsies from progression subjects and a lack of validation in different populations. Furthermore, the possible microbial mechanisms in gastric lesion progression still needs further investigation.

In conclusion, our high-risk population-based study suggests that microbiota may play different roles at different stages of GC development, including initial *H.pylori* infection at early stage, and replacement of *H.pylori* by other GC-related genera at later stages. The panel of *Helicobacter*, *Bacillus*, *Capnocytophaga*, and *Prevotella* may help to discriminate advanced gastric lesions and even show predictive value for lesion progression. Our findings provide new clues for microbial mechanisms of *H.pylori*-associated carcinogenesis, although further larger and multicenter validations are still needed.

Acknowledgements

This study is supported by grants from National Key R&D Program of China [2018YFC1313100]; National Natural Science Foundation of China [81572811]; International (regional) Cooperation and Exchange Project [NSFC-DFG, 8186-1138041]; Beijing Natural Science Foundation [7182032]; German Federal Ministry of Education and Research (BMBF) [German Research Presence in Asia: 01DO17022] to MG; German Research Foundation (DFG) [SFB 1371] to MG; Beijing Municipal Administration of Hospitals' Ascent Plan [DFL20181102]; Beijing Talents foundation [2018000021223ZK01].

Disclosure of conflict of interest

None.

Address correspondence to: Drs. Kai-Feng Pan and Yang Zhang, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Cancer Epidemiology, Peking University Cancer Hospital & Institute, 52 Fucheng Rd, Haidian District, Beijing 100142, China. Tel: +86-10-88196701; Fax: +86-10-88122437; E-mail: pankaifeng2002@yahoo.com (KFP); yzhang76@sina.com (YZ)

References

[1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018; 68: 394-424.

[2] Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ and He J. Cancer statistics in China, 2015. CA Cancer J Clin 2016; 66: 115-132.

[3] 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: 6735-6740.
[4] Lofgren JL, Whary MT, Ge Z, Muthupalani S, Taylor NS, Mobley M, Potter A, Varro A, Eibach D, Suerbaum S, Wang TC and Fox JG. Lack of commensal flora in Helicobacter pylori-infected INS-GAS mice reduces gastritis and delays intraepithelial neoplasia. Gastroenterology 2011; 140: 210-220.
[5] Lertpiriyapong K, Whary MT, Muthupalani S, Lofgren JL, Gamazon ER, Feng Y, Ge Z, Wang TC and Fox JG. 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.
[6] Coker OO, Dai Z, Nie Y, Zhao G, Cao L, Nakatsu G, Wu WK, Wong SH, Chen Z, Sung JJY and Yu J. Mucosal microbiome dysbiosis in gastric carcinogenesis. Gut 2017; 67: 1024-1032.
[7] Ferreira RM, Pereira-Marques J, Pinto-Ribeiro I, Costa JL, Carneiro F, Machado JC and Figueiredo C. Gastric microbial community profiling reveals a dysbiotic cancer-associated microbiota. Gut 2018; 67: 226-236.
[8] Guo Y, Zhang Y, Gerhard M, Gao JJ, Mejias-Luque R, Zhang L, Vieth M, Ma JL, Bajbouj M, Suchanek S, Liu WD, Ulm K, Quante M, Li ZX, Zhou T, Schmid R, Classen M, Li WQ, You WC and Pan KF. Effect of Helicobacter pylori on gastrointestinal microbiota: a population-based study in Linqu, a high-risk area of gastric cancer. Gut 2020; 69: 1598-1607.
[9] You W. Precancerous gastric lesions in a population at high risk of stomach cancer. Cancer Res 1993; 53: 1317-1321.
[10] Dixon MF, Genta RM, Yardley JH and Correa P. Classification and grading of gastritis. The updated Sydney System. International Workshop on the Histopathology of GASTRITIS, Houston 1994. Am J Surg Pathol 1996; 20: 1161-1181.
[11] Lagkouvardos I, Joseph D, Kapfhammer M, Giritli S, Horn M, Haller D and Clavel T. IMNGS: a comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. Sci Rep 2016; 6: 33721.
[12] Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods 2013; 10: 996-998.
[13] Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 2010; 26: 2460-2461.
[14] Wang Q, Garrity GM, Tiedje JM and Cole JR. Naive bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 2007; 73: 5261-5267.
[15] Kim J, Kim N, Jo HJ, Park JH, Nam RH, Seok YJ, Kim YR, Kim JS, Kim JM, Kim JM, Lee DH and Jung HC. An appropriate cutoff value for determining the colonization of helicobacter pylori by the pyrosequencing method: comparison with conventional methods. Helicobacter 2015; 20: 370-380.
[16] Lagkouvardos I, Fischer S, Kumar N and Clavel T. Rhea: a transparent and modular R pipeline for microbial profiling based on 16S rRNA gene amplicons. PeerJ 2017; 5: e2836.
[17] Benjamini Y and Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser C Appl Stat 1995; 57: 289-300.
[18] Asshauer KP, Wemheuer B, Daniel R and Meinicke P. Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. Bioinformatics 2015; 31: 2882-2884.
[19] Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J and Glockner FO. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 2013; 41: D590-596.
[20] Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M and Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res 2014; 42: D199-205.
[21] Parsons BN, Ijaz UZ, D'Amore R, Burkitt MD, Eccles R, Lenzi L, Duckworth CA, Moore AR, Tiszlavicz L, Varro A, Hall N and Pritchard DM. Comparison of the human gastric microbiota in hypochlorhydric states arising as a result of Helicobacter pylori-induced atrophic gastritis, autoimmune atrophic gastritis and proton pump inhibitor use. PLoS Pathog 2017; 13: e1006653.
[22] Ansari S and Yamaoka Y. Survival of Helicobacter pylori in gastric acidic territory. Helicobacter 2017; 22: e12386.
[23] Schulz C, Schutte K, Koch N, Vilchez-Vargas R, Wos-Oxley ML, Oxley APA, Vital M, Malfertheiner P and Pieper DH. The active bacterial assemblages of the upper GI tract in individuals with and without Helicobacter infection. Gut 2018; 67: 216-225.
[24] Polk DB and Peek RM Jr. Helicobacter pylori: gastric cancer and beyond. Nat Rev Cancer 2010; 10: 403-414.
[25] Schulz C, Schutte K, Mayerle J and Malfertheiner P. The role of the gastric bacterial microbiome in gastric cancer: helicobacter pylori and beyond. Therap Adv Gastroenterol 2019; 12: 1756284819894062.
[26] Sung JJY, Coker OO, Chu E, Szeto CH, Luk STY, Lau HCH and Yu J. Gastric microbes associated with gastric inflammation, atrophy and intestinal metaplasia 1 year after Helicobacter pylori eradication. Gut 2020; 69: 1572-1580.

[27] Karpinski TM. Role of oral microbiota in cancer development. Microorganisms 2019; 7: E20.
[28] Bonatti H, Rossboth DW, Nachbaur D, Fille M, Aspock C, Hend I, Hourmont K, White L, Malnick H and Allerberger FJ. A series of infections due to Capnocytophaga spp in immunosuppressed and immunocompetent patients. Clin Microbiol Infect 2003; 9: 380-387.
[29] Larsen JM. The immune response to prevotella bacteria in chronic inflammatory disease. Immunology 2017; 151: 363-374.
[30] Chen XH, Wang A, Chu AN, Gong YH and Yuan Y. Mucosa-associated microbiota in gastric cancer tissues compared with non-cancer tissues. Front Microbiol 2019; 10: 1261.
[31] Liu X, Shao L, Liu X, Ji F, Mei Y, Cheng Y, Liu F, Yan C, Li L and Ling Z. Alterations of gastric mucosal microbiota across different stomach microhabitats in a cohort of 276 patients with gastric cancer. EBioMedicine 2019; 40: 336-348.
[32] Ilinskaya ON, Ulyanova VV, Yarullina DR and Gataullin IG. Secretome of intestinal bacilli: a natural guard against pathologies. Front Microbiol 2017; 8: 1666.
[33] Neish AS. Microbes in gastrointestinal health and disease. Gastroenterology 2009; 136: 65-80.

Supplementary Table 1. General characteristics of subjects with various gastric lesions

	Total	normal/SG	CAG		IM		DYS/GC	
	n = 179	n = 35	n = 52	*p* value[a]	n = 67	*p* value[b]	n = 25	*p* value[c]
Age, years (Mean ± SD)[d]	55.9 ± 8.3	53.8 ± 7.8	53.4 ± 9.3	0.835	58.5 ± 7.7	0.005	57.6 ± 6.4	0.044
Sex[e]				0.088		0.009		0.001
Male	105 (58.7%)	13 (37.1%)	29 (55.8%)		43 (64.2%)		20 (80.0%)	
Female	74 (41.3%)	22 (62.9%)	23 (44.2%)		24 (35.8%)		5 (20.0%)	
Smoking[e]				0.331		0.019		0.023
No	134 (74.9%)	31 (88.6%)	42 (80.8%)		45 (67.2%)		16 (64.0%)	
Yes	45 (25.1%)	4 (11.4%)	10 (19.2%)		22 (32.8%)		9 (36.0%)	
Alcohol consumption[e]				0.035		0.031		0.010
No	122 (68.2%)	30 (85.7%)	34 (65.4%)		44 (65.7%)		14 (56.0%)	
Yes	57 (31.8%)	5 (14.3%)	18 (34.6%)		23 (34.3%)		11 (44.0%)	
H. pylori infection[e]				< 0.001		< 0.001		0.241
Negative	27 (15.1%)	9 (25.7%)	4 (7.7%)		4 (6.0%)		10 (40.0%)	
Positive	152 (84.9%)	26 (74.3%)	48 (92.3%)		63 (94.0%)		15 (60.0%)	

[a]CAG group *v.s.* normal/SG group; [b]IM group *v.s.* normal/SG group; [c]DYS/GC group *v.s.* normal/SG group; [d]*t*-test; [e]χ^2 test; CAG, chronic atrophic gastritis; DYS, dysplasia; GC, gastric cancer; *H.pylori*, *Helicobacter pylori*; IM, intestinal metaplasia; SG, superficial gastritis.

Supplementary Table 2. Differentially distributed taxa among various gastric lesions

Taxa[a]		normal/SG[b]	CAG[b]	IM[b]	DYS/GC[b]	CAG *v.s.* normal/SG		IM *v.s.* normal/SG		DYS/GC *v.s.* normal/SG		DYS/GC *v.s.* CAG		DYS/GC *v.s.* IM	
		n = 35	n = 52	n = 67	n = 25	OR[c]	*q* value[c]	OR[c]	*q* value[c]	OR[c]	*q* value[c]	OR[c]	*q* value[c]	OR[c]	*q* value[c]
H. pylori related taxa															
Phylum	*p__Proteobacteria*	36.39%	91.04%	90.46%	30.99%	1.05	0.002	1.04	0.002	0.99	0.993	0.95	0.001	0.95	0.001
Class	*c__Epsilonproteobacteria*	3.74%	89.28%	88.63%	2.57%	1.03	0.002	1.03	0.002	1.00	0.993	0.97	0.006	0.97	0.001
Order	*o__Campylobacterales*	3.74%	89.28%	88.63%	2.57%	1.03	0.002	1.03	0.002	1.00	0.993	0.97	0.006	0.97	0.001
Family	*f__Helicobacteraceae*	3.45%	89.27%	88.59%	1.92%	1.03	0.002	1.03	0.002	1.00	0.993	0.97	0.001	0.97	0.001
Genus	*g__Helicobacter*	3.45%	89.27%	88.59%	1.92%	1.03	0.002	1.03	0.002	1.00	0.993	0.97	0.001	0.97	0.001
Other candidate taxa															
Phylum	*p__Fusobacteria*	2.56%	0.22%	0.32%	2.00%	0.69	0.064	0.69	0.052	1.06	0.993	1.34	0.332	1.42	0.131
Phylum	*p__Actinobacteria*	2.72%	0.43%	0.50%	2.39%	0.74	0.079	0.74	0.082	0.93	0.993	1.36	0.130	1.37	0.060
Phylum	*p__Firmicutes*	12.94%	2.52%	2.91%	12.41%	0.91	0.079	0.95	0.284	1.02	0.993	1.11	0.089	1.06	0.172
Phylum	*p__Bacteroidetes*	14.18%	1.85%	2.65%	11.25%	0.92	0.064	0.92	0.100	1.00	0.993	1.08	0.251	1.11	0.046
Class	*c__Negativicutes*	1.31%	0.36%	0.39%	1.83%	0.51	0.064	0.58	0.102	1.21	0.993	2.36	0.076	1.98	0.039
Class	*c__Clostridia*	2.74%	0.68%	0.77%	1.94%	0.68	0.087	0.86	0.515	0.90	0.993	1.43	0.258	1.22	0.330
Class	*c__Fusobacteriia*	2.56%	0.22%	0.32%	2.00%	0.69	0.064	0.69	0.052	1.06	0.993	1.34	0.332	1.42	0.131
Class	*c__Actinobacteria*	2.72%	0.43%	0.50%	2.39%	0.74	0.079	0.74	0.089	0.93	0.993	1.36	0.130	1.37	0.061
Class	*c__Betaproteobacteria*	7.79%	0.76%	0.81%	7.21%	0.89	0.064	0.91	0.121	0.93	0.993	1.10	0.402	1.07	0.393
Class	*c__Bacteroidia*	12.39%	1.60%	2.63%	7.75%	0.92	0.064	0.91	0.102	0.98	0.993	1.07	0.466	1.09	0.157
Class	*c__Flavobacteriia*	0.52%	0.06%	0.05%	1.15%	0.43	0.247	0.53	0.324	1.53	0.993	5.90	0.023	4.91	0.012
Class	*c__Alphaproteobacteria*	1.20%	0.09%	0.08%	2.65%	0.94	0.934	0.82	0.544	1.18	0.993	1.15	0.829	1.59	0.032
Order	*o__Burkholderiales*	0.96%	0.10%	0.10%	1.43%	0.45	0.088	1.01	0.990	1.35	0.993	2.92	0.035	1.01	0.990
Order	*o__Selenomonadales*	1.31%	0.36%	0.39%	1.83%	0.51	0.064	0.58	0.102	1.21	0.993	2.36	0.076	1.98	0.039
Order	*o__Clostridiales*	2.74%	0.68%	0.77%	1.94%	0.68	0.087	0.86	0.515	0.90	0.993	1.43	0.258	1.22	0.330
Order	*o__Fusobacteriales*	2.56%	0.22%	0.32%	2.00%	0.69	0.064	0.69	0.052	1.06	0.993	1.34	0.332	1.42	0.131
Order	*o__Actinomycetales*	2.45%	0.30%	0.41%	1.92%	0.71	0.079	0.72	0.100	0.93	0.993	1.45	0.107	1.43	0.054
Order	*o__Neisseriales*	5.41%	0.58%	0.62%	5.09%	0.90	0.088	0.85	0.052	0.91	0.993	1.07	0.759	1.19	0.081
Order	*o__Bacteroidales*	12.39%	1.60%	2.63%	7.75%	0.92	0.064	0.91	0.102	0.98	0.993	1.07	0.466	1.09	0.157
Order	*o__Bacillales*	1.09%	0.24%	0.29%	1.36%	0.75	0.756	1.03	0.990	1.70	0.993	1.92	0.082	1.05	0.961
Order	*o__Flavobacteriales*	0.52%	0.06%	0.05%	1.15%	0.43	0.247	0.53	0.324	1.53	0.993	5.90	0.023	4.91	0.012
Order	*o__Sphingomonadales*	0.38%	0.02%	0.01%	0.92%	0.87	0.934	0.87	0.947	2.14	0.993	1.71	0.235	2.32	0.044
Order	*o__Rhizobiales*	0.17%	0.01%	0.02%	0.36%	1.01	0.992	0.17	0.121	1.15	0.993	0.92	0.991	15.17	0.018
Family	*f__Campylobacteraceae*	0.26%	0.01%	0.01%	0.17%	0.05	0.064	0.02	0.052	3.00	0.993	24.75	0.121	58.17	0.052
Family	*f__Peptostreptococcaceae*	0.27%	0.04%	0.04%	0.41%	0.12	0.064	0.18	0.190	1.14	0.993	9.50	0.148	8.32	0.081

Family	*f__Carnobacteriaceae*	0.59%	0.06%	0.10%	0.45%	0.22	0.064	0.42	0.209	1.73	0.993	4.33	0.114	2.32	0.137
Family	*f__Micrococcaceae*	1.34%	0.12%	0.16%	0.59%	0.49	0.064	0.54	0.052	0.88	0.993	2.36	0.058	1.87	0.052
Family	*f__Veillonellaceae*	1.30%	0.35%	0.39%	1.82%	0.49	0.064	0.57	0.102	1.23	0.993	2.44	0.075	2.07	0.036
Family	*f__Fusobacteriaceae*	1.86%	0.18%	0.26%	1.35%	0.61	0.064	0.62	0.052	0.95	0.993	1.56	0.187	1.57	0.090
Family	*f__Porphyromonadaceae*	2.23%	0.37%	0.49%	1.15%	0.67	0.064	0.86	0.512	0.76	0.993	1.15	0.991	1.04	0.990
Family	*f__Streptococcaceae*	3.42%	0.56%	0.97%	2.23%	0.76	0.064	0.92	0.399	1.05	0.993	1.31	0.070	1.08	0.378
Family	*f__Prevotellaceae*	6.47%	0.98%	0.91%	5.36%	0.89	0.090	0.85	0.052	1.03	0.993	1.12	0.289	1.17	0.046
Family	*f__Neisseriaceae*	5.41%	0.58%	0.62%	5.09%	0.90	0.088	0.85	0.052	0.91	0.993	1.07	0.759	1.19	0.081
Family	*f__Flavobacteriaceae*	0.52%	0.06%	0.05%	1.15%	0.43	0.256	0.53	0.333	1.53	0.993	5.91	0.023	4.92	0.012
Family	*f__Burkholderiaceae*	0.16%	0.00%	0.00%	0.75%	0.27	0.256	1.05	0.963	3.31	0.993	11.69	0.016	0.99	0.990
Family	*f__Actinomycetaceae*	0.22%	0.03%	0.05%	0.45%	0.11	0.172	0.10	0.121	2.01	0.993	21.23	0.053	29.01	0.005
Family	*f__Bacillaceae 1*	0.03%	0.00%	0.01%	0.19%	0.12	0.349	0.04	0.209	1.57	0.993	365.03	0.047	969.51	0.025
Family	*f__Sphingomonadaceae*	0.37%	0.02%	0.01%	0.92%	0.87	0.934	0.82	0.923	2.24	0.993	1.73	0.235	2.62	0.024
Family	*f__Pseudomonadaceae*	0.18%	0.01%	0.01%	0.35%	0.21	0.190	0.23	0.209	1.03	0.993	5.06	0.212	9.65	0.074
Family	*f__Comamonadaceae*	0.17%	0.02%	0.01%	0.13%	0.20	0.347	0.12	0.131	0.56	0.993	3.99	0.466	10.05	0.074
Genus	*g__Peptostreptococcus*	0.15%	0.02%	0.01%	0.20%	0.03	0.064	0.14	0.306	1.22	0.993	71.25	0.077	12.12	0.108
Genus	*g__Campylobacter*	0.24%	0.01%	0.01%	0.14%	0.04	0.068	0.01	0.052	3.18	0.993	31.84	0.130	82.31	0.052
Genus	*g__Actinomyces*	0.20%	0.03%	0.05%	0.34%	0.05	0.079	0.10	0.131	1.62	0.993	94.37	0.029	25.16	0.015
Genus	*g__Granulicatella*	0.55%	0.06%	0.10%	0.37%	0.19	0.064	0.42	0.230	1.65	0.993	4.09	0.146	2.12	0.210
Genus	*g__Rothia*	0.71%	0.09%	0.14%	0.56%	0.40	0.064	0.46	0.052	0.87	0.993	2.62	0.072	2.06	0.046
Genus	*g__Veillonella*	0.60%	0.23%	0.26%	1.27%	0.47	0.088	0.57	0.209	1.38	0.993	2.77	0.089	2.24	0.046
Genus	*g__Fusobacterium*	1.86%	0.18%	0.26%	1.35%	0.61	0.064	0.62	0.052	0.95	0.993	1.55	0.194	1.56	0.095
Genus	*g__Porphyromonas*	1.43%	0.15%	0.18%	0.80%	0.61	0.079	0.53	0.073	0.79	0.993	1.28	0.914	1.82	0.165
Genus	*g__Streptococcus*	3.30%	0.56%	0.95%	2.23%	0.76	0.064	0.92	0.432	1.05	0.993	1.30	0.071	1.08	0.386
Genus	*g__Neisseria*	5.16%	0.52%	0.55%	5.00%	0.90	0.112	0.84	0.052	0.91	0.993	1.07	0.829	1.20	0.081
Genus	*g__Ralstonia*	0.03%	0.00%	0.00%	0.58%	0.33	0.391	1.08	0.963	9.33	0.731	10.52	0.027	1.05	0.990
Genus	*g__Bacillus*	0.01%	0.00%	0.00%	0.18%	0.03	0.322	0.05	0.228	1.36	0.993	> 999.99	0.047	595.67	0.046
Genus	*g__Prevotella*	3.68%	0.54%	0.53%	3.33%	0.86	0.258	0.74	0.052	1.07	0.993	1.18	0.332	1.38	0.025
Genus	*g__Acinetobacter*	0.24%	0.02%	0.01%	0.48%	1.04	0.934	0.61	0.545	1.10	0.993	1.03	0.991	3.30	0.046
Genus	*g__Capnocytophaga*	0.17%	0.03%	0.03%	0.30%	0.21	0.185	0.30	0.218	0.81	0.993	7.22	0.180	5.80	0.088
Genus	*g__Sphingomonas*	0.21%	0.01%	0.01%	0.78%	0.79	0.934	0.52	0.599	3.01	0.993	2.34	0.146	7.25	0.001
Genus	*g__Pseudomonas*	0.15%	0.01%	0.01%	0.31%	0.20	0.202	0.15	0.158	0.99	0.994	4.78	0.280	16.32	0.058

[a]Candidate taxa were preliminarily selected with q < 0.10 after multiple-testing FDR correction and the altered taxa at least in two lesion groups comparison with relative abundance median > 0.1% were listed. [b]Relative abundance median of taxa. [c]Unconditional logistic regression adjusted for age, sex, smoking, and alcohol consumption status and *q* values were used after adjustment for multiple comparison by FDR. CAG, chronic atrophic gastritis; DYS, dysplasia; GC, gastric cancer; *H.pylori*, *Helicobacter pylori*; IM, intestinal metaplasia; OR, odds ratio; SG, superficial gastritis.

Supplementary Table 3. The distributions of 17 non-*Helicobacter* genera in *H.pylori* negative subjects with various gastric lesions

Taxa	normal/SG[a]	CAG[a]	IM[a]	DYS/GC[a]	CAG *v.s.* normal/SG		IM *v.s.* normal/SG		DYS/GC *v.s.* normal/SG		DYS/GC *v.s.* CAG		DYS/GC *v.s.* IM	
	n = 9	n = 4	n = 4	n = 10	OR[b]	*p* value[b]	OR[b]	*p* value[b]	OR[b]	*p* value[b]	OR[b]	*p* value[b]	OR[b]	*p* value[b]
g__Acinetobacter	0.60%	1.22%	0.67%	0.61%	0.64	0.495	0.85	0.891	0.47	0.332	0.32	0.339	1.53	0.787
g__Actinomyces	0.36%	0.35%	0.72%	0.82%	0.08	0.458	> 999.99	0.288	21.86	0.281	> 999.99	0.566	10.18	0.185
g__Bacillus	0.01%	0.00%	0.05%	0.23%	22.03	0.809	0.57	0.796	4.11	0.431	> 999.99	0.239	> 999.99	0.270
g__Campylobacter	0.45%	0.35%	0.50%	0.42%	< 0.001	0.439	0.38	0.670	2.60	0.502	22.15	0.280	6.86	0.323
g__Capnocytophaga	0.23%	0.23%	1.10%	0.74%	0.51	0.741	177.50	0.185	1.30	0.870	23.57	0.204	< 0.001	0.273
g__Fusobacterium	2.45%	2.16%	4.04%	2.05%	0.65	0.526	0.98	0.942	1.13	0.578	2.97	0.176	0.01	0.274
g__Granulicatella	1.17%	0.36%	0.69%	0.84%	0.09	0.361	0.11	0.282	2.78	0.186	13.66	0.292	8.07	0.294
g__Neisseria	9.02%	5.47%	6.89%	6.10%	0.95	0.585	0.99	0.921	0.90	0.190	0.98	0.859	0.70	0.622
g__Peptostreptococcus	0.29%	0.24%	0.49%	0.53%	0.01	0.279	> 999.99	0.249	58.50	0.122	125.27	0.264	1.12	0.940
g__Porphyromonas	2.42%	3.13%	2.47%	1.43%	6.02	0.211	0.88	0.710	0.83	0.642	0.94	0.928	0.54	0.651
g__Prevotella	4.99%	8.41%	5.99%	6.74%	> 999.99	0.528	1.05	0.892	1.30	0.276	1.12	0.726	1.30	0.444
g__Pseudomonas	0.41%	0.79%	0.41%	0.28%	0.67	0.562	1.53	0.637	0.33	0.408	< 0.001	0.246	< 0.001	0.504
g__Ralstonia	0.62%	0.98%	0.78%	0.95%	4.86	0.453	1.46	0.252	5.96	0.164	1.49	0.879	0.59	0.455
g__Rothia	1.32%	0.73%	1.91%	1.93%	4.19	0.514	2.28	0.449	0.94	0.896	4.14	0.218	0.85	0.829
g__Sphingomonas	0.78%	0.76%	0.77%	1.02%	0.57	0.599	0.40	0.642	1.34	0.728	0.09	0.186	11.63	0.259
g__Streptococcus	3.68%	5.02%	8.36%	5.89%	1.12	0.621	1.03	0.874	1.21	0.191	1.34	0.222	0.76	0.197
g__Veillonella	1.73%	1.34%	1.97%	2.03%	3.39	0.431	1.71	0.654	3.35	0.071	173.26	0.289	4.68	0.190

[a]Relative abundance median of non-*Helicobacter* genera. [b]Unconditional logistic regression adjusted for age, sex, smoking and alcohol consumption status. CAG, chronic atrophic gastritis; DYS, dysplasia; GC, gastric cancer; *H.pylori*, *Helicobacter pylori*; IM, intestinal metaplasia; OR, odds ratio; SG, superficial gastritis.

Supplementary Table 4. The validation of advanced lesion progression associated specific genera in progression and non-progression IM subjects

	Progression IM subjects			Non-progression IM subjects			*p* value[a,d]	*p* value[b,d]
	Initial biopsies n = 11	Follow-up biopsies n = 11	*p* value[c]	Initial biopsies n = 5	Follow-up biopsies n = 5	*p* value[c]		
g__Bacillus	0.03%	0.19%	0.328	0.00%	0.00%	0.893	0.069	0.005
g__Capnocytophaga	0.41%	0.30%	0.286	0.15%	0.03%	0.225	0.180	0.005
g__Helicobacter	2.82%	1.30%	0.328	62.57%	80.23%	0.177	0.145	0.009
g__Prevotella	3.27%	4.17%	0.594	0.32%	0.27%	0.225	0.221	0.027

[a]The comparison of initial biopsies between progression and non-progression IM subjects. [b]The comparison of follow-up biopsies between progression and non-progression IM subjects. [c]The self-comparison of paired initial and follow-up biopsies by Wilcoxon signed-rank test. [d]Mann-Whitney U test. IM, intestinal metaplasia.

Supplementary Table 5. Alterations of predicted metabolic pathways among various gastric lesions

	normal/SG[c]	CAG[c]	IM[c]	DYS/GC[c]	CAG *v.s.* normal/SG		IM *v.s.* normal/SG		DYS/GC *v.s.* normal/SG		DYS/GC *v.s.* CAG		DYS/GC *v.s.* IM	
	n = 35	n = 52	n = 67	n = 25	FC[d]	*q* value[e]	FC[d]	*q* value[e]	FC[d]	*q* value[e]	FC[d]	*q* value[e]	FC[d]	*q* value[e]
Up-regulated in CAG and IM compared to normal/SG[a]														
ko04975; Fat digestion and absorption	9.21E-06	3.40E-05	3.36E-05	6.04E-06	3.70	< 0.001	3.65	< 0.001	0.66	0.998	0.18	0.002	0.18	< 0.001
ko05120; Epithelial cell signaling in Helicobacter pylori infection	1.87E-02	6.88E-02	6.80E-02	1.12E-02	3.69	< 0.001	3.64	< 0.001	0.60	0.998	0.16	0.002	0.17	< 0.001
ko02040; Flagellar assembly	1.26E-02	3.74E-02	3.69E-02	1.21E-02	2.98	< 0.001	2.94	< 0.001	0.97	0.998	0.32	0.002	0.33	< 0.001
ko00592; alpha-Linolenic acid metabolism	3.06E-04	7.15E-04	7.09E-04	2.94E-04	2.34	< 0.001	2.32	< 0.001	0.96	0.998	0.41	0.002	0.41	< 0.001
ko03015; mRNA surveillance pathway	4.01E-05	9.06E-05	8.97E-05	3.04E-05	2.26	< 0.001	2.24	< 0.001	0.76	0.998	0.34	0.002	0.34	< 0.001
ko04260; Cardiac muscle contraction	1.79E-04	3.85E-04	3.81E-04	1.64E-04	2.15	< 0.001	2.13	< 0.001	0.91	0.998	0.43	0.002	0.43	< 0.001
ko05014; Amyotrophic lateral sclerosis (ALS)	3.09E-04	6.24E-04	6.21E-04	3.36E-04	2.02	< 0.001	2.01	< 0.001	1.09	0.998	0.54	0.002	0.54	< 0.001
ko05012; Parkinsons disease	2.25E-04	4.50E-04	4.46E-04	2.14E-04	2.00	< 0.001	1.98	< 0.001	0.95	0.998	0.48	0.002	0.48	< 0.001
ko05134; Legionellosis	3.82E-03	7.51E-03	7.44E-03	3.40E-03	1.97	< 0.001	1.95	< 0.001	0.89	0.998	0.45	0.002	0.46	< 0.001
ko00633; Nitrotoluene degradation	2.03E-03	3.92E-03	3.89E-03	1.92E-03	1.93	< 0.001	1.92	< 0.001	0.95	0.998	0.49	0.002	0.49	< 0.001
Down-regulated in CAG or IM compared to normal/SG[a]														
ko04622; RIG-I-like receptor signaling pathway	7.74E-05	1.78E-06	3.27E-06	7.99E-05	0.02	< 0.001	0.04	< 0.001	1.03	0.998	44.82	0.002	24.48	< 0.001
ko00364; Fluorobenzoate degradation	2.05E-04	4.97E-06	6.56E-06	2.66E-04	0.02	< 0.001	0.03	< 0.001	1.30	0.998	53.47	0.005	40.52	< 0.001
ko04011; MAPK signaling pathway - yeast	2.94E-04	7.29E-06	1.20E-05	3.48E-04	0.02	< 0.001	0.04	< 0.001	1.18	0.998	47.69	0.002	28.91	< 0.001
ko05203; Viral carcinogenesis	1.80E-04	4.61E-06	8.37E-06	2.30E-04	0.03	< 0.001	0.05	< 0.001	1.28	0.998	49.93	0.002	27.48	< 0.001
ko04930; Type II diabetes mellitus	1.67E-04	4.45E-06	7.89E-06	2.09E-04	0.03	< 0.001	0.05	< 0.001	1.25	0.998	47.04	0.002	26.50	< 0.001
ko05131; Shigellosis	1.77E-04	4.89E-06	5.71E-06	1.68E-04	0.03	< 0.001	0.03	< 0.001	0.95	0.998	34.39	0.010	29.46	0.002
ko03022; Basal transcription factors	8.55E-05	2.42E-06	3.81E-06	9.61E-05	0.03	< 0.001	0.04	< 0.001	1.12	0.998	39.79	0.002	25.20	< 0.001
ko05020; Prion diseases	3.11E-05	8.92E-07	1.16E-06	3.96E-05	0.03	< 0.001	0.04	< 0.001	1.27	0.998	44.37	0.003	33.99	< 0.001
ko05110; Vibrio cholerae infection	1.78E-04	5.20E-06	6.58E-06	1.83E-04	0.03	< 0.001	0.04	< 0.001	1.03	0.998	35.23	0.012	27.82	0.001
ko00785; Lipoic acid metabolism	1.33E-03	3.92E-05	5.99E-05	1.41E-03	0.03	< 0.001	0.05	< 0.001	1.06	0.998	35.93	0.002	23.53	< 0.001
Up-regulated in DYS/GC compared to IM[b]														
ko04080; Neuroactive ligand-receptor interaction	5.36E-07	1.36E-08	1.85E-08	9.99E-07	0.03	0.749	0.03	0.022	1.86	0.998	73.32	0.228	53.89	< 0.001
ko05210; Colorectal cancer	2.89E-06	8.24E-08	7.14E-08	3.69E-06	0.03	0.151	0.02	0.018	1.28	0.998	44.80	0.018	51.70	< 0.001
ko05416; Viral myocarditis	2.89E-06	8.24E-08	7.14E-08	3.69E-06	0.03	0.151	0.02	0.018	1.28	0.998	44.80	0.018	51.70	< 0.001
ko01057; Biosynthesis of type II polyketide products	9.14E-05	2.07E-06	2.44E-06	1.21E-04	0.02	0.001	0.03	< 0.001	1.33	0.998	58.64	0.004	49.77	< 0.001
ko00909; Sesquiterpenoid and triterpenoid biosynthesis	1.40E-04	3.13E-06	3.21E-06	1.54E-04	0.02	0.002	0.02	< 0.001	1.10	0.998	49.37	0.010	48.10	< 0.001
ko00100; Steroid biosynthesis	2.91E-05	6.51E-07	7.55E-07	3.58E-05	0.02	0.014	0.03	0.002	1.23	0.998	55.08	0.023	47.49	< 0.001
ko04916; Melanogenesis	3.18E-06	7.27E-08	8.51E-08	4.04E-06	0.02	0.005	0.03	0.001	1.27	0.998	55.57	0.020	47.47	< 0.001
ko00902; Monoterpenoid biosynthesis	1.08E-05	2.20E-07	2.80E-07	1.33E-05	0.02	0.064	0.03	0.009	1.23	0.998	60.39	0.019	47.45	< 0.001
ko00232; Caffeine metabolism	4.08E-05	1.04E-06	1.30E-06	5.93E-05	0.03	0.005	0.03	0.002	1.45	0.998	56.81	0.010	45.43	< 0.001
ko00522; Biosynthesis of 12-, 14- and 16-membered macrolides	1.86E-04	4.39E-06	5.37E-06	2.42E-04	0.02	0.001	0.03	< 0.001	1.30	0.998	55.19	0.005	45.08	< 0.001
Down-regulated in DYS/GC compared to IM[b]														
ko05120; Epithelial cell signaling in Helicobacter pylori infection	1.87E-02	6.88E-02	6.80E-02	1.12E-02	3.69	< 0.001	3.64	< 0.001	0.60	0.998	0.16	0.002	0.17	< 0.001
ko04975; Fat digestion and absorption	9.21E-06	3.40E-05	3.36E-05	6.04E-06	3.70	< 0.001	3.65	< 0.001	0.66	0.998	0.18	0.002	0.18	< 0.001
ko02040; Flagellar assembly	1.26E-02	3.74E-02	3.69E-02	1.21E-02	2.98	< 0.001	2.94	< 0.001	0.97	0.998	0.32	0.002	0.33	< 0.001

ko03015; mRNA surveillance pathway	4.01E-05	9.06E-05	8.97E-05	3.04E-05	2.26	< 0.001	2.24	< 0.001	0.76	0.998	0.34	0.002	0.34	< 0.001
ko00592; alpha-Linolenic acid metabolism	3.06E-04	7.15E-04	7.09E-04	2.94E-04	2.34	< 0.001	2.32	< 0.001	0.96	0.998	0.41	0.002	0.41	< 0.001
ko04260; Cardiac muscle contraction	1.79E-04	3.85E-04	3.81E-04	1.64E-04	2.15	< 0.001	2.13	< 0.001	0.91	0.998	0.43	0.002	0.43	< 0.001
ko05134; Legionellosis	3.82E-03	7.51E-03	7.44E-03	3.40E-03	1.97	< 0.001	1.95	< 0.001	0.89	0.998	0.45	0.002	0.46	< 0.001
ko05012; Parkinsons disease	2.25E-04	4.50E-04	4.46E-04	2.14E-04	2.00	< 0.001	1.98	< 0.001	0.95	0.998	0.48	0.002	0.48	< 0.001
ko00633; Nitrotoluene degradation	2.03E-03	3.92E-03	3.89E-03	1.92E-03	1.93	< 0.001	1.92	< 0.001	0.95	0.998	0.49	0.002	0.49	< 0.001
ko04612; Antigen processing and presentation	2.14E-04	3.43E-04	3.41E-04	1.75E-04	1.60	< 0.001	1.59	< 0.001	0.82	0.998	0.51	0.002	0.51	< 0.001

[a]The top 10 significantly up-regulated and down-regulated microbial metabolic pathways both in CAG and IM compared to normal/SG group (both $q < 0.001$) were listed. [b]The top 10 significantly up-regulated and down-regulated microbial metabolic pathways in DYS/GC compared to IM group ($q < 0.001$) were listed. [c]Relative frequency median of predicted microbiota functional capacity. [d]The fold change (FC) was calculated as the ratio of relative frequency median in advanced lesion to the relative frequency median of corresponding mild lesion. [e]Unconditional logistic regression adjusted for age, sex, smoking and alcohol consumption status and q values were used after adjustment for multiple comparison by FDR. CAG, chronic atrophic gastritis; DYS, dysplasia; FC, fold change; GC, gastric cancer; IM, intestinal metaplasia; SG, superficial gastritis.

Supplementary Table 6. The most significant up-regulated metabolic pathways after lesion progression from IM to DYS/GC

	Progression subjects from IM to DYS/GC				Non-progression IM subjects				Comparison of initial biopsies between progression and non-progression subjects		Comparison of follow-up biopsies between progression and non-progression subjects	
	Initial biopsies n = 11	Follow-up biopsies n = 11	FC[a]	*p* value[c]	Initial biopsies n = 5	Follow-up biopsies n = 5	FC[a]	*p* value[c]	FC[b]	*p* value[d]	FC[b]	*p* value[d]
ko04974; Protein digestion and absorption	1.28E-03	1.41E-03	1.10	0.248	5.97E-04	2.19E-05	0.04	0.063	2.14	0.234	64.22	0.006
ko00785; Lipoic acid metabolism	1.32E-03	1.54E-03	1.16	0.248	3.85E-04	3.58E-05	0.09	0.063	3.44	0.234	43.03	0.006
ko01057; Biosynthesis of type II polyketide products	1.04E-04	1.44E-04	1.39	0.248	1.91E-05	3.45E-06	0.18	0.063	5.43	0.126	41.75	0.006
ko00522; Biosynthesis of 12-, 14- and 16-membered macrolides	1.85E-04	2.61E-04	1.41	0.248	2.68E-05	6.47E-06	0.24	0.313	6.89	0.079	40.26	0.006
ko00100; Steroid biosynthesis	3.85E-05	5.10E-05	1.32	0.310	5.32E-06	1.29E-06	0.24	0.313	7.25	0.126	39.57	0.006
ko00909; Sesquiterpenoid and triterpenoid biosynthesis	1.49E-04	1.86E-04	1.25	0.304	1.99E-05	4.78E-06	0.24	0.438	7.46	0.100	38.87	0.006
ko04726; Serotonergic synapse	1.35E-05	1.62E-05	1.19	0.248	2.09E-06	4.50E-07	0.22	0.063	6.47	0.100	35.92	0.006
ko00984; Steroid degradation	3.85E-04	4.70E-04	1.22	0.248	9.14E-05	1.60E-05	0.18	0.063	4.21	0.126	29.40	0.006
ko04920; Adipocytokine signaling pathway	6.12E-04	6.90E-04	1.13	0.248	1.30E-04	3.19E-05	0.24	0.063	4.70	0.126	21.64	0.006
ko03320; PPAR signaling pathway	1.16E-03	1.30E-03	1.12	0.248	2.71E-04	9.28E-05	0.34	0.063	4.28	0.157	14.01	0.006
ko00351; DDT degradation	7.39E-05	8.84E-05	1.20	0.403	2.14E-05	1.38E-05	0.64	0.313	3.46	0.100	6.41	0.006

[a]The fold change (FC) was calculated as the ratio of the relative frequency median in follow-up biopsies to that in initial biopsies. [b]The fold change (FC) was calculated as the ratio of the relative frequency median in progression subjects to that in non-progression subjects. [c]Wilcoxon signed-rank test. [d]Mann-Whitney U test. DYS, dysplasia; FC, fold change; GC, gastric cancer; IM, intestinal metaplasia.

Supplementary Table 7. Significant changes in predicted nitrite related orthologs using KEGG among various gastric lesions

	normal/SG[a]	CAG[a]	IM[a]	DYS/GC[a]	CAG *v.s.* normal/SG		IM *v.s.* normal/SG		DYS/GC *v.s.* normal/SG		DYS/GC *v.s.* CAG		DYS/GC *v.s.* IM	
	n = 35	n = 52	n = 67	n = 25	FC[b]	*q* value[c]	FC[b]	*q* value[c]	FC[b]	*q* value[c]	FC[b]	*q* value[c]	FC[b]	*q* value[c]
K04747; nitric oxide reductase NorF protein	4.51E-07	9.05E-09	7.45E-09	7.65E-07	0.02	0.298	0.02	0.013	1.70	0.950	84.50	0.047	102.64	< 0.001
K02305; nitric oxide reductase subunit C	9.35E-06	2.02E-07	2.14E-07	1.47E-05	0.02	0.012	0.02	0.003	1.57	0.950	72.73	0.005	68.70	< 0.001
K15864; nitrite reductase (NO-forming)/hydroxylamine reductase	2.72E-05	5.65E-07	6.12E-07	3.65E-05	0.02	0.001	0.02	0.006	1.34	0.950	64.69	0.003	59.67	< 0.001
K02164; nitric oxide reductase NorE protein	9.82E-06	2.17E-07	2.36E-07	1.34E-05	0.02	0.013	0.02	0.001	1.36	0.950	61.64	0.012	56.54	< 0.001
K01721; nitrile hydratase	1.28E-05	3.08E-07	3.19E-07	1.69E-05	0.02	0.030	0.02	0.001	1.33	0.950	54.95	0.029	53.18	< 0.001
K00372; nitrate reductase catalytic subunit	1.18E-04	2.91E-06	3.52E-06	1.72E-04	0.02	0.004	0.03	0.006	1.46	0.955	59.08	0.006	48.84	< 0.001
K02448; nitric oxide reductase NorD protein	2.92E-05	6.57E-07	6.82E-07	3.33E-05	0.02	0.031	0.02	0.003	1.14	0.950	50.64	0.014	48.77	< 0.001
K00368; nitrite reductase (NO-forming)	9.26E-05	2.43E-06	2.66E-06	1.22E-04	0.03	0.001	0.03	0.001	1.32	0.985	50.27	0.006	45.80	< 0.001
K04748; nitric oxide reductase NorQ protein	3.25E-05	7.80E-07	1.09E-06	4.78E-05	0.02	0.003	0.03	0.003	1.47	0.950	61.22	0.006	43.81	< 0.001
K02571; periplasmic nitrate reductase NapE	1.31E-06	3.49E-08	3.80E-08	1.61E-06	0.03	0.013	0.03	0.003	1.23	0.950	46.12	0.012	42.29	< 0.001
K00363; nitrite reductase (NAD(P)H) small subunit	2.54E-05	6.81E-07	9.05E-07	3.65E-05	0.03	0.002	0.04	0.002	1.44	0.973	53.60	0.006	40.33	< 0.001
K05916; nitric oxide dioxygenase	9.58E-05	2.50E-06	3.20E-06	1.22E-04	0.03	0.001	0.03	0.001	1.27	0.999	48.87	0.004	38.15	< 0.001
K00362; nitrite reductase (NAD(P)H) large subunit	2.38E-04	6.49E-06	9.01E-06	3.43E-04	0.03	0.001	0.04	0.002	1.44	0.968	52.82	0.005	38.09	< 0.001
K13771; Rrf2 family transcriptional regulator, nitric oxide-sensitive transcriptional repressor	4.87E-05	1.24E-06	1.45E-06	5.50E-05	0.03	0.001	0.03	0.001	1.13	0.950	44.43	0.009	37.95	< 0.001
K07684; two-component system, NarL family, nitrate/nitrite response regulator NarL	7.11E-05	1.75E-06	2.15E-06	8.04E-05	0.02	0.001	0.03	0.001	1.13	0.950	45.97	0.007	37.33	< 0.001
K02575; MFS transporter, NNP family, nitrate/nitrite transporter	2.42E-04	5.97E-06	9.64E-06	3.36E-04	0.02	0.001	0.04	0.001	1.39	0.972	56.38	0.004	34.89	< 0.001
K12266; anaerobic nitric oxide reductase transcription regulator	6.05E-05	1.96E-06	2.18E-06	7.20E-05	0.03	0.001	0.04	0.001	1.19	0.985	36.82	0.006	33.10	< 0.001
K04561; nitric oxide reductase subunit B	2.93E-04	7.28E-06	8.85E-06	2.90E-04	0.02	< 0.001	0.03	0.001	0.99	0.980	39.86	0.003	32.79	< 0.001
K00373; nitrate reductase 1, delta subunit	5.22E-05	1.32E-06	2.24E-06	6.56E-05	0.03	< 0.001	0.04	0.001	1.26	0.959	49.48	0.003	29.27	< 0.001
K00371; nitrate reductase 1, beta subunit	1.73E-04	4.80E-06	6.90E-06	1.89E-04	0.03	< 0.001	0.04	0.001	1.09	0.950	39.41	0.003	27.41	< 0.001
K00374; nitrate reductase 1, gamma subunit	7.22E-05	1.99E-06	2.99E-06	8.16E-05	0.03	< 0.001	0.04	0.001	1.13	0.950	41.02	0.003	27.30	< 0.001
K00370; nitrate reductase 1, alpha subunit	4.25E-04	1.29E-05	1.78E-05	4.66E-04	0.03	< 0.001	0.04	0.001	1.10	0.950	36.16	0.003	26.13	< 0.001
K00366; ferredoxin-nitrite reductase	2.50E-05	1.12E-06	1.28E-06	3.23E-05	0.04	0.001	0.05	0.001	1.29	0.950	28.78	0.003	25.21	< 0.001
K12264; anaerobic nitric oxide reductase flavorubredoxin	4.89E-05	1.83E-06	3.05E-06	3.65E-05	0.04	< 0.001	0.06	0.001	0.75	0.950	19.99	0.003	11.96	< 0.001

[a]Relative frequency median of predicted nitrite related orthologs. [b]The fold change (FC) was calculated as the ratio of the relative frequency medians in different gastric lesions. [c]Unconditional logistic regression adjusted for age, sex, smoking and alcohol consumption status and *q* values were used after adjustment for multiple comparison by FDR. CAG, chronic atrophic gastritis; DYS, dysplasia; FC, fold change; GC, gastric cancer; IM, intestinal metaplasia; SG, superficial gastritis.

Supplementary Table 8. The predicted nitrite related orthologs using KEGG associated with gastric lesion progression to DYS/GC from IM

	Progression subjects from IM to DYS/GC				Non-progression IM subjects				Comparison of initial biopsies between progression and non-progression subjects		Comparison of follow-up biopsies between progression and non-progression subjects	
	Initial biopsies n = 11	Follow-up biopsies n = 11	FC[a]	*p* value[c]	Initial biopsies n = 5	Follow-up biopsies n = 5	FC[a]	*p* value[c]	FC[b]	*p* value[d]	FC[b]	*p* value[d]
K00368; nitrite reductase (NO-forming)	1.10E-04	1.42E-04	1.30	0.365	2.88E-05	3.33E-06	0.12	0.625	3.81	0.079	42.62	0.006
K05916; nitric oxide dioxygenase	8.43E-05	1.25E-04	1.49	0.010	1.46E-05	3.08E-06	0.21	0.313	5.78	0.336	40.74	0.020
K07684; two-component system, NarL family, nitrate/nitrite response regulator NarL	6.11E-05	8.57E-05	1.40	0.083	9.20E-06	2.14E-06	0.23	0.625	6.64	0.126	40.03	0.008
K00363; nitrite reductase (NAD(P)H) small subunit	3.23E-05	4.29E-05	1.33	0.320	4.77E-06	1.14E-06	0.24	0.313	6.77	0.126	37.67	0.008
K13771; Rrf2 family transcriptional regulator, nitric oxide-sensitive transcriptional repressor	4.25E-05	5.68E-05	1.34	0.240	5.81E-06	1.51E-06	0.26	0.625	7.31	0.126	37.65	0.008
K02575; MFS transporter, NNP family, nitrate/nitrite transporter	2.96E-04	3.49E-04	1.18	0.465	4.06E-05	9.54E-06	0.23	0.625	7.29	0.126	36.56	0.015
K00362; nitrite reductase (NAD(P)H) large subunit	2.92E-04	3.80E-04	1.30	0.320	4.21E-05	1.07E-05	0.26	0.438	6.93	0.126	35.38	0.008
K12266; anaerobic nitric oxide reductase transcription regulator	6.79E-05	7.93E-05	1.17	0.413	1.61E-05	2.27E-06	0.14	0.313	4.21	0.157	35.02	0.006
K04748; nitric oxide reductase NorQ protein	4.98E-05	5.47E-05	1.10	> 0.999	6.52E-06	1.59E-06	0.24	0.125	7.63	0.126	34.47	0.008
K02164; nitric oxide reductase NorE protein	1.44E-05	1.46E-05	1.02	0.831	2.38E-06	4.73E-07	0.20	0.125	6.05	0.126	30.96	0.006
K15864; nitrite reductase (NO-forming)/hydroxylamine reductase	3.59E-05	4.21E-05	1.17	> 0.999	7.13E-06	1.38E-06	0.19	0.313	5.03	0.193	30.52	0.006
K00372; nitrate reductase catalytic subunit	1.92E-04	1.99E-04	1.04	> 0.999	2.70E-05	6.88E-06	0.25	0.313	7.11	0.126	28.94	0.008

[a]The fold change (FC) was calculated as the ratio of the relative frequency median in follow-up biopsies to that in initial biopsies. [b]The fold change (FC) was calculated as the ratio of the relative frequency median in progression subjects to that in non-progression subjects. [c]Wilcoxon signed-rank test. [d]Mann-Whitney U test. DYS, dysplasia; FC, fold change; GC, gastric cancer; IM, intestinal metaplasia.