Journal of Pathology
J Pathol November 2022; **258:** 300–311
Published online 16 September 2022 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/path.6000

ORIGINAL ARTICLE

Genomic features of *Helicobacter pylori*–naïve diffuse-type gastric cancer

Ken Namikawa[1†], Norio Tanaka[2†], Yuki Ota[3], Manabu Takamatsu[4], Mayuko Kosugi[2], Yoshitaka Tokai[1], Shoichi Yoshimizu[1], Yusuke Horiuchi[1], Akiyoshi Ishiyama[1], Toshiyuki Yoshio[1], Toshiaki Hirasawa[1], Sayuri Amino[3], Rie Furuya[3], Osamu Gotoh[2], Tomoko Kaneyasu[2], Izuma Nakayama[5], Yu Imamura[6], Tetsuo Noda[7], Junko Fujisaki[1*] and Seiichi Mori[2,8*]

[1] *Department of Gastroenterology, Cancer Institute Hospital, Japanese Foundation for Cancer Research, Tokyo, Japan*
[2] *Project for Development of Innovative Research on Cancer Therapeutics, Cancer Precision Medicine Center, Japanese Foundation for Cancer Research, Tokyo, Japan*
[3] *Project for Development of Genomics-based Cancer Medicine, Cancer Precision Medicine Center, Japanese Foundation for Cancer Research, Tokyo, Japan*
[4] *Division of Pathology, Cancer Institute, Japanese Foundation for Cancer Research, Tokyo, Japan*
[5] *Department of Gastroenterological Chemotherapy, Cancer Institute Hospital, Japanese Foundation for Cancer Research, Tokyo, Japan*
[6] *Department of Gastroenterological Surgery, Cancer Institute Hospital, Japanese Foundation for Cancer Research, Tokyo, Japan*
[7] *Cancer Institute, Japanese Foundation for Cancer Research, Tokyo, Japan*
[8] *Department of Genetic Diagnosis, Cancer Institute Hospital, Japanese Foundation for Cancer Research, Tokyo, Japan*

**Correspondence to: J Fujisaki, Department of Gastroenterology, Cancer Institute Hospital, Japanese Foundation for Cancer Research, Ariake 3-8-31, Koto-ku, Tokyo, 135-8550, Japan. E-mail: junko.fujisaki@jfcr.or.jp; S Mori, Project for Development of Innovative Research on Cancer Therapeutics, Cancer Precision Medicine Center, Japanese Foundation for Cancer Research, Ariake 3-8-31, Koto-ku, Tokyo, 135-8550, Japan. E-mail: seiichi.mori@jfcr.or.jp*

[†]Equal first authors.

Abstract

Helicobacter pylori **(HP) is a major etiologic driver of diffuse-type gastric cancer (DGC). However, improvements in hygiene have led to an increase in the prevalence of HP-naïve DGC; that is, DGC that occurs independent of HP. Although multiple genomic cohort studies for gastric cancer have been conducted, including studies for DGC, distinctive genomic differences between HP-exposed and HP-naïve DGC remain largely unknown. Here, we employed exome and RNA sequencing with immunohistochemical analyses to perform binary comparisons between 36 HP-exposed and 27 HP-naïve DGCs from sporadic, early-stage, and intramucosal or submucosal tumor samples. Among the samples, 33 HP-exposed and 17 HP-naïve samples had been preserved as fresh-frozen samples. HP infection status was determined using stringent criteria. HP-exposed DGCs exhibited an increased single nucleotide variant burden (HP-exposed DGCs; 1.97 [0.48–7.19] and HP-naïve DGCs; 1.09 [0.38–3.68] per megabase; $p = 0.0003$) and a higher prevalence of chromosome arm-level aneuploidies ($p < 0.0001$). *CDH1* was mutated at similar frequencies in both groups, whereas the RHOA–ARHGAP pathway misregulation was exclusive to HP-exposed DGCs ($p = 0.0167$). HP-exposed DGCs showed gains in chromosome arms 8p/8q ($p < 0.0001$), 7p ($p = 0.0035$), and 7q ($p = 0.0354$), and losses in 16q ($p = 0.0167$). Immunohistochemical analyses revealed a higher expression of intestinal markers such as CD10 ($p < 0.0001$) and CDX2 ($p = 0.0002$) and a lower expression of the gastric marker, MUC5AC ($p = 0.0305$) among HP-exposed DGCs. HP-naïve DGCs, on the other hand, had a purely gastric marker phenotype. This work reveals that HP-naïve and HP-exposed DGCs develop along different molecular pathways, which provide a basis for early detection strategies in high incidence settings.**

Keywords: gastric cancer; *Helicobacter pylori*; diffuse-type gastric cancer; exome sequencing; RNA sequencing; immunohistochemistry; RHOA–ARHGAP pathway; chromosome arm-level aneuploidy; gastric marker; intestinal marker

Received 19 April 2022; Revised 25 July 2022; Accepted 9 August 2022

No conflicts of interest were declared.

Introduction

According to the conventional Lauren classification, gastric cancer is divided into two major histological types: intestinal-type gastric cancer (IGC) and diffuse-type gastric cancer (DGC) [1]. IGC exhibits intestinal differentiation that morphologically mimics intestinal glandular cells and develops through the "intestinal

metaplasia–dysplasia–carcinoma" sequence [1,2]. DGC, on the other hand, consists of individually infiltrating tumor cells with signet-ring and/or poorly differentiated histology, and is considered to develop directly from gastric epithelial cells [1]. *Helicobacter pylori* (HP) is a well-known major cause of gastric cancer, particularly IGC, but also DGC, with various epidemiological studies documenting that a fraction of DGC arises in patients with concomitant HP infection [3–5]. However, given that the prevalence of HP infection globally decreases with improvements in hygiene [6], HP-naïve gastric cancer has now become an increasingly relevant consideration, accounting for 0.4 to 1.3% of gastric cancer. HP-naïve gastric cancer is a heterogeneous entity, and includes sporadic DGC, hereditary DGC, and gastric adenocarcinoma of the fundic gland type [5]. Stringent criteria are used to confirm HP negativity in HP-naïve gastric cancers, relying on multimodal findings from endoscopy, histopathology, and laboratory tests for serum anti-HP IgG antibody levels [7,8], serum pepsinogen levels [9], and urea breath test results [5,10]. Recent efforts have identified several clinicopathological features to be associated with sporadic HP-naïve DGC: young age of onset, signet-ring histology with gastric-type mucin expression, localization in the lower and middle regions of the body of the stomach, and early-stage cancer; this is presumably due to its indolent progression and slow proliferative capacity [5,11,12]. Numerous genomic cohort studies have been conducted for gastric cancer, including DGC [13–24], and these have led to the identification of relevant driver genes and the development of molecular subtyping schemes. Nevertheless, due to the rarity of the disease and other technical hurdles associated with determining HP negativity, the molecular etiologies and genomic features of HP-naïve DGC remain largely unknown. In the current study we applied stringent criteria to select DGC patients with or without HP infection, and then conducted comparative exome, transcriptome, and immunohistochemical analyses to better understand the tumorigenic process and identify the genomic profiles common or specific to DGC patients with and without HP infection. We show that, compared with HP-exposed DGC, HP-naïve DGC presents with a stable and salient genomic nature and a pure gastric phenotype. Although further genomic and epigenetic studies are still required, these findings provide a molecular basis toward gaining an understanding of the tumorigenic program and the relevant knowledge for designing novel detection and preventive methods for sporadic DGC in the presence or absence of HP infection.

Materials and methods

Ethics and informed consent

Gastric carcinoma patients underwent tumor excision by endoscopic submucosal dissection or surgery at the Cancer Institute Hospital of Japanese Foundation for Cancer Research (JFCR) between 2010 and 2019. Recruited patients provided written informed consent. Ethical approval was obtained from the internal review boards of the JFCR (approval number 2018-1074).

Terminology of clinicopathological characteristics

Clinicopathological characteristics such as tumor location, macroscopic type, tumor invasion depth, histological classification, and lymphovascular invasion were classified using the "Japanese Classification of Gastric Carcinoma: 3rd English Edition" [25]. The variables for each case are provided in supplementary material, Table S1.

Inclusion criteria

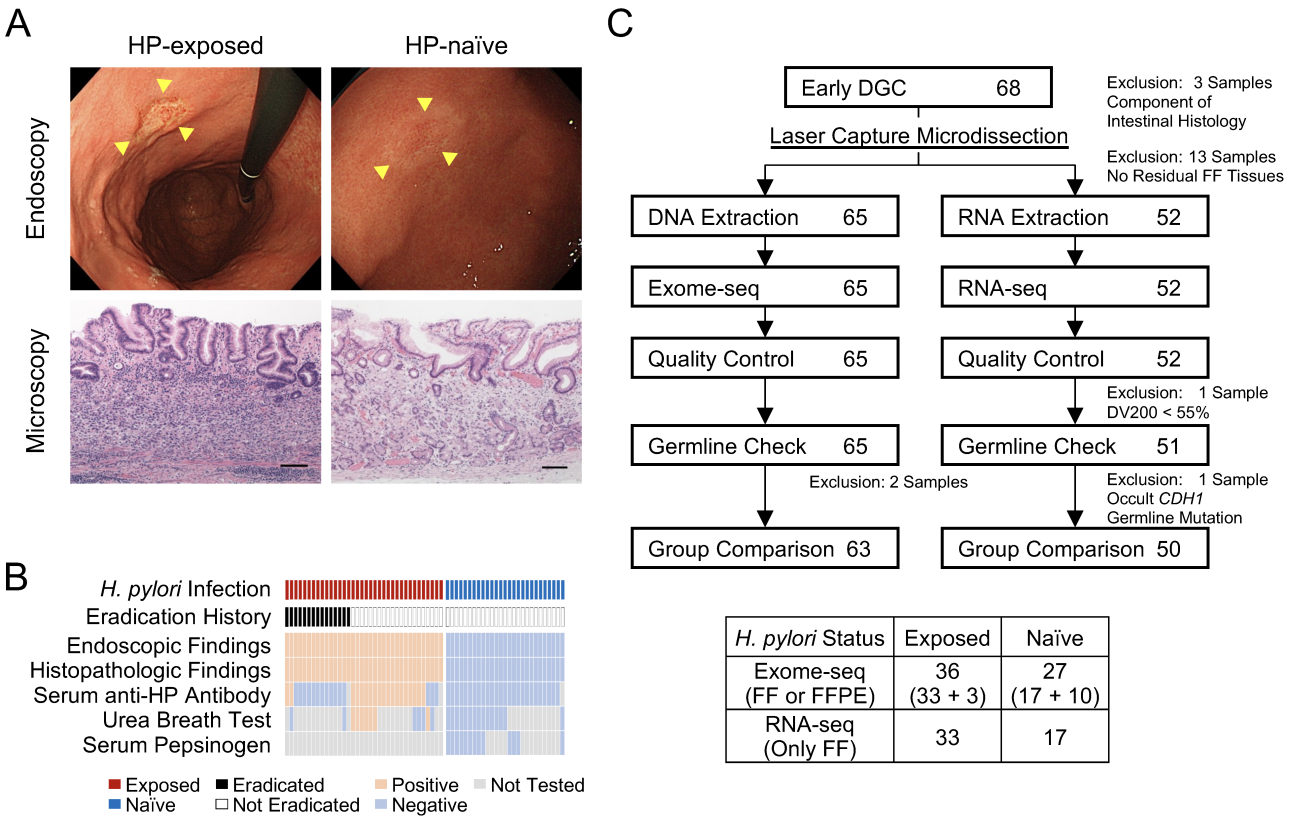
Subjects who presented with sporadic early diffuse gastric carcinoma with or without HP infection were selected for inclusion based on strict lesion criteria, as follows: (1) single lesion; (2) size ≥8 mm in endoscopically-estimated long diameter; (3) invasion depth penetrating the intramucosal or submucosal layers; (4) diffuse histology (signet-ring cell carcinoma or poorly differentiated adenocarcinoma) [1,25]; and (5) sporadic case.

Criteria for *Helicobacter pylori* infection classification

HP infection status was assessed and classified using the following criteria: (1) history of HP eradication therapy (i.e. previous treatment with antibiotics elsewhere); (2) endoscopic findings for mucosal atrophy (pale mucosa, submucosal vessel prominence, and loss of regular arrangement of collecting venules in the lower gastric body) or gastritis (diffuse or spotty redness, fold enlargement, sticky mucus, xanthoma, and intestinal metaplasia) [26–28]; (3) histopathologic findings for evidence of mucosal atrophy (loss of glandular cells), the presence of HP bacterial cells, or gastritis (inflammatory-cell infiltration in mucosa and intestinal metaplasia) [5,26,27]; and (4) laboratory tests (serum anti-HP antibody [7,8], serum pepsinogen [9], and urea breath test [10]; see also Supplementary materials and methods and supplementary material, Table S1). HP-naïve cases were defined as those with no history of HP eradication, no sign of atrophic gastritis or HP infection in endoscopic and histopathologic examinations, and one or more negative results in laboratory tests. HP-exposed cases were those who had at least one positive result for any of the items listed above (Figure 1B).

Processing tumor samples

Tumor specimens were preserved as formalin-fixed paraffin-embedded (FFPE) or fresh-frozen (snap-frozen) (FF) samples. FFPE tissue samples were processed for histopathological, immunohistochemical, and fluorescence *in situ* hybridization (FISH) examinations.



H. pylori Status	Exposed	Naïve
Exome-seq (FF or FFPE)	36 (33 + 3)	27 (17 + 10)
RNA-seq (Only FF)	33	17

Figure 1. Genomic analysis of *Helicobacter pylori* (HP)-exposed and HP-naïve diffuse gastric carcinoma (DGC). (A) Endoscopic and histopathologic appearance of HP-exposed and HP-naïve DGCs. Upper and lower panels indicate endoscopic and microscopic images from the same patients (left; HP-exposed, and right; HP-naïve DGCs). Arrowheads in the endoscopic image point to the edge of the tumor lesion. The horizontal bar in the microscopic image indicates 100 µm. Note that pale mucosa, submucosal vessel prominence, and intestinal metaplasia are observed in the photograph of HP-exposed stomach. In the HP-exposed DGC microscopic image, inflammatory-cell infiltration and loss of glandular cells can be seen. (B) HP infection status, eradication history, and positivity and negativity in clinical tests to classify HP infection status. DGC patients were classified into HP-exposed and HP-naïve cases via stringent criteria (see Materials and Methods). (C) REMARK diagram of the current study. The sample number of early DGC is shown at each step: pathological review, sample preparation, sequencing, and informatics/data quality control. Starting with 68 early DGCs, 63 (36 HP-exposed and 27 HP-naïve) and 50 (33 HP-exposed and 17 HP-naïve) samples finally passed stringent quality assessment and were taken for exome sequencing and RNA-seq analyses, respectively. Whereas exome data were generated from fresh-frozen (FF) or formalin-fixed paraffin-embedded (FFPE) samples, RNA-seq data were prepared only from FF samples. H, *Helicobacter*.

The exome analysis was also carried out using FFPE samples when FF samples were unavailable. Only FF samples were rendered to RNA-seq (supplementary material, Table S2). FF and FFPE tissues were cut into 10-µm-thick sections, and cancer cells were selectively enriched using an LMD7000 laser-capture microdissection system (Leica Microsystems, Shinjuku, Tokyo, Japan), following the manufacturer's protocol.

Exome sequencing depth

Exome sequencing data were generated with a median depth of coverage of 283 (154 [minimum]–406 [maximum]) per tumor from FF samples ($n = 50$; 33 HP-exposed and 17 HP-naïve samples), 272 (130–331) per tumor from FFPE samples ($n = 13$), and 173 (124–228) per paired-normal sample ($n = 63$).

The following sections are described in Supplementary materials and methods: Laboratory tests to determine HP infection status; DNA/RNA preparation for genomic analysis; Exome sequencing analysis; RNA sequencing analysis; Alignment, local realignment, somatic variant call, and the other informatics analyses for exome sequencing; Bioinformatics analyses for RNA-seq data; Fusion gene detection, FISH analysis for chromosome arm aneuploidy; Immunohistochemistry; and Statistical analyses.

Results

Genomic cohort of *Helicobacter pylori* (HP)-exposed and HP-naïve diffuse gastric carcinoma

Based on the stringent criteria described in the Materials and methods, patients with sporadic intramucosal or submucosal DGC were classified into HP-exposed and HP-naïve cases (Figure 1A–C and supplementary material, Table S1). With an original starting cohort of 68 DGC samples (40 HP-exposed and 28 HP-naïve), three

samples were excluded because the tumor contained tissue of intestinal histology (moderately differentiated tubular histology [25]). We further excluded another two cases following exome analyses, as these cases had occult germline mutations in the *CDH1* locus. After quality checking that the samples were appropriate for exome- and RNA-sequencing, a comparative analysis of exome data was finally conducted with 36 HP-exposed and 27 HP-naïve samples, whereas 33 HP-exposed and 17 HP-naïve snap-frozen samples were used for RNA-seq data analysis (Figure 1C).

The clinicopathological characteristics of the patients and samples are presented in Table 1 and supplementary material, Table S1. In the cohort, there were 15 cases with a history of HP eradication (Figure 1B and supplementary material, Table S1). Among the clinicopathological attributes, we found significant differences between HP-exposed and HP-naïve cases for Brinkman index, tumor size, tumor location, macroscopic type, resection method, tumor invasion depth, and histological classification; these findings are consistent with previous observations (Table 1) [5,11,29] and validate both the stringency of our criteria for HP infection status and the accuracy of clinicopathologic information in the current cohort.

Genomic landscape of HP-naïve DGCs

Exome analyses revealed a significantly higher number of single nucleotide variants (SNVs) in HP-exposed DGCs than in HP-naïve DGCs (HP-exposed DGCs; 1.97 [0.48 minimum–7.19 maximum] and HP-naïve DGCs; 1.09 [0.38–3.68] per megabase, $p = 0.0003$). The numbers of indels and abnormal segments with copy number aberrations were comparable between the two groups. There was a significantly higher number of chromosomal arms with copy number aberrations among the HP-exposed DGC samples (Figure 2A,B). Two HP-exposed DGCs were classified as chromosomal instability (CIN) subtype tumors, while the remaining HP-exposed tumors and all of HP-naïve tumors were classified as genomically stable (GS) subtype tumors. Only one sample among the HP-exposed cases was tetraploid, with all of the other DGCs diploid (Figure 2A). C>A nucleotide substitution was more frequently observed among HP-exposed samples, whereas T>A and T>C substitutions were more evident among the HP-naïve tumors (Figure 2A,B). These observations indicate that genomic aberration profiles are distinct with respect to HP infection status, even with similar signet-ring cell histology or with a GS molecular subtype.

Table 1. Patient demographics and clinicopathological characteristics.

Characteristics	HP-exposed ($n = 36$)	HP-naïve ($n = 27$)	*P* value
Sex			
Female	20	10	$p = 0.2034$*
Male	16	17	
Age			
Median (range)	54 (38–85)	51.5 (27–86)	$p = 0.1260$†
Alcohol consumption			
Yes	20	20	$p = 0.1871$*
No	16	7	
Brinkman index‡			
Median (range)	0 (0–960)	300 (0–2,220)	$p = 0.0057$†
Tumor size (mm)			
Median (range)	20.5 (8–44)	13 (3–25)	$p < 0.0001$†
Tumor location			
Upper/Middle third	11	1	$p = 0.0087$*
Lower third	25	26	
Macroscopic type			
Type 0–IIb	2	15	$p < 0.0001$*
Type 0–IIc/0–IIc + IIa	34	12	
Resection method			
ESD	19	25	$p = 0.0007$*
Surgery	17	2	
Tumor invasion depth			
M	24	25	$p = 0.0163$*
SM1/2/3	12	2	
Histological classification			
sig	17	25	$p = 0.0001$*
sig + por	19	2	
Lymphovascular invasion			
ly0, v0	32	27	$p = 0.7090$*
ly0, v1	3	0	
ly1, v0	1	0	

ESD, endoscopic submucosal dissection; HP, *Helicobacter pylori*; ly0, no lymphatic invasion; ly1, minimal lymphatic invasion; M, mucosa; por, poorly differentiated adenocarcinoma; sig, signet-ring cell carcinoma; SM, submucosa; Type 0–IIa, superficial elevated; Type 0–IIb, superficial flat; Type 0-IIc, superficial depressed; v0, no venous invasion; v1, minimal venous invasion.

*Computed using Fisher's exact test.

†Computed using Mann–Whitney *U*-test.

‡Brinkman index is calculated as (number of cigarettes smoked per day) multiplied by (number of years smoked).

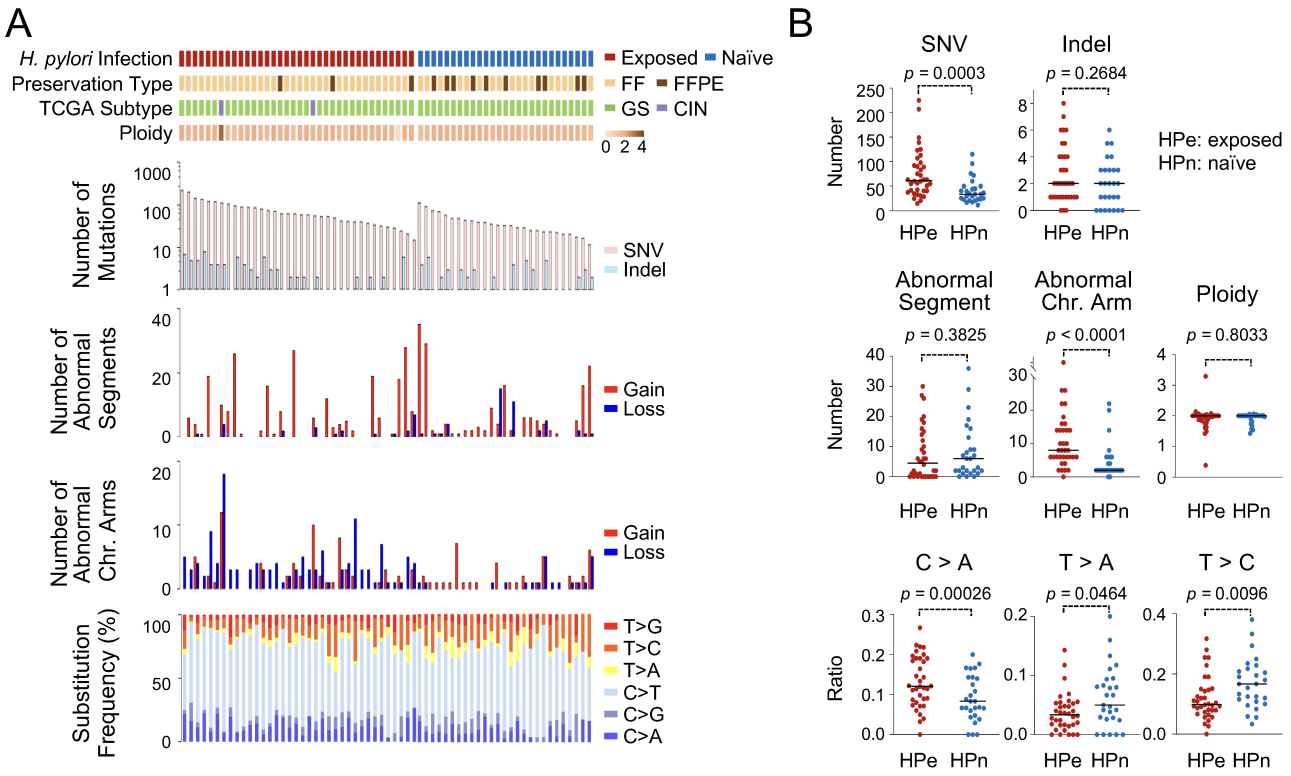


Figure 2. Genomic aberration patterns between HP-exposed and HP-naïve DGCs. (A) Overview of genomic aberrations between HP-exposed and HP-naïve DGCs. Panels from top to bottom: HP-infection status; preservation type (FF or FFPE); TCGA molecular subtype; tumor ploidy; bar plots for number of SNVs and indels; bar plots for number of segments with copy-number aberration (gain or loss); bar plots for number of chromosomal arms with copy-number aberration (gain or loss); and rates (in percent) of nucleotide substitutions. (B) Binary comparisons of genomic aberrations between HP-exposed and HP-naïve DGCs. Top panels, numbers of SNVs and indels; middle panels, numbers of segments with copy-number aberration, chromosomal arms with copy-number aberration, and ploidy; and bottom panels, ratio of C>A, T>A, and T>C substitutions. *P* values were computed using the Mann–Whitney *U*-test. Bars indicate median. H, *Helicobacter*; FF, fresh frozen; FFPE, formalin-fixed paraffin-embedded; TCGA, the Cancer Genome Atlas; HPi, *Helicobacter pylori*-exposed DGCs; HPu, *Helicobacter pylori*-naïve DGCs.

Driver gene alterations in HP-exposed and HP-naïve DGCs

Figure 3 shows protein-truncating SNVs/indels (frameshift indels, nonsense SNVs/indels, and splice-site SNVs/indels), missense SNVs, and CNVs detected among driver genes in HP-exposed and HP-naïve DGCs. Truncating SNVs/indels and missense SNVs were detected in *CDH1*, *ARID1A*, *RHOA*, *MUC6*, *ERBB3*, *KMT2D*, *TGFBR1*, and *TP53* genes in more than three tumor samples in the current cohort; these prevalence rates were comparable with those among samples with signet-ring histology in the TCGA cohort, except for *TP53* mutation [14]. The frequency of *CDH1* mutation did not differ between HP-exposed (22/36; 61.1%) and HP-naïve (17/27; 63.0%) DGCs ($p > 0.9999$ using Fisher's exact test) and although *ARID1A* and *RHOA* exhibited seemingly biased distributions in HP-exposed DGCs, these differences were not significant by Fisher's exact tests (for *ARID1A*; $p = 0.0968$; missense SNVs plus truncating variants were regarded as functional, and for *RHOA*; $p = 0.0651$; only missense SNVs were considered as functional) (Figure 3A). For driver genes other than *CDH1* ($p = 0.0434$ by Mann–Whitney *U*-test), the average numbers of truncating SNVs/indels and missense SNVs per sample were 1.61 and 0.85 in HP-exposed and HP-naïve DGCs, respectively. Of note, *ARID1A* was significantly more frequently comutated with *CDH1* in HP-exposed tumors than in HP-naïve tumors ($p = 0.0124$ by Fisher's exact test) (Figure 3A). Regarding copy number aberrations, there were no differences between the two groups at the segment level, with all aberrations detected as focal amplification events (Figure 3B).

Next, we investigated the presence of fusion genes by split probe FISH for 36 HP-exposed and 27 HP-naïve tumors, and by RNA-seq fusion transcript analysis for 33 HP-exposed and 17 HP-naïve tumors. The assessment revealed two driver fusion transcripts in two HP-exposed tumors: *CLDN18–ARHGAP26* and *CLDN18–ARHGAP18* transcripts (Figures 3C and 4C). HP-exposed DGCs were more frequently associated with the combination of *RHOA* missense SNVs and *CLDN18–ARHGAP* fusion ($p = 0.0167$ by Fisher's exact test; Figure 4B), which may imply a link between RHOA–ARHGAP pathway misregulation and HP-induced DGC tumorigenesis. The Firth bias-reduced logistic regression analysis also revealed a significant correlation in RHOA–ARHGAP pathway misregulation and

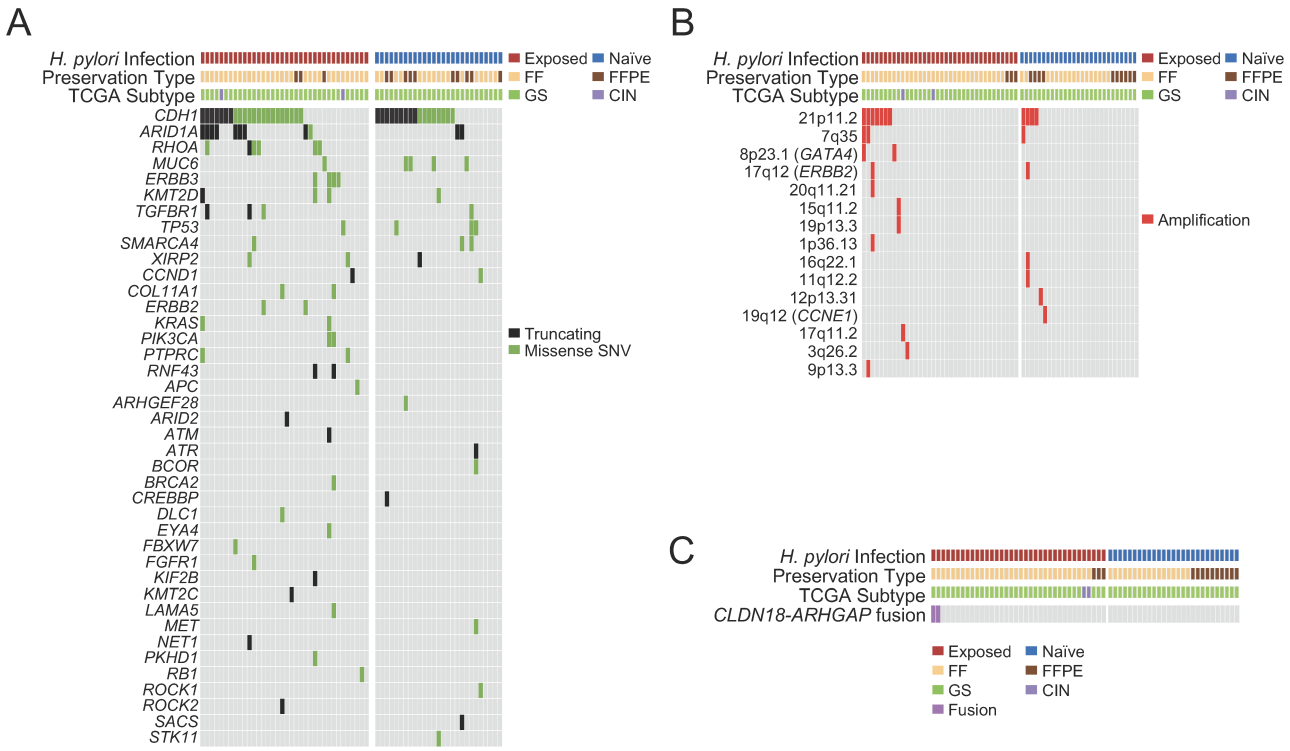


Figure 3. Distribution of driver mutations in HP-exposed and HP-naïve DGCs. (A) Oncoprint of driver mutations (SNVs/indels). (B) Oncoprint of significant copy number alterations detected by GISTIC in the cytoband level. The analysis detected only the regions with amplification (copy number [CN] ≥ 4) and not those with homozygous deletion (CN = 0). Copy number gain (CN = 3) or loss (CN = 1) was not regarded as functionally significant in this analysis. (C) Oncoprint of *CLDN18–ARHGAP* fusion. Sample labeling for HP infection status, preservation type (FF or FFPE), TCGA molecular subtype, and eradication history is shown above the Oncoprints. FF, fresh-frozen; FFPE, formalin-fixed paraffin-embedded; TCGA; the Cancer Genome Atlas.

HP infection status in the univariate analysis ($p = 0.0130$); however, this correlation did not hold in the multivariate model (Table 2). It is noteworthy that this pathway misregulation was neither associated with tumor invasion depth nor histological classification (supplementary material, Table S4). Aside from these gastric cancer drivers, there were no other differentially mutated SNVs/indels (truncating SNVs/indels and missense SNVs), CNVs, or fusion genes between HP-exposed and HP-naïve DGCs in binary comparisons.

Chromosome arm-level aneuploidies in HP-exposed and HP-naïve DGCs

A GISTIC analysis (Genomic Identification of Significant Targets in Cancer) revealed chromosome arm aneuploidies (CAAs) among the samples (Figure 4) [30]: one copy gain for chromosomes 7p, 7q, 8p, and 8q, and one copy loss of chromosome 16q were detected as significantly enriched CAAs in HP-exposed DGCs (Figure 4A, B). The copy number gain of chromosome 8 was validated by FISH analysis using randomly selected tumor tissues (5 HP-exposed samples with chromosome 8 gain and 5 HP-naïve samples without chromosome 8 aberrations) and a centromere probe to chromosome 8 (Figure 4D). Gain of 8p/8q was significantly associated with histological classification, whereas 16q loss was significantly correlated with tumor invasion depth (supplementary material, Table S5). 8p/8q gain and 7p/7q gain were significantly associated with HP infection status, even in the multivariate analyses with clinicopathological parameters, using Firth bias-reduced logistic regression models (Table 2). Collectively, these associations are independent of tumor invasion depth or histological classification.

Transcriptomic and immunohistochemical characteristics of HP-exposed and HP-naïve DGCs

To identify distinctive features between HP-exposed and HP-naïve DGCs, we conducted transcriptomic analyses using RNA-seq. A consensus clustering analysis was conducted with 4,087 genes; these genes were selected because of their highly variable expression across the cohort. The analysis revealed two major clusters: 1A/1B and 2A/2B (Figure 5A). The 1A/1B cluster contained predominantly HP-exposed tumors, whereas cluster 2A/2B comprised all but one of the HP-naïve samples (Figure 5A). Among nine gene clusters, gene clusters I and J included genes encoding chemokine ligands and receptors (*CCL19/20/21/24*, *CXCL1/13/5/8*, and *CXCR2*) and matrix metalloproteases (*MMP1/10/3/7/9*) (cluster I), and B-cell markers (*CD19*, *CD79A*, *IGHG2-4*, *IGLC1/3/7*, *IGLL1/5*, and *TNFRSF13B*) (cluster J), implicating the relevance of inflammatory/immune processes to the tumor clusters (Figure 5A).

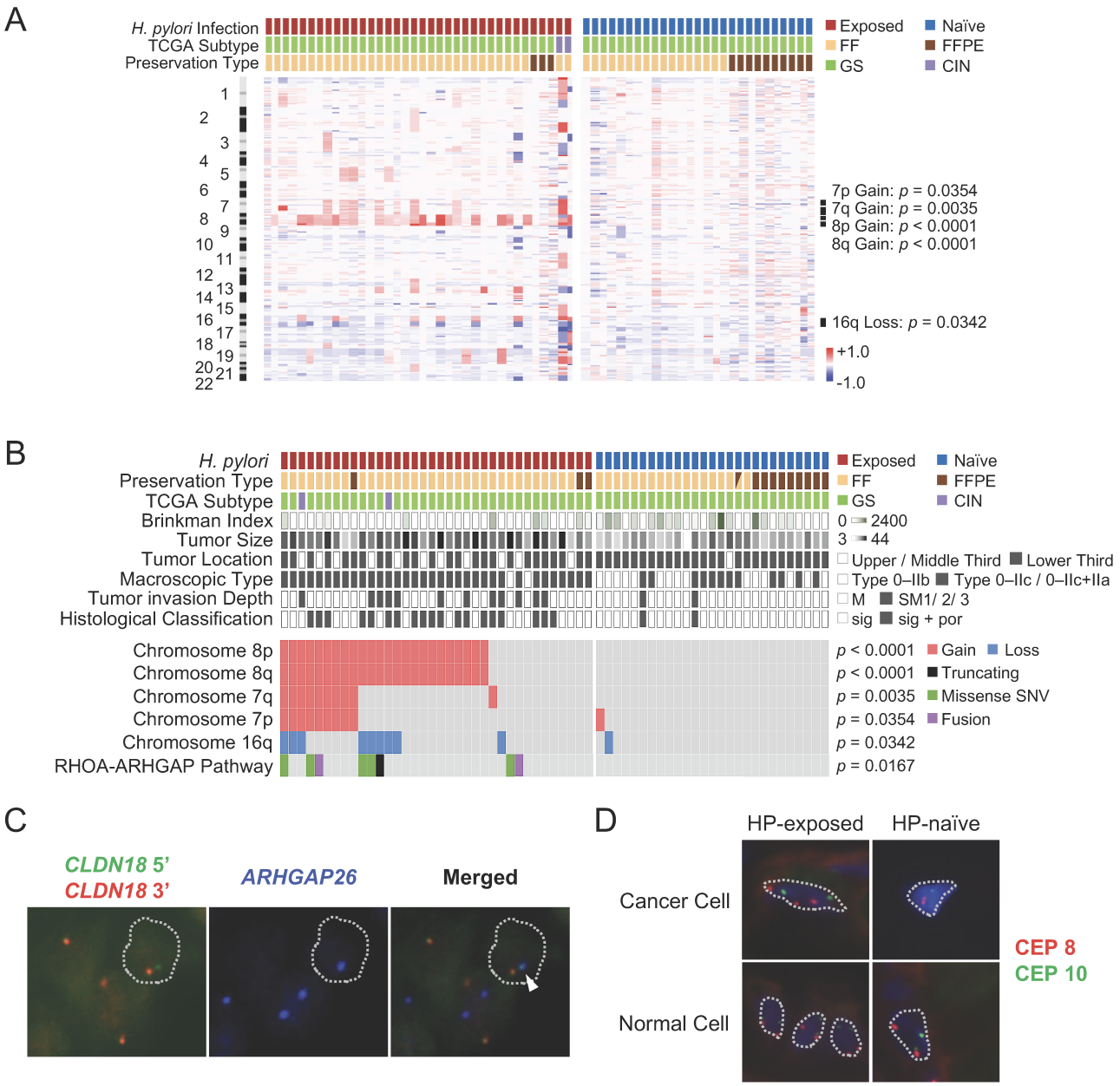


Figure 4. Genomic aberrations associated with HP infection status. (A) Copy number alterations at the chromosome arm level between HP-exposed and HP-naïve DGCs. A GISTIC output presented as a heatmap shows gained (red) or lost (blue) copy number segments at the chromosome arm level. Chromosome position is shown to the left. *P* values <0.05 in binary comparisons by Fisher's exact tests are shown on the right. (B) Oncoprint of CAAs of 8q, 8p, 16q, 7p, and 7q, SNVs/indels for *RHOA* genes, and *CLDN18–ARHGAP* fusion genes. The *P* values in binary comparisons by Fisher's exact tests are shown on the right. Also shown are the *P* values for the combination of *RHOA* missense SNVs and *CLDN18–ARHGAP* fusions. Sample labeling for HP infection status, preservation type (FF or FFPE), TCGA molecular subtype, Brinkman index, tumor size, tumor location, macroscopic type, tumor invasion depth, and histological classification are shown above the Oncoprints. (C) Images of FISH analysis of *CLDN18–ARHGAP26* fusion gene. We show images of split probe assay for *CLDN18* (left panel) and FISH assay for *ARHGAP26* (middle panel), and the merged image of both signals (right panel) in one tumor with the *CLDN18–ARHGAP26* fusion gene. Dashed lines indicate nuclear outlines. Note that the signal for the 5' part of *CLDN18* (green) is colocalized with that for *ARHGAP26* (blue). (D) Representative images of FISH signals for chromosomes #8 and #10 in HP-exposed and HP-naïve cancer cells. FISH analyses were conducted with probes specific for the centromeres of chromosome #8 (CEP8) and #10 (CEP10). Normal cells are foveolar epithelial cells. Nuclei were counterstained with DAPI. Dashed lines indicate nuclear outlines. Note that three and two dots of CEP8 and CEP10 were detected in the nucleus of a cancer cell with one extra copy of chromosome #8 from an HP-exposed patient. FF, fresh-frozen; FFPE, formalin-fixed paraffin-embedded.

Binary comparisons of the gene expression values were assessed using significance analysis of microarrays - receiver operating curve (SAM-ROC) analysis. We found 148 and 285 differentially expressed genes for HP-exposed and HP-naïve DGCs, respectively (SAM *q* value <0.02, and area under the curve [AUC] <0.2 or >0.8; Figure 5B). Of note, markers of intestinal differentiation, *CDX2* and *MUC2*, and CDX2 downstream genes, *SI* and *HOXA9* [31], were highly expressed in HP-exposed samples (*MUC2* transcription is also

Table 2. Univariate and multivariate Firth bias-reduced logistic regression analysis of *Helicobacter pylori* infection status with multiple clinicopathological variables and chromosome arm aneuploidies or *RHOA–ARHGAP* misregulation.

Variables	Univariate OR (CI)	*P* value	Multivariate OR (CI)	*P* value
Brinkman index	1.002 (1.001–1.004)	0.0048	1.000 (0.999–1.003)	0.6751*
Tumor size	0.824 (0.724–0.908)	<0.0001	0.905 (0.793–0.994)	0.0364*
Tumor location	0.126 (0.013–0.588)	0.0062	0.317 (0.019–3.608)	0.3565*
Macroscopic type	0.058 (0.010–0.224)	<0.0001	0.676 (0.091–4.938)	0.6890*
Tumor invasion depth	0.192 (0.035–0.731)	0.0139	5.143 (0.121–651.634)	0.3962*
Histological classification	0.088 (0.016–0.325)	0.0001	0.088 (0.001–2.829)	0.1755*
8p/8q gain	0.009 (0.000–0.077)	<0.0001	0.031 (0.000–0.345)	0.0020*
7q gain	0.046 (0.000–0.389)	0.0016	0.039 (0.000–0.497)	0.0082†
7p gain	0.164 (0.017–0.789)	0.0219	0.107 (0.007–0.892)	0.0381†
16q loss	0.164 (0.017–0.789)	0.0219	0.133 (0.005–1.720)	0.1276†
RHOA-ARHGAP misregulation	0.072 (0.001–0.635)	0.0130	0.156 (0.001–2.368)	0.2070†
MUC2	0.142 (0.017–0.413)	<0.0001	0.627 (0.117–1.800)	0.4182†
CDX2	0.306 (0.115–0.620)	0.0004	1.082 (0.391–3.067)	0.8742†
CD10	0.067 (0.007–0.279)	<0.0001	0.216 (0.019–0.893)	0.0292†

Firth logistic regression analysis was conducted if the attribute showed significant correlation with *Helicobacter pylori* infection status using Mann–Whitney *U*-test or Fisher's exact test.
CI, confidence interval; OR, odds ratio.
*Multivariate Firth's bias-reduced logistic regression analysis of clinicopathological variables with 8p/8q gain.
†For multivariate Firth bias-reduced logistic regression, each molecular event (chromosome arm aberration, RHOA-ARHGAP misregulation, or immunohistochemical protein positivity) was independently analyzed with the other clinicopathological variables (Brinkman index, tumor size, tumor location, macroscopic type, tumor invasion depth, histological classification) from the remaining molecular events.

reported regulated by CDX2 [31]). Conversely, the expression of *MUC5AC*, a gastric phenotype marker, was more evident among HP-naïve samples (Figure 5B, left). These expression patterns are consistent with the finding that HP-exposed and HP-naïve DGCs exhibit, respectively, a higher proportion of intestinal differentiation (additional poorly differentiated histology) and gastric differentiation (pure signet-ring histology) (Table 1) [12]. In addition, we found a higher proportion of immunoglobulin gene component expression among HP-exposed tumors (Figure 5B, left): this is in accordance with the abundance of B-lymphocytic infiltration in typical HP-exposed gastric mucosa (Figure 1A) [32].

Using gene set analysis with ssGSEA along with subsequent statistical evaluations with SAM-ROC, we found gene sets to be differentially regulated between the two groups (712 and 222 gene sets for HP-exposed and HP-naïve DGCs, respectively; SAM q value <0.02, and AUC <0.2 or >0.8; Figure 5B, right). These gene sets included 'Liu CDX2 Targets Up' and multiple gene sets with 'B-cell' annotations (Figure 5B, right); these findings support the results from the analyses using genes (Figure 5B, left). Using informatics analysis of tumor-infiltrating microenvironmental cells by MCPcounter [33], we observed enrichment of B-lineage cells and neutrophils in HP-exposed tumors and higher expression of fibroblastic and endothelial markers among HP-naïve samples (Figure 5C); these findings further support the distinct histological features of HP-exposed and HP-naïve DGCs [32].

HP-exposed DGCs exhibited a significantly higher proportion of intestinal marker protein expression (MUC2, CD10, and CDX2 for protein presence) (Figure 5D), as determined through immunohistochemical analyses, which is consistent with the results from the transcriptome analysis. In line with previous observations [11], HP-exposed DGCs also showed a higher proliferative capacity (higher Ki-67 indices using a Mann–Whitney *U*-test; Figure 5D). The markers, MUC5AC and CLDN18, showed a similar frequency in positivity within the two groups (Figure 5D). Similarly, CDH1-positivity was not significantly different between HP-exposed and HP-naïve DGCs (Figure 5D). To validate these findings, we performed an analysis using an independent cohort of an additional 10 HP-exposed and 10 HP-naïve tumor samples. Among the various markers, many of the findings were reproducible, with significant differences again detected for the expression of CDX2, and for that of Ki67 (supplementary material, Figure S1). In the original cohort, the expression of intestinal markers and Ki-67 (%)—but not gastric markers—were correlated with submucosal invasion or poorly differentiated histology (supplementary material, Table S6). This subgroup analysis revealed intestinal marker positivity to be significantly associated with HP-infection status in intramucosal and signet-ring cell tumors (supplementary material, Table S7), as was the Ki-67-positive proliferative fraction (supplementary material, Table S7).

Collectively, our observations indicate that HP-exposed DGCs expressed a higher proportion of intestinal markers at the protein level, yet maintained baseline gastric cell features, and that intestinal differentiation and proliferative features were directly associated with HP infection.

Discussion

The tumorigenic process in HP-naïve DGCs has been linked with multiple etiological factors, including alcohol abuse, tobacco smoking, the consumption of salt-preserved food, autoimmunity, and germline pathogenic variants in the *CDH1* gene. Yet the tumorigenic program

and any associated genomic aberrations have remained largely unknown [5]. In the current study, clinicopathologically we found a higher proportion of poorly differentiated tumor cells and submucosal tumors with lymphovascular invasion within the HP-exposed tumor group. Genomically, HP-exposed DGCs exhibited a higher proportion of SNVs and aberrant chromosome arms, including gains for 8p/8q and 7p/7q, and

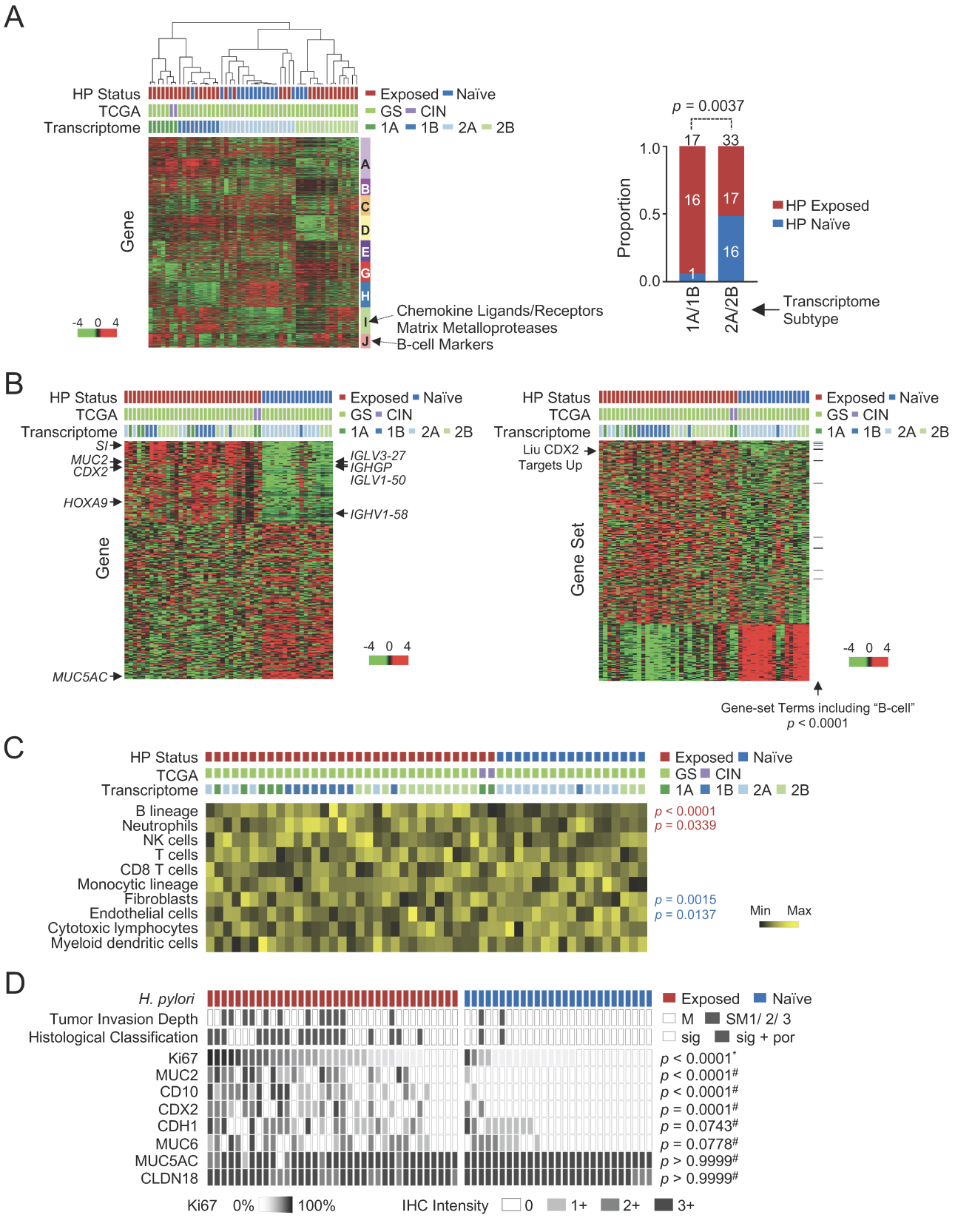


Figure 5 Legend on next page.

Table 3. Summary of the genomic, transcriptomic, and immunohistochemical features.

Features	HP-exposed DGC versus HP-naïve DGC
SNV burden	Elevated in HP-exposed DGCs*
RHOA-ARHGAP misregulation	Present versus absent
CAA (7p, 7q, 8p, 8q gain, or 16q loss)	Frequently aberrated in HP-exposed DGCs*
B-lineage infiltration	Present versus absent
Proliferative (Ki-67 expression)	Elevated in HP-exposed DGCs*
Marker expression	Mixed gastric and intestinal versus Pure gastric

CAA, chromosome arm aneuploidy; DGC, diffuse gastric carcinoma; HP, *Helicobacter pylori*; SNV, single nucleotide variant.
*Observed in HP-exposed DGCs relative to HP-naïve DGCs.

mutations in the RHOA–ARHGAP signaling pathway. We detected significant differences in intestinal-type and gastric-type marker gene expression between HP-exposed and HP-naïve DGCs (Table 3) by assessing the transcriptomes and using immunohistochemistry.

Several previous genomic studies of DGCs provided data pertaining to HP infection status, including TCGA [13,14,18,34]; albeit, the definition of HP infection status tended to be based on the presence or absence of the pathogen, as identified through the detection of bacterial bodies [13] or genomes [14,18,34], which are of low sensitivity. In addition, these tests cannot be used to distinguish past infection [5]. Indeed, given that HP-associated gastric cancer can arise even after the disappearance of HP infection [32], the absence of bacteria does not indicate a "never exposed" status [5], and thus the data from such cases are not suitable for studying the distinctive features of HP-exposed and HP-naïve gastric cancer genomes. Recently, two studies by Kiso *et al* and Nikaido *et al* were published, both of which focused on genomic alterations in HP-naïve gastric cancer [35,36]. In both studies, subjects were selected under comparably stringent criteria for HP infection status to those used in the current study. However, the detected frequency of driver mutations was very low in the former study [35], and HP-exposed DGCs were not exome-sequenced in the latter study [36]; these differences in study findings and design hinder us from making a direct comparison with our data.

We noted a higher number of mutated driver genes among HP-exposed DGCs, with the exception of *CDH1*; this higher number may simply reflect a larger somatic SNV burden. Nevertheless, the biased distribution of RHOA-ARHGAP pathway misregulation among our HP-exposed DGCs may indicate a growth advantage provided by these aberrations. CAA was also recognized as a salient characteristic of HP-exposed DGCs. CAA is considered to be caused by chromosome segregation errors [37], which can be derived from mitotic impairment by CagA, a *Helicobacter*-produced virulence factor [38], or by deregulated activation of the RHOA-ARHGAP signaling pathway [39].

Using exome analysis, we detected no significantly recurrent driver event in HP-naïve DGCs aside from *CDH1* mutation; this extremely stable genomic feature of the tumor may result in the indolent phenotype of HP-naïve DGCs, as documented elsewhere [36]. Alternatively, HP-naïve DGCs may be driven by epigenetic or noncoding drivers, which were unable to be identified in the current study. We therefore consider that further genomic and epigenomic studies using whole-genome sequencing, along with other advanced technologies, such as spatial epigenomics/transcriptomics, in combination with organoid/xenograft models, are warranted to aid in identifying the driver event(s) in the tumorigenic process of HP-naïve DGCs. Such driver event(s) will be useful biomarker(s) for the early detection and prevention of sporadic DGCs.

Figure 5. Distinct gene expression patterns between HP-exposed and HP-naïve DGCs. (A) Gene expression heatmap and proportions of HP-exposed and HP-naïve samples for transcriptomic subtypes. Left panel. Gene expression heatmap of two major sample clusters revealed by consensus clustering. Green and red colors indicate over- and underexpression of the genes with highly variable expression across the samples in the cohort, respectively. Sample clusters are shown with a dendrogram, and as color codes for status of HP infection, TCGA subtype, and transcriptomic subtype. Transcriptomic subtypes are designated 1A, 1B, 2A, and 2B according to the dendrogram. The labels for gene clusters are also shown on the right with characteristic gene components in gene clusters I and J. Proportions of HP-exposed and HP-naïve DGCs for transcriptomic subtypes. Sample number is shown on the stacked bar plots. The *P* value was computed by Fisher's exact test. (B) Genes (left panel) and gene sets (right panel) between HP-exposed and HP-naïve DGCs are shown as heatmaps. Green and red colors indicate over- and underexpression, respectively, of the genes or gene sets. Gene symbols indicated on the left to the gene heatmap are *CDX2*, its downstream genes (*SI*, *MUC2*, and *HOXA9*), and *MUC5AC*. Immunoglobulin-component related genes are shown to the right of the gene heatmap. The position of 'Liu CDX2 Targets Up' is shown to the left of the gene-set heatmap. Ticks marked along the right side of the gene-set map indicate the positions for gene-set annotations that include the term 'B-cell.' The *P* value was computed by hypergeometric test. (C) Heatmap for MCPcounter scores. *P* values were computed using Mann–Whitney *U*-tests to compare the scores between HP-exposed and HP-naïve DGCs. Scores for which there was a significant difference are indicated with *P* values in red font (HP-exposed > HP-naïve) or blue font (HP-exposed < HP-naïve). Sample labeling for HP-infection status, TCGA molecular subtype, and transcriptomic subtype is shown above the heatmaps. (D) Immunohistochemical analyses of HP-exposed and HP-naïve DGCs. Protein expression levels are shown as a heatmap. We also indicate the tumor invasion depth (M versus SM1/2/3) and histological classification (sig versus sig + por). The level of Ki-67 protein expression was evaluated by percentage; the levels of MUC2, CD10, CDX2, MUC6, MUC5AC, and CLDN18 were evaluated using a four-level scale (0, 1+, 2+, and 3+, as indicated). Samples were sorted according to HP-infection status and Ki-67 expression level. The *P* value for Ki-67 was computed using a Mann–Whitney *U*-test (marked with *); *P* values for other proteins were computed by Fisher's exact test (marked with #), as shown on the right. M; mucosa, SM; submucosa, sig; signet-ring cell carcinoma, and por; poorly differentiated adenocarcinoma.

Acknowledgements

The authors thank Kazuma Kiyotani, Toru Hirota, Yasuo Uemura, Koichiro Inaki, and Yusuke Nakamura for helpful discussions. The authors also thank Megumi Nakai, Rika Nishiko, Junko Kanayama, Nobuyuki Fukui, Akihisa Takahara, Shohei Konno, Taichi Hagio, Noriko Yaguchi, Satoko Baba, Motoyoshi Iwakoshi, Tomoyo Kakita, and Kumiko Sakurai for technical assistance; Minako Hoshida for administrative assistance; and Rebecca Jackson for editing a draft of this article as a commercial service.

This work was supported by JSPS KAKENHI; Grant Numbers JP21K15393 (MT), JP21H02793 (SM), JP20K09634 (OG), JP18K07338 (SM), JP17K18337 (OG), and JP15K06861 (SM), Princess Takamatsu Cancer Research Fund; Grant Number 11-24317 (SM), and Takeda Japan Medical Office Funded Research Grant 2018 (JF).

Author contributions statement

KN, NT, YO, IN, MK and YI performed the analyses and wrote the article. MT performed the histopathological diagnoses, the analyses, and wrote the article. SA and RF performed wet experiments. OG and TK performed informatics analyses. YT, SY, YH, AI, TY and TH collected the specimens and provided clinical information. TN, JF and SM conceived the study and wrote the article.

Data availability statement

The data generated in this study have been submitted to the National Bioscience Database Center (NBDC; https://biosciencedbc.jp/en/) and NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession numbers hum0338 (exome BAM files [raw data] are available through controlled-access of NBDC) and GSE129219 (RNA-seq TPM file; the raw data will not be available through controlled-access resources such as dbGaP or EGA).

References

1. Lauren P. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. *Acta Pathol Microbiol Scand* 1965; **64:** 31–49.
2. 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.
3. Parsonnet J, Vandersteen D, Goates J, *et al*. *Helicobacter pylori* infection in intestinal- and diffuse-type gastric adenocarcinomas. *J Natl Cancer Inst* 1991; **83:** 640–643.
4. Wadhwa R, Song S, Lee JS, *et al*. Gastric cancer-molecular and clinical dimensions. *Nat Rev Clin Oncol* 2013; **10:** 643–655.
5. Yamamoto Y, Fujisaki J, Omae M, *et al*. *Helicobacter pylori*-negative gastric cancer: characteristics and endoscopic findings. *Dig Endosc* 2015; **27:** 551–561.
6. Hooi JKY, Lai WY, Ng WK, *et al*. Global prevalence of *Helicobacter pylori* infection: systematic review and meta-analysis. *Gastroenterology* 2017; **153:** 420–429.
7. Tanaka S, Goto A, Yamagishi K, *et al*. Long-term response of *Helicobacter pylori* antibody titer after eradication treatment in middle-aged Japanese: JPHC-NEXT Study. *J Epidemiol* 2021. https://doi.org/10.2188/jea.JE20200618.
8. Tokai Y, Fujisaki J, Ishizuka N, *et al*. Usefulness of the l-type Wako *Helicobacter pylori* antibody J test. *JGH Open* 2021; **5:** 673–678.
9. Di Mario F, Cavallaro LG, Moussa AM, *et al*. Usefulness of serum pepsinogens in *Helicobacter pylori* chronic gastritis: relationship with inflammation, activity, and density of the bacterium. *Dig Dis Sci* 2006; **51:** 1791–1795.
10. Ferwana M, Abdulmajeed I, Alhajiahmed A, *et al*. Accuracy of urea breath test in *Helicobacter pylori* infection: meta-analysis. *World J Gastroenterol* 2015; **21:** 1305–1314.
11. Horiuchi Y, Fujisaki J, Yamamoto N, *et al*. Biological behavior of the intramucosal *Helicobacter pylori*-negative undifferentiated-type early gastric cancer: comparison with *Helicobacter pylori*-positive early gastric cancer. *Gastric Cancer* 2016; **19:** 160–165.
12. Matsuo T, Ito M, Takata S, *et al*. Low prevalence of *Helicobacter pylori*-negative gastric cancer among Japanese. *Helicobacter* 2011; **16:** 415–419.
13. Wang K, Yuen ST, Xu J, *et al*. Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat Genet* 2014; **46:** 573–582.
14. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 2014; **513:** 202–209.
15. Kakiuchi M, Nishizawa T, Ueda H, *et al*. Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. *Nat Genet* 2014; **46:** 583–587.
16. Wong SS, Kim KM, Ting JC, *et al*. Genomic landscape and genetic heterogeneity in gastric adenocarcinoma revealed by whole-genome sequencing. *Nat Commun* 2014; **5:** 5477.
17. Cristescu R, Lee J, Nebozhyn M, *et al*. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat Med* 2015; **21:** 449–456.
18. Cho SY, Park JW, Liu Y, *et al*. Sporadic early-onset diffuse gastric cancers have high frequency of somatic CDH1 alterations, but low frequency of somatic RHOA mutations compared with late-onset cancers. *Gastroenterology* 2017; **153:** 536–549.e26.
19. Ge S, Li B, Li Y, *et al*. Genomic alterations in advanced gastric cancer endoscopic biopsy samples using targeted next-generation sequencing. *Am J Cancer Res* 2017; **7:** 1540–1553.
20. Shu Y, Zhang W, Hou Q, *et al*. Prognostic significance of frequent CLDN18-ARHGAP26/6 fusion in gastric signet-ring cell cancer. *Nat Commun* 2018; **9:** 2447.
21. Kwon CH, Kim YK, Lee S, *et al*. Gastric poorly cohesive carcinoma: a correlative study of mutational signatures and prognostic significance based on histopathological subtypes. *Histopathology* 2018; **72:** 556–568.
22. Nemtsova MV, Kalinkin AI, Kuznetsova EB, *et al*. Clinical relevance of somatic mutations in main driver genes detected in gastric cancer patients by next-generation DNA sequencing. *Sci Rep* 2020; **10:** 504.
23. Wang R, Song S, Harada K, *et al*. Multiplex profiling of peritoneal metastases from gastric adenocarcinoma identified novel targets and molecular subtypes that predict treatment response. *Gut* 2020; **69:** 18–31.
24. Yeoh KG, Tan P. Mapping the genomic diaspora of gastric cancer. *Nat Rev Cancer* 2022; **22:** 71–84.

25. Japanese Gastric Cancer Association. Japanese classification of gastric carcinoma: 3rd English edition. *Gastric Cancer* 2011; **14:** 101–112.
26. Adamu MA, Weck MN, Gao L, *et al*. Incidence of chronic atrophic gastritis: systematic review and meta-analysis of follow-up studies. *Eur J Epidemiol* 2010; **25:** 439–448.
27. Craanen ME, Dekker W, Blok P, *et al*. Intestinal metaplasia and *Helicobacter pylori*: an endoscopic bioptic study of the gastric antrum. *Gut* 1992; **33:** 16–20.
28. Yagi K, Honda H, Yang JM, *et al*. Magnifying endoscopy in gastritis of the corpus. *Endoscopy* 2005; **37:** 660–666.
29. Horiuchi Y, Fujisaki J, Ishizuka N, *et al*. Study on clinical factors involved in *Helicobacter pylori*-uninfected, undifferentiated-type early gastric cancer. *Digestion* 2017; **96:** 213–219.
30. Mermel CH, Schumacher SE, Hill B, *et al*. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* 2011; **12:** R41.
31. Liu T, Zhang X, So CK, *et al*. Regulation of Cdx2 expression by promoter methylation, and effects of Cdx2 transfection on morphology and gene expression of human esophageal epithelial cells. *Carcinogenesis* 2007; **28:** 488–496.
32. Piscione M, Mazzone M, Di Marcantonio MC, *et al*. Eradication of *Helicobacter pylori* and gastric cancer: a controversial relationship. *Front Microbiol* 2021; **12:** 630852.
33. Becht E, Giraldo NA, Lacroix L, *et al*. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol* 2016; **17:** 218.
34. Zhang X, Liu F, Bao H, *et al*. Distinct genomic profile in *h. pylori*-associated gastric cancer. *Cancer Med* 2021; **10:** 2461–2469.
35. Kiso M, Urabe Y, Ito M, *et al*. Clinical and genomic characteristics of mucosal signet-ring cell carcinoma in *Helicobacter pylori*-uninfected stomach. *BMC Gastroenterol* 2020; **20:** 243.
36. Nikaido M, Kakiuchi N, Miyamoto S, *et al*. Indolent feature of *Helicobacter pylori*-uninfected intramucosal signet ring cell carcinomas with CDH1 mutations. *Gastric Cancer* 2021; **24:** 1102–1114.
37. Soto M, Raaijmakers JA, Medema RH. Consequences of genomic diversification induced by segregation errors. *Trends Genet* 2019; **35:** 279–291.
38. Umeda M, Murata-Kamiya N, Saito Y, *et al*. *Helicobacter pylori* CagA causes mitotic impairment and induces chromosomal instability. *J Biol Chem* 2009; **284:** 22166–22172.
39. Maddox AS, Burridge K. RhoA is required for cortical retraction and rigidity during mitotic cell rounding. *J Cell Biol* 2003; **160:** 255–265.
40. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011; **12:** 323.
41. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25:** 1754–1760.
42. DePristo MA, Banks E, Poplin R, *et al*. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; **43:** 491–498.
43. Koboldt DC, Zhang Q, Larson DE, *et al*. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 2012; **22:** 568–576.
44. Cibulskis K, Lawrence MS, Carter SL, *et al*. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013; **31:** 213–219.
45. Ueda H. Karkinos. [Accessed 26 August 2022]. Available from: http://sourceforge.net/projects/karkinos/.
46. Magi A, Tattini L, Cifola I, *et al*. EXCAVATOR: detecting copy number variants from whole-exome sequencing data. *Genome Biol* 2013; **14:** R120.
47. Tanaka N, Takahara A, Hagio T, *et al*. Sequencing artifacts derived from a library preparation method using enzymatic fragmentation. *PLoS One* 2020; **15:** e0227427.
48. Kato M, Nakamura H, Nagai M, *et al*. A computational tool to detect DNA alterations tailored to formalin-fixed paraffin-embedded samples in cancer clinical sequencing. *Genome Med* 2018; **10:** 44.
49. Park H, Cho SY, Kim H, *et al*. Genomic alterations in BCL2L1 and DLC1 contribute to drug sensitivity in gastric cancer. *Proc Natl Acad Sci U S A* 2015; **112:** 12492–12497.
50. Shen R, Seshan VE. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res* 2016; **44:** e131.
51. Alexandrov LB, Kim J, Haradhvala NJ, *et al*. The repertoire of mutational signatures in human cancer. *Nature* 2020; **578:** 94–101.
52. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 2010; **26:** 1572–1573.
53. Takeuchi K, Soda M, Togashi Y, *et al*. RET, ROS1 and ALK fusions in lung cancer. *Nat Med* 2012; **18:** 378–381.
54. McPherson A, Hormozdiari F, Zayed A, *et al*. deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. *PLoS Comput Biol* 2011; **7:** e1001138.
55. Kim D, Salzberg SL. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol* 2011; **12:** R72.
56. Shiraishi Y, Okada A, Ohta S, Chiba K. fusionfusion. [Accessed 26 August 2022]. Available from: https://github.com/Genomon-Project/fusionfusion.
57. Danecek P, Bonfield JK, Liddle J, *et al*. Twelve years of SAMtools and BCFtools. *Gigascience* 2021; **10:** giab008.

References 40–57 are cited only in the supplementary material.

SUPPLEMENTARY MATERIAL ONLINE

Supplementary materials and methods

Figure S1. Immunohistochemical analyses of additional HP-exposed and HP-naïve DGC samples in the validation cohort

Table S1. Clinicopathological features of each patient

Table S2. Sample information

Table S3. List of gastric cancer driver genes

Table S4. Association of clinicopathological variables with RHOA–ARHGAP signaling misregulation computed by Fisher's exact test

Table S5. Association of clinicopathological variables with chromosome arm aneuploidies

Table S6. Association of clinicopathological variables with immunohistochemical characteristics

Table S7. Subgroup analyses of intestinal or gastric marker profiles regarding HP infection status