



UNIVERSITY OF LEEDS

This is a repository copy of *Accumulation of somatic mutations in TP53 in gastric epithelium with Helicobacter pylori infection.*

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/86682/>

Version: Accepted Version

Article:

Shimizu, T, Marusawa, H, Matsumoto, Y et al. (9 more authors) (2014) Accumulation of somatic mutations in TP53 in gastric epithelium with Helicobacter pylori infection. *Gastroenterology*, 147 (2). 407 - 417.e3. ISSN 0016-5085

<https://doi.org/10.1053/j.gastro.2014.04.036>

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

**Accumulation of Somatic Mutations in TP53 in Gastric Epithelium
with Helicobacter pylori infection.**

Takahiro Shimizu¹, Hiroyuki Marusawa¹, Yuko Matsumoto¹, Tadashi Inuzuka¹,

Atsuyuki Ikeda¹, Yosuke Fujii¹, Sachiko Minamiguchi², Shin'ichi Miyamoto¹,

Tadayuki Kou³, Yoshiharu Sakai⁴, Jean E Crabtree⁵ and Tsutomu Chiba¹

1. Department of Gastroenterology and Hepatology,
Graduate School of Medicine, Kyoto University, Kyoto, Japan
2. Department of Diagnostic Pathology, Graduate School of Medicine,
Kyoto University, Kyoto, Japan
3. Digestive Disease Center, The Tazuke Kofukai Medical Research Institute,
Kitano Hospital, Osaka, Japan
4. Department of Gastrointestinal Surgery, Graduate School of Medicine,
Kyoto University, Kyoto, Japan
5. Leeds Institute Molecular Medicine, University of Leeds, Leeds, United Kingdom

Corresponding & Reprint Author: Hiroyuki Marusawa at:

Department of Gastroenterology and Hepatology,
Graduate School of Medicine, Kyoto University,
54 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan
Phone; +81-75-751-4319, Fax; +81-75-751-4303
E-mail; maru@kuhp.kyoto-u.ac.jp

Short Title: Mutations in *H. pylori* (+) inflamed gastric mucosa.

Grant Support: This work was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI (25130706, 24229005 and 24659363), and Grants-in-aid for Scientific Research on Innovative Areas from the Ministry of Education, Culture, Sports Science and Technology

(MEXT) of Japan.

Abbreviations: AID: activation-induced cytidine deaminase, H. pylori: Helicobacter pylori, Hupki: human TP53 knock-in, indel: insertion and deletion, MSI: microsatellite instability, MSS: microsatellite stability, SNV: single nucleotide variant

Disclosures: The authors have no conflicts of interest.

Data Profiling: Sequence reads were deposited in the DNA Data Bank of Japan Sequence Read Archive (http://trace.ddbj.nig.ac.jp/dra/index_e.shtml) under accession numbers DRA001035.

Author contributions: T.S. conducted all the experiments and contributed to manuscript preparation. H.M. designed the study, supervised the project and contributed to manuscript preparation. Y.M., T.I. and A.I. performed deep sequencing analyses. Y.F. performed the bioinformatics data analysis. S.M. performed histological examination. S.M., T.K. and Y.S. collected human samples. J.E.C. and T.C. supervised the project.

Abstract

Background & Aims: *Helicobacter pylori* infection is a risk factor for gastric cancer. To explore the genetic basis of gastric cancer that develops in inflamed gastric mucosa, we investigated genetic aberrations that latently accumulate in non-tumorous gastric epithelium with *H pylori* infection.

Methods: We performed whole exome sequencing of gastric tumors, non-cancerous tissues with gastritis, and peripheral lymphocytes from 5 patients. We performed additional deep-sequencing analyses of selected tumor-related genes using 34 gastritis mucosal samples from patients with or without gastric cancer. We also performed deep sequencing analyses of gastric mucosal tissues from mice that express transgenic human TP53 and constitutively express activation-induced cytidine deaminase (AICDA or AID) (Hupki/AID-Tg mice).

Results: Whole-exome sequencing revealed that somatic mutations accumulated in various genes in inflamed gastric tissues. Additional deep sequencing analyses of tissues from regions of gastritis confirmed non-synonymous low-abundance mutations in TP53 in 15 cases (44.1%) and ARID1A in 5 cases (14.7%). The mutations that accumulated in gastric mucosal tissues with *H pylori*-induced gastritis, as well as gastric tumors, were predominantly C:G>T:A transitions in GpCpX motifs—a marker of cytidine deamination induced by AID. Constitutive expression of AID in the gastric mucosa of mice led to mutations in the human TP53, at amino acid coding positions identical to those detected in human gastric cancers.

Conclusions: Studies of gastric tumors and tissues from humans and mice indicate that somatic mutations accumulate in various genes in gastric mucosal tissues with *H pylori* infection. Increased cytidine deaminase activity in these tissues appears to promote the accumulation of these mutations and might promote gastric carcinogenesis in patients with *H pylori* infection.

Keywords: stomach cancer; somatic hypermutation; pathogenesis; bacteria

Introduction

The recent introduction of deep sequencing technology, capable of detecting sequences of hundreds of thousands of individual molecules, is rapidly changing the landscape of cancer genome research. Comprehensive genome analyses using the deep sequencing technique have clarified that each tumor cell has numerous nucleotide alterations¹. Among the genetic aberrations, driver mutations are thought to contribute to tumorigenesis by conferring a selective advantage to the cell. On the other hand, passenger mutations are defined as those that do not alter fitness but occur in a cell that coincidentally, or subsequently, acquired a driver mutation². Recent reports demonstrated that human gastric cancer has approximately 66 to 212 mutations in coding regions, and only a few of them are considered driver mutations³⁻⁶. Among them, mutations in the tumor suppressor TP53 gene are the most common driver mutations in human gastric cancer tissues. A second common putative driver mutation was determined in the chromatin remodeling complexes gene, ARID1A³. It remains unclear, however, when and how those driver mutations emerge in the gastric epithelium during tumorigenesis.

Epidemiologic studies revealed that patients with *Helicobacter pylori* (*H pylori*) infection have a significantly increased risk of gastric cancer compared to *H pylori*-negative individuals^{7,8}. The more prevalent form of gastric adenocarcinoma frequently progresses through a series of histologic steps that are initiated by chronic gastritis, which then leads to atrophic gastritis and intestinal metaplasia, and finally to dysplasia and adenocarcinoma in individuals with *H pylori* infection⁹. Multicentric tumor development characterized in *H pylori*-associated gastric cancer¹⁰ suggests that the *H pylori*-infected stomach could include epithelial cells and/or intestinal metaplasia with genetic damage that possess high susceptibility to tumor development¹¹. Consistent with these findings, animal experiments using Big Blue mice revealed a high mutation frequency in the *H pylori*-infected gastric mucosa¹², which are increased in

Helicobacter-infected Big Blue-p53 hemizygous mice¹³. The precise mechanism for the emergence of genetic alterations in H pylori-infected gastric mucosa, however, has not been elucidated. We recently demonstrated that H pylori infection and the resultant chronic inflammation with mononuclear cell infiltration induces genetic alterations, including somatic mutations and chromosomal aberrations, in gastric epithelial cells through the expression of a DNA-mutator enzyme, activation-induced cytidine deaminase (AID)¹⁴⁻¹⁷. In vitro studies revealed that H pylori infection elicited the ectopic expression of AID in association with a high mutation frequency of the TP53 genome sequences¹⁴. To date, however, there has not been enough evidence showing that AID expression in gastric mucosa in vivo causes the accumulation of driver mutations in the human TP53 sequences required for malignant transformation.

To clarify the mechanisms underlying the contribution of inflammation to carcinogenesis, it is important to identify the genetic alterations that occur in the inflamed epithelium before the onset of tumorigenesis. The main obstacle to detecting latently accumulated genetic alterations in non-tumorous inflamed tissue is the diversity and low frequency of the mutated genes¹⁸. In the present study, to unveil the landscape of accumulated genetic aberrations in chronically inflamed gastric epithelium, we performed whole exome sequencing and additional deep sequencing of the inflamed gastric mucosa with chronic H pylori infection. Furthermore, to confirm whether constitutive AID expression in vivo contributes to the emergence of the TP53 mutations, we determined the mutation profile of the TP53 gene in the AID-expressing gastric mucosa of the human TP53 knock-in mouse model.

Materials and methods

Whole exome capture and massively-parallel sequencing.

Massively-parallel sequencing was performed as described previously^{19, 20}. Fragmented DNA (>5 µg) was used to prepare each DNA sequencing library. The DNA libraries were prepared according to the instructions provided with the Illumina Preparation Kit (Illumina, San Diego, CA). Whole exome sequence capture was then performed using SeqCap EZ Human Exome Library v2.0 (Roche, Madison, WI) according to the manufacturer's instructions. Cluster generation was performed on the Illumina cluster station using their TruSeq PE Cluster Kit v5. Paired-end sequencing for 2 × 76 base pair (bp) reads was performed on the Illumina Genome Analyzer Iix using their SBS Kits v5. Data collection and base-calling were performed using SCS v2.9/RTA 1.9 and resultant data files were converted to the FASTQ format.

Sequencing data were analyzed with NextGENe 2nd Generation Sequence Analysis Software v2.2.0 (SoftGenetics, State College, PA). We identified somatic mutations by the strict variant filtering process (Supplementary Figure 1 and Supplementary materials and methods). To validate somatic mutations, we performed PCR-based deep-sequencing using a different deep-sequencing platform (the GS Junior System, Roche) on the non-silent mutations detected by whole exome sequencing, according to the validation protocol reported in the recent study²¹. We estimated that the reliabilities of whole exome sequencing were 88.6% for somatic mutations detected in the tumor tissues and 86.4% for somatic mutations detected in the non-tumor tissues (data not shown). Primer sequences are shown in Supplementary Table 1.

Deep sequencing on selected genes.

The mutational status of TP53, ARID1A, MLL2, SUV39H1, CTNNB1, PIK3CA, and TP53 cDNA was investigated by high-coverage sequencing analyses. Target regions were designed within the range from 330 to 520 bp. The primer sequences are described in Supplementary

Table 2. Each region was PCR-amplified with Phusion High-Fidelity PCR Enzyme (FINZYMES, Espoo, Finland), purified by gel-extraction methods. Five hundred nanograms of each sample was dA-tailed and ligated with adaptors containing tag sequences, followed by emulsion PCR and sequencing using the GS Junior System (Roche) according to the manufacturer's protocol.

Deep sequencing data were analyzed with NextGENe software. We identified low-abundance somatic mutations by the strict variant filtering process (Supplementary Figure 2 and Supplementary materials and methods). Candidate low-abundance mutations were validated by repeated deep sequencing using independent amplicons derived from the same samples.

Tissue specimens and DNA extraction.

Animal Experiments.

Whole exome sequencing analysis and variant filtering.

Deep sequencing analysis and variant filtering.

Functional prediction analysis and pathway analysis.

Statistical analysis.

These procedures and information are described in the Supplementary Materials and Methods.

Results

Landscape of genetic alterations that accumulated in H pylori-associated gastric cancers.

To overview the landscape of genetic alterations that accumulated in both gastric cancer and the adjacent background inflamed mucosa with H pylori infection, we performed whole exome sequencing of 5 matched pairs of gastric cancer and non-cancerous gastritis tissues in patients with H pylori infection. To subtract the normal variants of each individual from the somatic mutations, we also determined the whole exome sequences of the matched peripheral lymphocytes in each patient. Microsatellite analysis revealed that one gastric cancer had microsatellite instability (MSI) and the remaining four tumors had the characteristics of microsatellite stability (MSS). All background gastric mucosa had the features of chronic gastritis with mucosal atrophy and intestinal metaplasia and none of the gastric specimens included dysplastic lesions (Supplementary Table 3).

We targeted the whole exons of approximately 20,000 genes, sequenced 3.33Gb on average for each sample, and achieved 41.2-fold coverage per samples as the mean coverage of each base in the target regions (Supplementary Table 4). According to the variant filtering process (Supplementary Figure 1), we identified single nucleotide variants (SNVs), short insertions and deletions (indels) in cancer tissues as well as in the non-cancerous gastritis mucosa.

We identified 892 somatic mutations in 793 genes in the 5 gastric cancer tissues. The identified mutations included 654 non-silent mutations (552 missense, 25 nonsense, 1 stop codon loss, 11 essential splice site mutations, and 65 indels) and 238 synonymous mutations (Supplementary Tables 5, 6). The number of non-silent mutations in the MSI gastric cancer sample was significantly higher than those of any of the MSS tumor samples (Figure 1A). In addition, the MSI tumor also had a large number of indels, while the MSS tumors had only a few indels (Figure 1C, Supplementary Table 5). Furthermore, the mutation frequency at each

nucleotide position in the MSI tumor tended to be higher than those in the MSS tumors (Figure 1B). These findings indicated that MSI gastric cancers have enhanced genetic aberrations compared with MSS tumors.

Mutation profiles of MSS and MSI gastric cancer tissues.

Among the mutated genes identified, 26 genes were recurrently mutated as non-synonymous mutations or indels in two or more gastric cancers (Table 1). These genes included TP53 (mutated in 3 of 5 cases), and ARID1A (mutated in 2 of 5 cases), mutations of which were mutually exclusive^{3,4}. All of the TP53 mutations were present as a type of nucleotide substitution in three MSS tumor tissues, whereas the ARID1A mutations were detected as deletions in the one remaining MSS and in one MSI gastric cancer. Pathway analyses in the five gastric cancers revealed that cell adhesion⁴ was the most significantly enriched biological process, followed by metabolic processes, endocytosis and extracellular structure organization (Supplementary Table 7).

Recent studies showed that the pattern of nucleotide alterations could provide a clue to determine the molecular process underlying mutation occurrence²²⁻²⁷. We first confirmed that over half of the SNVs that accumulated in the exome sequences of gastric cancer tissues were enriched with a C:G>T:A transition ($p < 0.01$ by ANOVA, Figure 1C). The prevalence of these C:G>T:A transitions in the MSI gastric cancer was substantially higher than that in the MSS tumors (65.1% vs. 48.9%, respectively). These mutational signatures suggested that cytidine to uracil deamination is deeply implicated in the mutational process underlying H pylori-associated gastric carcinogenesis. Among the cytidine deaminase family proteins, AID exhibits a strong preference for deaminating C residues flanked by a 5'-purine (G or A)^{22, 28, 29}. In contrast, APOBEC3 family enzymes and APOBEC1 favor C residues flanked by 5'-T^{30, 31}. Therefore, we

further evaluated the sequence characteristics flanking the mutated C residues in tumor tissues. Consistent with the previous report⁵, we found that the GpCpX (C is the mutated nucleotide and X is any base, GpCpA: 23.9%, GpCpC: 24.3%, GpCpG: 38.4%, GpCpT: 13.4%) pattern dominated across all of the tissues examined (p<0.01 by ANOVA), followed by CpCpX and ApCpX (Figure 1D), suggesting that deamination achieved by AID activity is likely to be involved in the mutational signature of H pylori-associated gastric cancers.

Overview of genetic alterations that accumulated in non-cancerous gastritis mucosa with H pylori infection.

Whole exome sequencing clarified that various mutations were also present in the non-cancerous H pylori-infected gastritis mucosa. All of the mutations detected in the non-tumorous gastric mucosa were SNVs and no indels were observed in any of the non-tumorous gastritis tissues. The mean number of mutated genes in the non-tumorous tissues was 28.4 per sample, which was lower than that in the matched tumor tissues of all 5 patients (Supplementary Tables 5, 8). In addition, the mutation frequency at each nucleotide position represented less than 20% of the total reads of the majority of the altered genes in non-tumorous gastric mucosa (Figure 1B), indicating that the mutation frequency at each nucleotide position tended to be lower in the non-cancerous tissues than in the cancer tissues. The total number of mutated genes and mutant allele frequencies in the non-tumorous tissues did not significantly differ between the patient with MSI and those with MSS cancers (Figure 1A, 1B, Supplementary Table 5).

Among the 142 mutations identified in the 5 non-tumorous gastritis tissues, 91 mutations (64.1%) were non-silent. Sorting Intolerant From Tolerant (SIFT) functional impact predictions³² revealed that the mean percentages of somatic mutations predicted to be “damaging” or

frame-shift, nonsense and splice site mutations in non-tumorous gastritis and cancer tissues were 29.6% and 44.7%, respectively ($p < 0.01$ by χ^2 test), suggesting that the SNVs that accumulated in the non-tumorous tissues more frequently included “passenger” mutations with less functional significance than those in the tumor tissues (data not shown). Pathway analyses revealed that transcription-related genes, such as MLL2 and SUV39H1 that are involved in chromatin or histone modification, were frequently altered in non-tumorous gastritis tissues (Supplementary Table 9). Whole exome sequencing also revealed 25 genes that were commonly mutated in both the tumor and the matched background gastritis mucosa of the same patients (Supplementary Tables 10 and 11). Among them, 16 genes acquired non-synonymous mutations and Sanger sequencing confirmed that those SNVs were not detected in the matched control lymphocytes, while the biological role of these mutated genes remains unclear.

Establishment of deep sequencing platform for detecting low-abundance mutations in tumor-related genes.

Whole exome sequencing clarified that somatic mutations accumulated in various genes in H pylori-infected gastritis mucosa, while the majority of the mutated genes detected were supposed to be “passenger” mutations. Those findings prompted us to examine whether somatic mutations latently accumulate in “driver” genes that could contribute to gastric tumorigenesis, and thus deep sequencing analysis of the selected amplicons of tumor-related genes were applied to the non-tumorous gastric mucosa and the matched peripheral lymphocytes from H pylori (+) patients with and without gastric cancers.

First, we conducted control experiments to validate the efficacy and error rates in deep sequencing using a plasmid encoding the TP53 cDNA sequence as a template. The deep sequencing platform provided us the TP53 sequences derived from the plasmids with a mean

coverage of 4024 at each nucleotide site and the error rate per nucleotide position was below 0.07%. To estimate the accuracy of detecting nucleotide alterations using reads filtered by mean base quality and mapping quality, we introduced expression plasmids with a single point mutation within this wild-type sequence with a ratio of 1%(1:100), 0.1%(1:1000), or 0.01%(1:10000). Duplicate control experiments revealed that mutations present at an input ratio of 0.1% were detected ranging between 0.11% and 0.12% (Supplementary Table 12). Based on these results, the limit of mutant allele frequencies detectable by this deep sequencing was considered to be more than 0.1% of the total reads. Finally, to exclude the errors during the sequencing process, repeated experiments were performed and we determined that nucleotide alterations validated by at least more than two different deep sequencings were low-abundance somatic mutations.

Somatic mutations accumulated in tumor-related genes in H pylori infection-related gastritis mucosa.

We performed deep sequencing on the selected amplicons of tumor-related genes in H pylori-related gastritis mucosa in 28 patients with gastric cancer and 6 patients without gastric cancer (Supplementary Table 13). Consistent with the previous studies^{3, 4}, our whole exome sequencing on 5 tumor tissues identified that TP53 and ARID1A were frequently mutated in gastric cancer tissues. In addition, the previous studies clarified that CTNNB1 and PIK3CA are also frequently mutated in gastric cancer tissues^{3, 4}. Based on those findings, we selected these four genes (TP53, ARID1A, CTNNB1 and PIK3CA) for further deep-sequencing analyses. In addition, we targeted MLL2 and SUV39H1, which were identified by current whole exome sequencing in non-cancerous gastritis tissues. The mean coverage ranged from 2319 to 12,934 fold at each targeted gene.

We first verified that no nucleotide alterations were detected in normal gastric mucosa without *H pylori* infection or any lymphocytes of the enrolled patients. Deep sequencing on 34 gastritis mucosa revealed that a lot of non-synonymous mutations were present at a very low frequency (0.11-1.12%) in various genes ($p < 0.05$ by χ^2 test). We validated the results by repeated deep sequencings and confirmed that 83.7% of the nucleotide changes detected by the initial analyses were also detected by the independent deep sequencing analyses (Supplementary Table 14). In the gastritis mucosa of 28 patients with gastric cancer, we identified non-synonymous low-abundance mutations in 13 nucleotide positions of TP53 in 11 cases, 6 positions of ARID1A in 4 cases, and 3 positions of MLL2 in 2 cases, while no mutations were detected in CTNNB1, PIK3CA, and SUV39H1 (Table 2). In addition, in the gastritis mucosa of 6 patients without gastric cancer, non-synonymous low-abundance mutations were detected in 6 positions of TP53 in 4 cases and in 1 position of ARID1A in 1 case (Table 3).

Mutational status in non-cancerous gastritis mucosa was not significantly associated with the stage, differentiation, MSI status, or mutational status of the corresponding gastric cancers (data not shown). Regarding the mutation signature in non-tumorous gastritis mucosa, deaminating C residues flanked by a 5'-purine (G or A) were strongly preferred, similar to that in gastric cancer tissues. Indeed, 15 of 19 (78.9%) low-abundance TP53 mutations in *H pylori*-related gastritis mucosa with or without gastric cancer were C:G>T:A transitions ($p < 0.01$ by χ^2 test) and all of these transitions accumulated in the context of GpCpX (Figure 2A). Numbers of the mutated TP53 nucleotide positions observed in the gastritis mucosa were identical to those that accumulated in human gastric cancer tissues (IARC TP53 database, R16, November 2012, <http://p53.iarc.fr/>), including codons 245 and 273 (Figure 2B, 2C).

Mutation signature induced by AID activity that accumulated in human TP53 genes in vivo.

The mutation spectrum determined in the TP53 sequences supported the possibility that AID-mediated deamination activity is deeply involved in the induction of driver mutations in tumor-related genes during human gastric carcinogenesis. To gain further insight into the role of AID in the emergence of somatic mutations in vivo, we determined the nucleotide sequences of the human TP53 genes in the gastric mucosa of the human TP53 knock-in (Hupki) mice³³ after crossing them with AID transgenic (Tg) mice using deep sequencing analysis. Although the mean coverage reached 6434-fold, only one low-abundance mutation was detected in the human TP53 sequences from the gastric mucosa of 6 control Hupki mice with any type of codon 72 polymorphic variant (codon72 Arg/Arg, Pro/Pro and Arg/Pro) (Table 4). In contrast, low-abundance mutations accumulated in the human TP53 gene sequences in the gastric mucosa of 6 of 7 (85.7%) Hupki/AID-Tg mice ($p=0.053$ by χ^2 test, Table 4). 8 of 9 (88.9%) genetic alterations determined in the gastric mucosa of 6 Hupki/AID-Tg mice were C:G>T:A transitions at the GpCpX motif ($p<0.01$ by χ^2 test). Moreover, numbers of the low-abundance TP53 mutations were identical to the amino acid positions altered in human gastric cancer tissues^{34,35} (Figure 2D). These findings provide evidence that the AID-mediated cytidine deamination could be deeply involved in the nucleotide alterations of the human TP53 gene that accumulate in the gastric mucosa during gastric tumorigenesis.

Discussion

Chronic inflammation plays a critical role as a background of cancer development^{16, 36, 37}. Therefore, it is reasonable to assume that stepwise accumulation of genetic alterations that contribute to tumorigenesis occurs in chronically inflamed epithelial cells. Consistently, several studies have reported that genetic aberrations are frequently detected in non-tumorous inflamed epithelial tissues where the risk of cancer development is remarkably high³⁸⁻⁴⁰. For example, TP53 mutations are detected in the inflamed mucosa of the colonic epithelium of patients with inflammatory bowel disease^{41, 42}. A recent sequencing study on Barrett's esophagus revealed that the mutational profiles of Barrett's esophagus and esophageal adenocarcinoma are remarkably similar, suggesting that genetic alterations in the metaplasia–carcinoma sequence of Barrett's esophagus occur much earlier than the histologic changes of frank dysplasia⁴⁰. In the present study, using a whole exome sequencing technique, we demonstrated that somatic mutations accumulated not only in tumors but also in the non-tumorous inflamed gastric mucosa infected with H pylori.

Regarding the cancer tissues, whole exome sequencing determined a remarkable difference in the mutation signature between MSI and MSS tumors. Namely, MSS tumors possessed a relatively lower number of total mutations, including SNVs and indels, compared with the MSI tumor. Indeed, the 4 MSS gastric cancer samples had an average of 56 non-silent somatic mutations with few indels, while the MSI gastric cancer sample had 427 non-silent somatic mutations including a large number of indels, consistent with the previous reports^{3, 6}. Moreover, the frequency of mutations at each nucleotide position in the MSS cancers tended to be lower than that in the MSI tumor. Notably, the mutation signature in non-tumorous background gastric mucosa did not differ significantly between patients with MSS and MSI tumors. These findings suggest that the outstanding features of genetic instability characterized as MSS and MSI

phenotypes were not evident in the premalignant condition of gastric mucosa underlying H pylori infection and resultant chronic inflammation, and would only become evident in the tumor cells after malignant transformation. Since the allelic fraction is highly dependent on the cell purity in the tissue samples, we should pay attention to the possibility that the tumor or epithelial cell purity could cause differences in the mutation frequencies in each tissue specimen analyzed.

Whole exome sequencing also clarified that the somatic mutations in various genes were latently accumulated in the non-tumorous gastritis mucosa of patients with H pylori infection. It is notable that the total number of mutated genes of the non-tumorous gastric mucosa was smaller than that of the matched tumor tissues and the majority of the mutations detected in gastritis mucosa were predicted to have no influence on cell behavior, suggesting that most of the mutations in gastritis mucosa were passenger mutations. To unveil whether mutations of driver genes were present in gastritis mucosa, we conducted deep sequencing of the selected target genes on H pylori (+) gastric epithelium. The sensitivity of ultra-deep sequencing analysis is primarily limited by errors introduced during any process of the sample preparation and the sequencing reaction, including PCR amplification, cluster amplification, cycle sequencing, image analysis, and others; thus, it is a challenge to distinguish rare mutations from sequencing artifacts¹⁸. In the present study, we optimized the deep sequencing by defining a strict filter setting and confirmed that no nucleotide alterations over the cut-off were detected in the control lymphocytes or in normal gastric mucosa. Furthermore, we validated the results by repeating the deep sequencing using independent amplicons derived from the same patients, although errors cannot completely be ruled out. Through these strict validation processes, we reproducibly detected low-abundance somatic mutations in non-tumorous gastric mucosa with H pylori infection. Interestingly, TP53 was the most frequently mutated gene in gastritis mucosa among the six tumor-related genes analyzed. Strikingly, numbers of the mutated TP53 nucleotide

positions observed in gastritis mucosa were located at the nucleotide positions of the mutations that accumulated in human gastric cancer tissues. It is not known whether the low-abundance TP53 mutations latently present in gastric mucosa with H pylori infection derive from the clinically-undetectable small tumor cells or untransformed epithelial cells under premalignant conditions. It is reasonable to assume, however, that the low-abundance TP53 mutations in the background gastric mucosa with H pylori infection could be the source of gastric cancer cells.

Recent studies showed that the mutation signature that accumulates in tumor tissues provides the clue to identifying the cause of genetic alterations during tumor development^{22-27, 43}. It was shown that C:G>T:A transitions at XpCpG trinucleotides were especially prominent in many cancer types^{27, 44}. Among the C:G>T:A transitions, the footprint of AID is characterized by C:G>T:A alterations that occur in GpCpX or ApCpX sequences^{22, 28, 29, 45-47}. In the current study, we found that the frequency of C:G>T:A transitions was the highest pattern of SNVs accumulated in the whole exome sequences of all five cancer tissues examined. Interestingly, the C:G>T:A mutations in gastric cancer genomes predominantly accumulated in the context of GpCpX or ApCpX, suggesting the involvement of AID-mediated cytidine deamination in the induction of somatic mutations during gastric carcinogenesis.

We previously revealed a causal relationship between mutagenic AID activity and the accumulation of TP53 mutations during the development of human gastric cancers¹⁴. The present findings showing the detection of low-abundance TP53 mutations with the typical footprint of AID activity in inflamed human gastric mucosa further support the putative role of AID in the generation of genetic alterations in an inflammatory microenvironment of gastric mucosa with H pylori infection. The differences in p53 gene sequences between humans and mice, however, made it difficult to confirm whether constitutive AID expression in gastric mucosa in vivo contributes to the emergence of TP53 mutations that are identical to those observed in human

gastritis and gastric cancer tissues. It is well established that Hupki mice, in which exons 4-9 of human TP53 were knocked in the corresponding mouse exons, are useful models for studying human TP53 mutagenesis^{33, 48, 49}. Therefore, in this study we performed deep sequencing analyses on human TP53 gene of gastric mucosa derived from Hupki mice with, and without, constitutive AID expression. AID expression in gastric mucosa resulted in the accumulation of low-abundance human TP53 mutations with the typical signature of AID-mediated cytidine deamination. Strikingly, numbers of the amino acid changes determined in AID-expressing gastric mucosa of Hupki mice accumulated at positions identical to those altered in human gastric cancers. Taken together, these findings strongly suggest that constitutive AID expression in gastric epithelium plays a critical role in the accumulation of human TP53 mutations during the development of gastric cancers (Figure 3). To clarify the role of AID in the accumulation of somatic mutations and tumorigenesis in the stomach in vivo, further analyses of aberrant AID expression and resultant Trp53 mutations in gastric inflamed mucosa are required using previously established mouse models with inflammation-associated gastric carcinogenesis⁵⁰⁻⁵².

In summary, we demonstrated that various mutations were latently accumulated in the gastric mucosa in patients with H pylori infection. The combination of whole exome sequencing with deep sequencing enabled us to unveil the landscape of mutations that latently accumulated in non-tumorous gastritis mucosa. Interestingly, the mutation signature not only in gastric cancer genomes but also in gastritis mucosa suggested that AID-mediated deamination events were predominant during the process of gastric carcinogenesis. In particular, a causal relationship between AID activity and the accumulation of TP53 mutations was also validated by in vivo models. It is expected that uncovering the landscape of genetic alterations in gastritis mucosa will contribute to clarifying the mechanism of gastric cancer development.

Acknowledgments

We thank A. Takaori-Kondo, K. Shimizu and C. Kakimoto for deep sequencing analyses, M. Hollstein for Hupki mice, T. Honjo for AID Tg mice, and K. Kinoshita for his helpful suggestion.

Figure legends

Figure 1. Somatic mutations in tumor and non-tumorous gastric tissues determined by whole exome sequencing.

(A) Number of non-synonymous and synonymous mutations in each sample. C001-C004 had gastric cancers with MSS features, and C005 had gastric cancer with MSI features.

(B) Distribution of the mutation frequency determined in each sample (at a frequency of $\geq 10\%$ of reads). The proportions of mutations whose frequencies were less than 20% of the total reads significantly differed between cancer tissues and non-cancerous tissues in C001, C003, C004, and C005 ($p < 0.01$ by χ^2 test).

(C) Mutation signature determined in one MSI and four MSS gastric cancer tissues. C:G>T:A transitions in MSS cancer samples were significantly enriched ($p < 0.01$ by ANOVA).

(D) Sequence context of the C:G>T:A transitions in one MSI and four MSS gastric cancer tissues. GpCpX patterns in MSS cancer samples were significant enriched ($p < 0.01$ by ANOVA).

*MSI status was determined from gastric cancer tissue.

**T: tumor, NT: non-tumorous gastritis mucosa

Figure 2. Low-abundance mutations accumulated in TP53 genes in human gastritis mucosa and human TP53-knock-in mice stomach.

(A) Mutation signature of 19 low-abundance TP53 mutations in *H. pylori*-related gastritis mucosa of 34 patients with or without gastric cancer determined by deep sequencing. These low-abundance TP53 mutations included 15 C:G>T:A transitions, 3 T:A>C:G transitions and 1 C:G>A:T transversion. C:G>T:A transitions were significantly enriched ($p < 0.01$ by ANOVA).

(B) Distribution of low-abundance TP53 mutations in human gastritis mucosa of 34 patients with or without gastric cancer identified by deep sequencing.

(C) Distribution of TP53 mutations in human gastric cancers obtained from IARC TP53 database (<http://p53.iarc.fr/>).

(D) Distribution of low-abundance TP53 mutations in gastric mucosa of seven 12-month-old Hupki/AID-Tg mice.

Figure 3. Schematic summary for AID-mediated TP53 mutations and gastric cancer development.

H pylori infection and inflammatory signals such as TNF- α induce aberrant expression of a DNA-mutator enzyme, AID, in gastric epithelial cells. AID-mediated mutagenesis of driver genes including TP53 could contribute to gastric carcinogenesis.

Reference

1. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nature Reviews Genetics* 2010;11:685-696.
2. Bozic I, Antal T, Ohtsuki H, et al. Accumulation of driver and passenger mutations during tumor progression. *Proc Natl Acad Sci U S A* 2010;107:18545-18550.
3. **Wang K, Kan J**, Yuen ST, et al. Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nat Genet* 2011;43:1219-1223.
4. **Zang ZJ, Cutcutache I**, Poon SL, et al. Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. *Nat Genet* 2012;44:570-574.
5. **Nagarajan N, Bertrand D, Hillmer AM**, et al. Whole-genome reconstruction and mutational signatures in gastric cancer. *Genome Biol* 2012;13:R115.
6. **Yoon K, Lee S, Han TS**, et al. Comprehensive genome- and transcriptome-wide analyses of mutations associated with microsatellite instability in Korean gastric cancers. *Genome Res* 2013;23:1109-1117.
7. Uemura N, Okamoto S, Yamamoto S, et al. Helicobacter pylori infection and the development of gastric cancer. *N Engl J Med* 2001;345:784-789.
8. Chiba T, Marusawa H, Seno H, et al. Mechanism for gastric cancer development by Helicobacter pylori infection. *J Gastroenterol Hepatol* 2008;23:1175-1181.
9. Polk DB, Peek RM, Jr. Helicobacter pylori: gastric cancer and beyond. *Nat Rev Cancer* 2010;10:403-414.
10. Aoi T, Marusawa H, Sato T, et al. Risk of subsequent development of gastric cancer in patients with previous gastric epithelial neoplasia. *Gut* 2006;55:588-589.
11. Gutierrez-Gonzalez L, Graham TA, Rodriguez-Justo M, et al. The clonal origins of

- dysplasia from intestinal metaplasia in the human stomach. *Gastroenterology* 2011;140:1251-1260 e1-6.
12. Touati E, Michel V, Thiberge JM, et al. Chronic *Helicobacter pylori* infections induce gastric mutations in mice. *Gastroenterology* 2003;124:1408-1419.
 13. Jenks PJ, Jeremy AH, Robinson PA, et al. Long-term infection with *Helicobacter felis* and inactivation of the tumour suppressor gene p53 cumulatively enhance the gastric mutation frequency in Big Blue transgenic mice. *J Pathol* 2003;201:596-602.
 14. Matsumoto Y, Marusawa H, Kinoshita K, et al. *Helicobacter pylori* infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. *Nat Med* 2007;13:470-476.
 15. Matsumoto Y, Marusawa H, Kinoshita K, et al. Up-regulation of activation-induced cytidine deaminase causes genetic aberrations at the CDKN2b-CDKN2a in gastric cancer. *Gastroenterology* 2010;139:1984-1994.
 16. Chiba T, Marusawa H, Ushijima T. Inflammation-associated cancer development in digestive organs: mechanisms and roles for genetic and epigenetic modulation. *Gastroenterology* 2012;143:550-563.
 17. Nagata N, Akiyama J, Marusawa H, et al. Enhanced expression of activation-induced cytidine deaminase in human gastric mucosa infected by *Helicobacter pylori* and its decrease following eradication. *J Gastroenterol* 2014; e-pub ahead of print 18 April 2013; doi: 10.1007/s00535-013-0808-z.
 18. Schmitt MW, Kennedy SR, Salk JJ, et al. Detection of ultra-rare mutations by next-generation sequencing. *Proc Natl Acad Sci U S A* 2012;109:14508-14513.
 19. Nasu A, Marusawa H, Ueda Y, et al. Genetic heterogeneity of hepatitis C virus in association with antiviral therapy determined by ultra-deep sequencing. *PLoS One*

- 2011;6:e24907.
20. Nishijima N, Marusawa H, Ueda Y, et al. Dynamics of hepatitis B virus quasispecies in association with nucleos(t)ide analogue treatment determined by ultra-deep sequencing. *PLoS One* 2012;7:e35052.
 21. **Sato Y, Yoshizato T, Shiraishi Y**, et al. Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat Genet* 2013;45:860-867.
 22. Schmitz KM, Petersen-Mahrt SK. AIDing the immune system-DIAbolic in cancer. *Semin Immunol* 2012;24:241-245.
 23. Nik-Zainal S, Alexandrov LB, Wedge DC, et al. Mutational processes molding the genomes of 21 breast cancers. *Cell* 2012;149:979-993.
 24. **Stephens PJ, Tarpey PS**, Davies H, et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature* 2012;486:400-404.
 25. Roberts SA, Lawrence MS, Klimczak LJ, et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet* 2013;45:970-976.
 26. Burns MB, Temiz NA, Harris RS. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat Genet* 2013;45:977-983.
 27. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature* 2013;500:415-421.
 28. Bransteitter R, Pham P, Calabrese P, et al. Biochemical analysis of hypermutational targeting by wild type and mutant activation-induced cytidine deaminase. *J Biol Chem* 2004;279:51612-51621.
 29. Pham P, Bransteitter R, Petruska J, et al. Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature* 2003;424:103-107.
 30. **Beale RC, Petersen-Mahrt SK**, Watt IN, et al. Comparison of the differential

context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. *J Mol Biol* 2004;337:585-596.

31. **Burns MB, Lackey L**, Carpenter MA, et al. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* 2013;494:366-370.
32. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009;4:1073-1081.
33. Besaratinia A, Pfeifer GP. Applications of the human p53 knock-in (Hupki) mouse model for human carcinogen testing. *FASEB J* 2010;24:2612-2619.
34. Petitjean A, Mathe E, Kato S, et al. Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Hum Mutat* 2007;28:622-629.
35. Olivier M, Hollstein M, Hainaut P. TP53 mutations in human cancers: origins, consequences, and clinical use. *Cold Spring Harb Perspect Biol* 2010;2:a001008.
36. Shimizu T, Marusawa H, Endo Y, et al. Inflammation-mediated genomic instability: roles of activation-induced cytidine deaminase in carcinogenesis. *Cancer Sci* 2012;103:1201-1206.
37. Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. *Cell* 2010;140:883-899.
38. Kou T, Marusawa H, Kinoshita K, et al. Expression of activation-induced cytidine deaminase in human hepatocytes during hepatocarcinogenesis. *Int J Cancer* 2007;120:469-476.
39. Barrett MT, Sanchez CA, Prevo LJ, et al. Evolution of neoplastic cell lineages in Barrett oesophagus. *Nat Genet* 1999;22:106-109.
40. Streppel MM, Lata S, Delabastide M, et al. Next-generation sequencing of endoscopic

- biopsies identifies ARID1A as a tumor-suppressor gene in Barrett's esophagus. *Oncogene* 2014;33:347-357.
41. Yin J, Harpaz N, Tong Y, et al. p53 point mutations in dysplastic and cancerous ulcerative colitis lesions. *Gastroenterology* 1993;104:1633-1639.
 42. Kern SE, Redston M, Seymour AB, et al. Molecular genetic profiles of colitis-associated neoplasms. *Gastroenterology* 1994;107:420-428.
 43. Pleasance ED, Stephens PJ, O'Meara S, et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* 2010;463:184-190.
 44. **Lawrence MS, Stojanov P, Polak P**, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 2013;499:214-218.
 45. Muramatsu M, Kinoshita K, Fagarasan S, et al. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* 2000;102:553-563.
 46. Cascalho M. Advantages and disadvantages of cytidine deamination. *J Immunol* 2004;172:6513-6518.
 47. Liu M, Schatz DG. Balancing AID and DNA repair during somatic hypermutation. *Trends Immunol* 2009;30:173-181.
 48. **Luo JL, Yang Q**, Tong WM, et al. Knock-in mice with a chimeric human/murine p53 gene develop normally and show wild-type p53 responses to DNA damaging agents: a new biomedical research tool. *Oncogene* 2001;20:320-328.
 49. Kucab JE, Phillips DH, Arlt VM. Linking environmental carcinogen exposure to TP53 mutations in human tumours using the human TP53 knock-in (Hupki) mouse model. *FEBS J* 2010;277:2567-2583.
 50. Tye H, Kennedy CL, Najdovska M, et al. STAT3-driven upregulation of TLR2 promotes

- gastric tumorigenesis independent of tumor inflammation. *Cancer Cell* 2012;22:466-478.
51. Oshima H, Oshima M. Mouse models of gastric tumors: Wnt activation and PGE2 induction. *Pathol Int* 2010;60:599-607.
 52. **Okumura T, Ericksen RE**, Takaishi S, et al. K-ras mutation targeted to gastric tissue progenitor cells results in chronic inflammation, an altered microenvironment, and progression to intraepithelial neoplasia. *Cancer Res* 2010;70:8435-8445.

Author names in bold designate shared co-first authorship.

Figure 1. Shimizu et al.

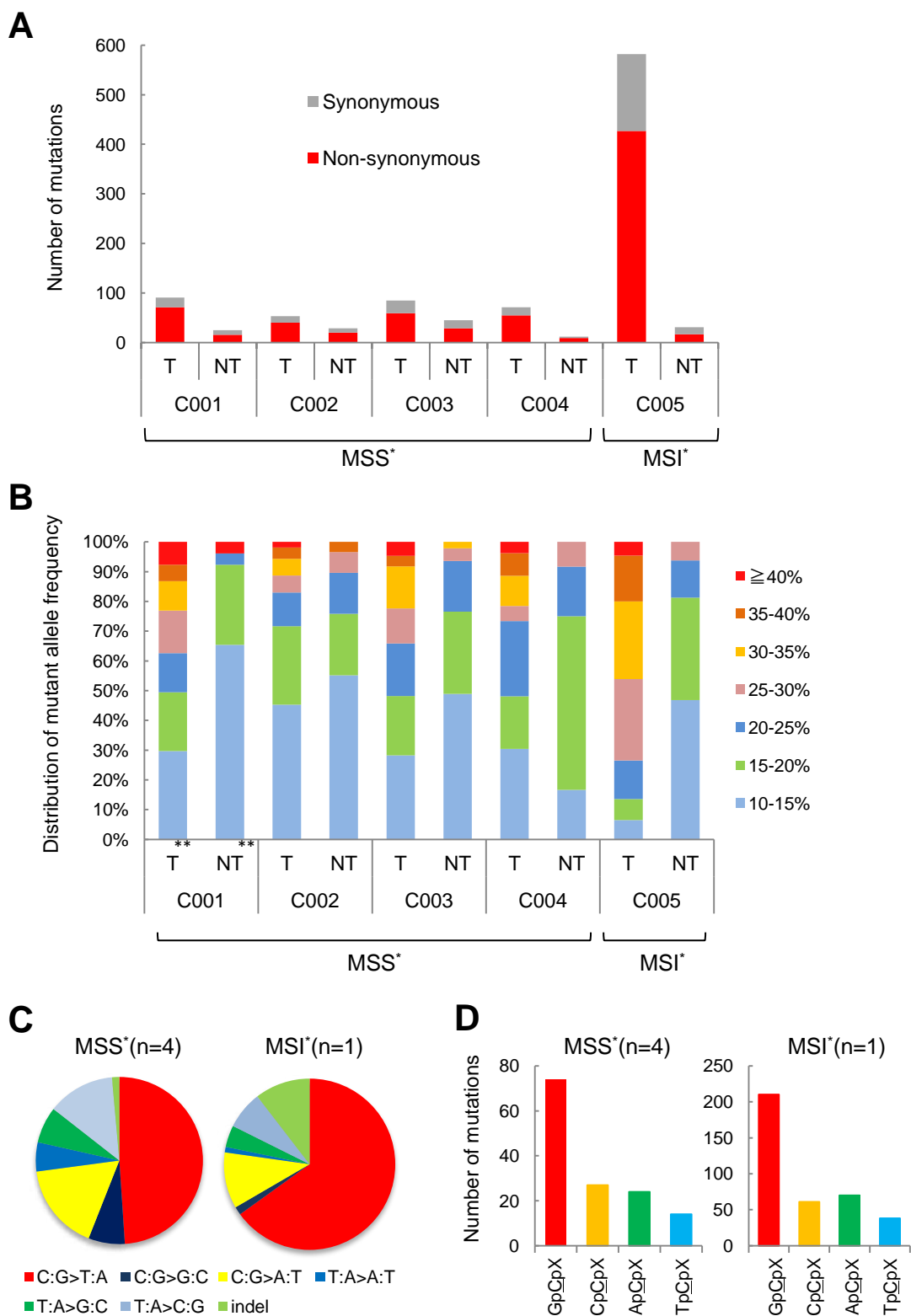
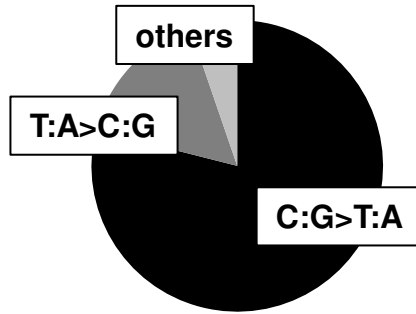
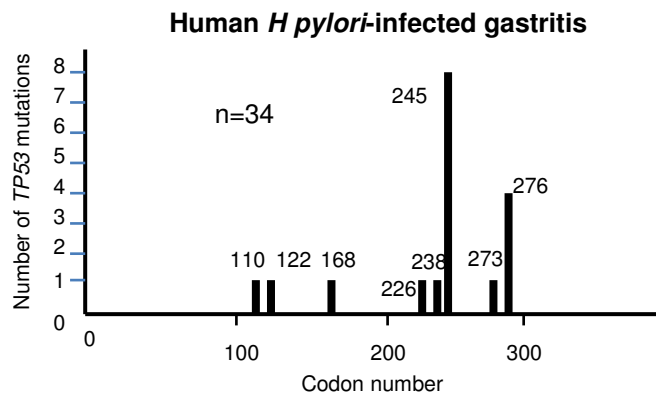


Figure 2. Shimizu et al.

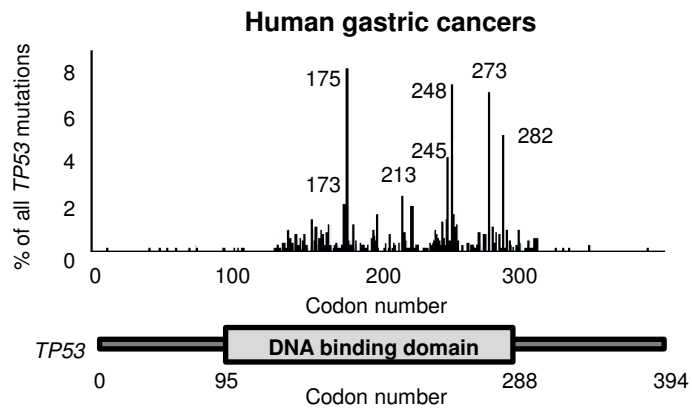
A



B



C



D

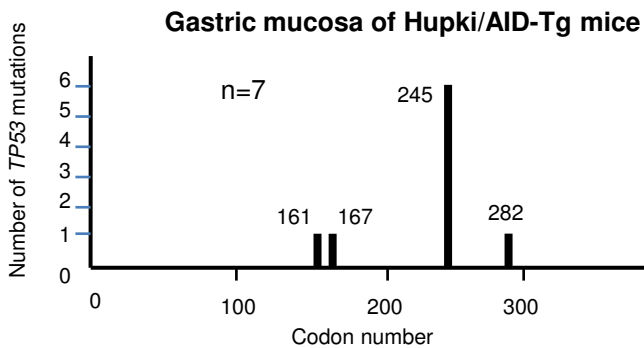


Figure 3. Shimizu et al.

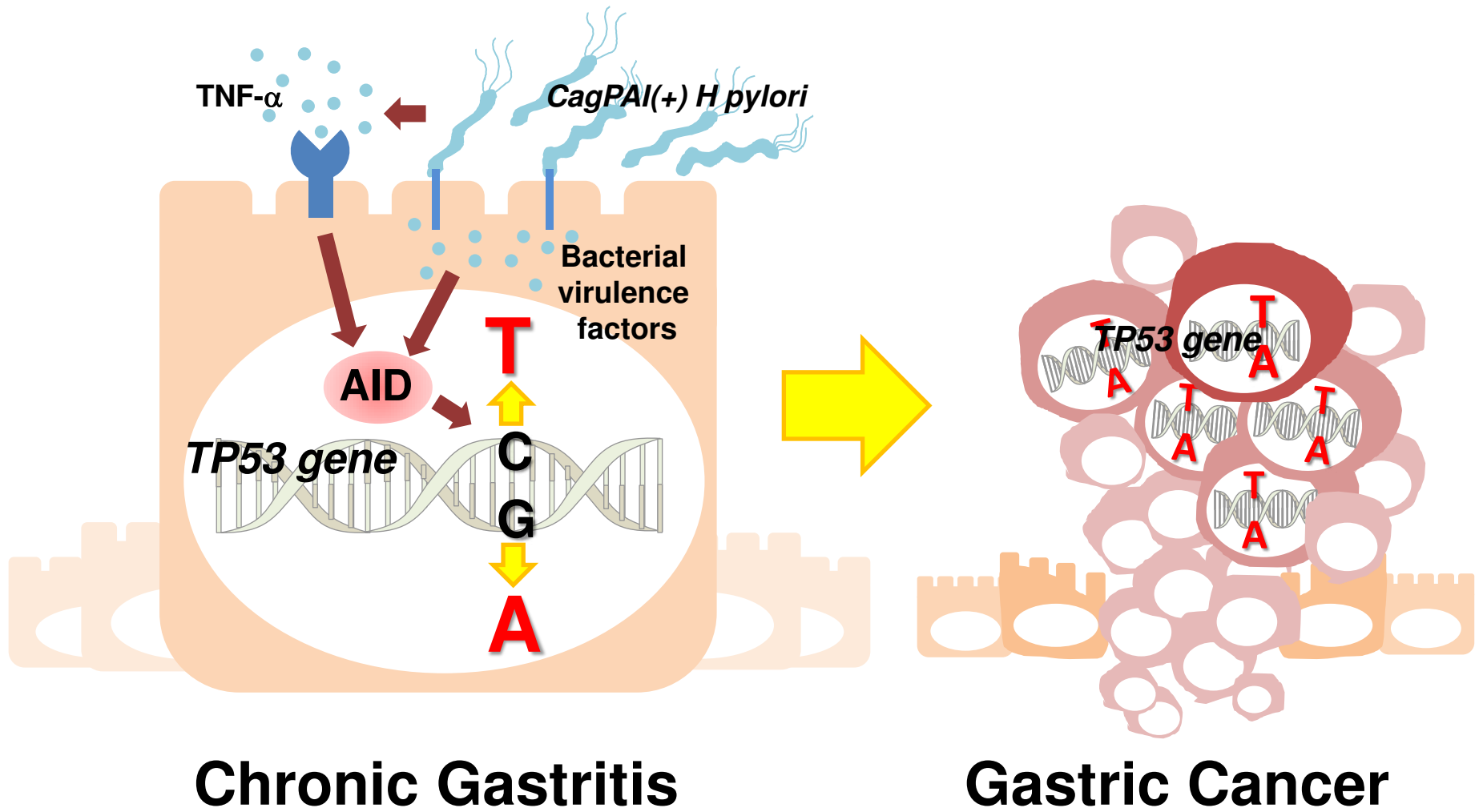


Table 1. List of 26 genes recurrently mutated in two or more tumor tissues of 5 gastric cancer patients.

Gene symbol	Gene name	Gene size (bp)	# of mutated samples	# of mutations
OBSCN	obscurin, cytoskeletal calmodulin and titin-interacting RhoGEF	23,907	3	4
TP53	tumor protein p53	1,182	3	3
MUC17	mucin 17, cell surface associated	13,482	2	2
HRNR	hornerin	8,553	2	3
FLG	filaggrin	12,186	2	2
PCDHB5	protocadherin beta 5	2,388	2	3
PRB1	proline-rich protein BstNI subfamily 1	996	2	2
RORB	RAR-related orphan receptor B	1,380	2	2
SLC35G6	solute carrier family 35, member G6	1,017	2	2
ARID1A	AT rich interactive domain 1A (SWI-like)	6,858	2	2
CACNA1B	calcium channel, voltage-dependent, N type, alpha 1B subunit	7,020	2	2
CD163	CD163 molecule	3,471	2	2
CEP164	centrosomal protein 164kDa	4,383	2	2
CNGB1	cyclic nucleotide gated channel beta 1	900	2	2
COL5A1	collagen, type V, alpha 1	5,517	2	2
CYP2A6	cytochrome P450, family 2, subfamily A, polypeptide 6	1,485	2	2
FCGBP	Fc fragment of IgG binding protein	16,218	2	2
GPC6	glypican 6	1,668	2	2
HCN2	hyperpolarization activated cyclic nucleotide-gated potassium channel 2	2,670	2	2
PCDHA3	protocadherin alpha 3	2,853	2	2
RABGGTA	Rab geranylgeranyltransferase, alpha subunit	1,704	2	2
RAD54L2	RAD54-like 2 (<i>S. cerevisiae</i>)	4,404	2	2
ROPN1B	rhophilin associated tail protein 1B	639	2	2
SPDYE5	speedy homolog E5 (<i>Xenopus laevis</i>)	1,014	2	2
SUPT6H	suppressor of Ty 6 homolog (<i>S. cerevisiae</i>)	5,181	2	2
TUBGCP2	tubulin, gamma complex associated protein 2	2,793	2	2

Table 2. Low-abundance mutations in H pylori-related gastritis mucosa of patients with gastric cancer detected by deep sequencing.

Case*	TP53 DNA binding domain	ARID1A exon20	MLL2 SET domain	SUV39H1 SET domain	CTNNB1 exon3	PIK3CA exon21
C001	None	None	None	None	None	None
C002	c.365T>C p.V122A c.734G>A p.G245D	None	None	None	None	None
C003	None	c.5763C>T p.T1921T	None	None	None	None
C004	c.734G>A G245D	None	None	None	None	None
C005	None	None	None	None	None	None
C006	None	c.5920C>T p.Q1974X c.6001A>G p.M2001V	None	None	None	None
C007	c.712T>C p.C238R	None	None	None	None	None
C008	None	None	c.16373A>G p.E5458G c.16438A>G p.N5480D	None	None	None
C009	c.734G>A p.G245D	None	None	None	None	None
C010	None	None	None	None	None	None
C011	c.734G>A p.G245D	None	None	None	None	None
C012	c.329G>T p.R110L	c.5983C>T p.P1995S	None	None	None	None
C013	c.503A>G p.H168R c.827C>T p.A276V	None	None	None	None	None
C014	None	None	None	None	None	None
C015	None	c.5307A>G p.L1769L c.5607G>A p.R1869R	None	None	None	None
C016	None	None	c.16428C>T p.S5476S	None	c.99T>C p.S33S	None
C017	c.734G>A p.G245D	c.5406G>A p.E1802E	None	None	None	None
C018	None	None	None	None	None	None
C019	None	None	None	None	None	None
C020	None	None	c.16379T>A p.V5460E	None	None	None
C021	None	None	None	None	None	None
C022	None	None	None	None	None	None
C023	c.817C>T p.R273C	None	None	None	None	None
C024	c.827C>T p.A276V	c.5239C>T p.P1747S c.5503C>T p.Q1835X	None	None	None	None
C025	None	c.5482T>C p.S1828P	None	None	None	None
C026	c.734G>A p.G245D	None	None	None	None	None
C027	None	None	None	None	None	None
C028	None	None	None	None	None	None

All low-abundance mutations in this table were validated by repeated deep sequencing.

Low-abundance mutations written in bold letters are non-synonymous, and those written in small letters are synonymous

*C001-C028 are H pylori-related gastritis patients with gastric cancer.

Table 3. Low-abundance mutations in H pylori-related gastritis mucosa of patients without gastric cancer and normal gastric mucosa detected by deep sequencing.

Case*	TP53 DNA binding domain	ARID1A exon20	MLL2 SET domain	SUV39H1 SET domain	CTNNB1 exon3	PIK3CA exon21
A029	c.734G>A p.G245D c.827C>T p.A276V	None	None	c.1029G>A p.R343R	None	None
A030	c.412G>A p.A138T c.677G>A p.G226D	None	None	None	None	None
A031	c.827C>T p.A276V	None	None	None	None	None
A032	c.734G>A p.G245D	None	None	None	None	None
A033	None	None	None	None	None	None
A034	None	c.5983C>T p.P1995S	None	None	None	None
N035	None	None	None	None	None	None
N036	None	None	None	None	None	None
N037	None	None	None	None	None	None
N038	None	None	None	None	None	None
N039	None	None	None	None	None	None

All low-abundance mutations in this table were validated by repeated deep sequencing.

Low-abundance mutations written in bold letters are non-synonymous, and those written in small letters are synonymous

*A029-A034 are H pylori-related gastritis patients without gastric cancer. N035-N039 are individuals without H pylori infection and gastric cancer.

Table 4. Low-abundance TP53 mutations in gastric mucosa of Hupki/AID-Tg mice detected by deep sequencing.

Mouse number	Codon 72 polymorphism *	AID expression **	Average Coverage	low-abundance TP53 mutation ***		
				Nucleotide/Amino acid change	Coverage	# of Mutant Allele
1	Arg/Arg	+	3082	c.734G>A p.G245D	3056	3
2	Arg/Arg	+	4057	c.734G>A p.G245D	4662	8
3	Pro/Pro	+	4374	c.734G>A p.G245D	9160	13
				c.844C>T p.R282W	4223	5
4	Pro/Pro	+	6015	c.734G>A p.G245D	11098	12
5	Pro/Pro	+	7570	c.734G>A p.G245D	7694	10
6	Arg/Pro	+	4676	c.481G>A p.A161T	4551	6
				c.501G>A p.Q167Q	4572	5
				c.734G>A p.G245D	8061	11
7	Arg/Pro	+	8130		None	
8	Arg/Arg	-	8853	c.734G>A p.G245D	6170	10
9	Arg/Arg	-	6470		None	
10	Pro/Pro	-	5090		None	
11	Pro/Pro	-	11036		None	
12	Arg/Pro	-	4052		None	
13	Arg/Pro	-	10237		None	

All mice were male and sacrificed at 12 months of age.

*Hupki mouse strain with arginine(Arg)/Arg, proline(Pro)/Pro or Arg/Pro variants at codon 72 of human TP53. A common polymorphism is frequently observed at codon 72 in TP53, resulting in either a Pro or an Arg, and the polymorphism at codon 72 affect critical biochemical properties of p53 protein (Ref. 49)

**Constitutive AID expression of the stomach was achieved by crossing Hupki mice with AID-Tg mice.

***Low-abundance somatic mutations accumulated in human TP53 gene sequences in the gastric mucosa of Hupki mice.