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

Takahiro Shimizu[1], Hiroyuki Marusawa[1], Yuko Matsumoto[1], Tadashi Inuzuka[1], Atsuyuki Ikeda[1], Yosuke Fujii[1], Sachiko Minamiguchi[2], Shin'ichi Miyamoto[1], Tadayuki Kou[3], Yoshiharu Sakai[4], Jean E Crabtree[5] and Tsutomu Chiba[1]

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 *TP5*3 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 alterations[1]. 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 coincidently, or subsequently, acquired a driver mutation[2]. 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 mutations[3-6]. 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, *ARID1A*[3]. 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* infection[9]. Multicentric tumor development characterized in *H pylori*-associated gastric cancer[10] suggests that the *H pylori*-infected stomach could include epithelial cells and/or intestinal metaplasia with genetic damage that possess high susceptibility to tumor development[11]. Consistent with these findings, animal experiments using Big Blue mice revealed a high mutation frequency in the *H pylori*-infected gastric mucosa[12], which are increased in

Helicobacter-infected Big Blue-p53 hemizygous mice[13]. 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 sequences[14]. 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 genes[18]. 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 study[21]. 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 adhesion[4] 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 occurrence[22-27]. 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 report[5], 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 predictions[32] 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) mice[33] 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 high[38-40]. 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 dysplasia[40]. 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 artifacts[18]. 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 cancers[14]. 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 carcinogenesis[50-52].

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 particularly, 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 ≥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.

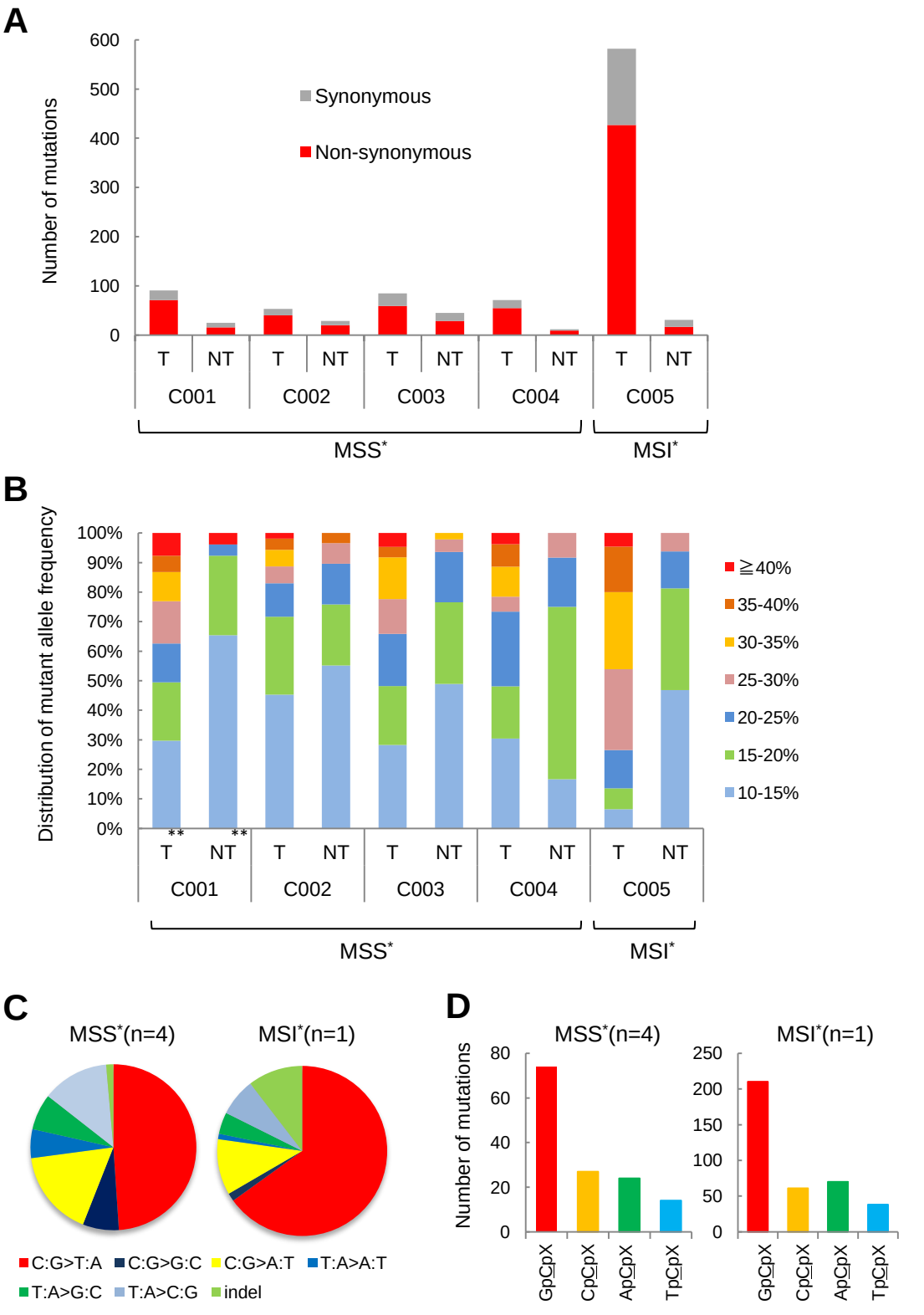


Figure 2. Shimizu et al.

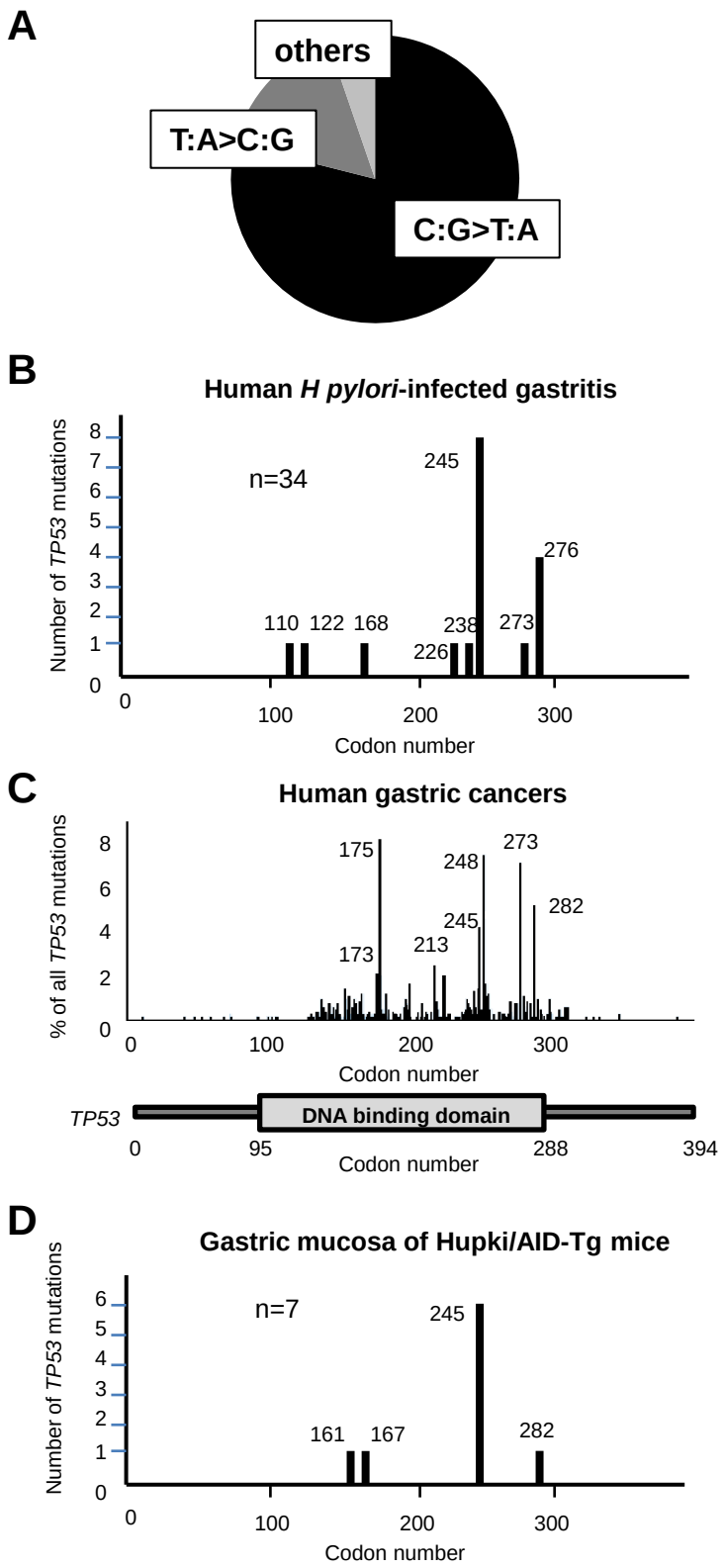


Figure 3. Shimizu et al.

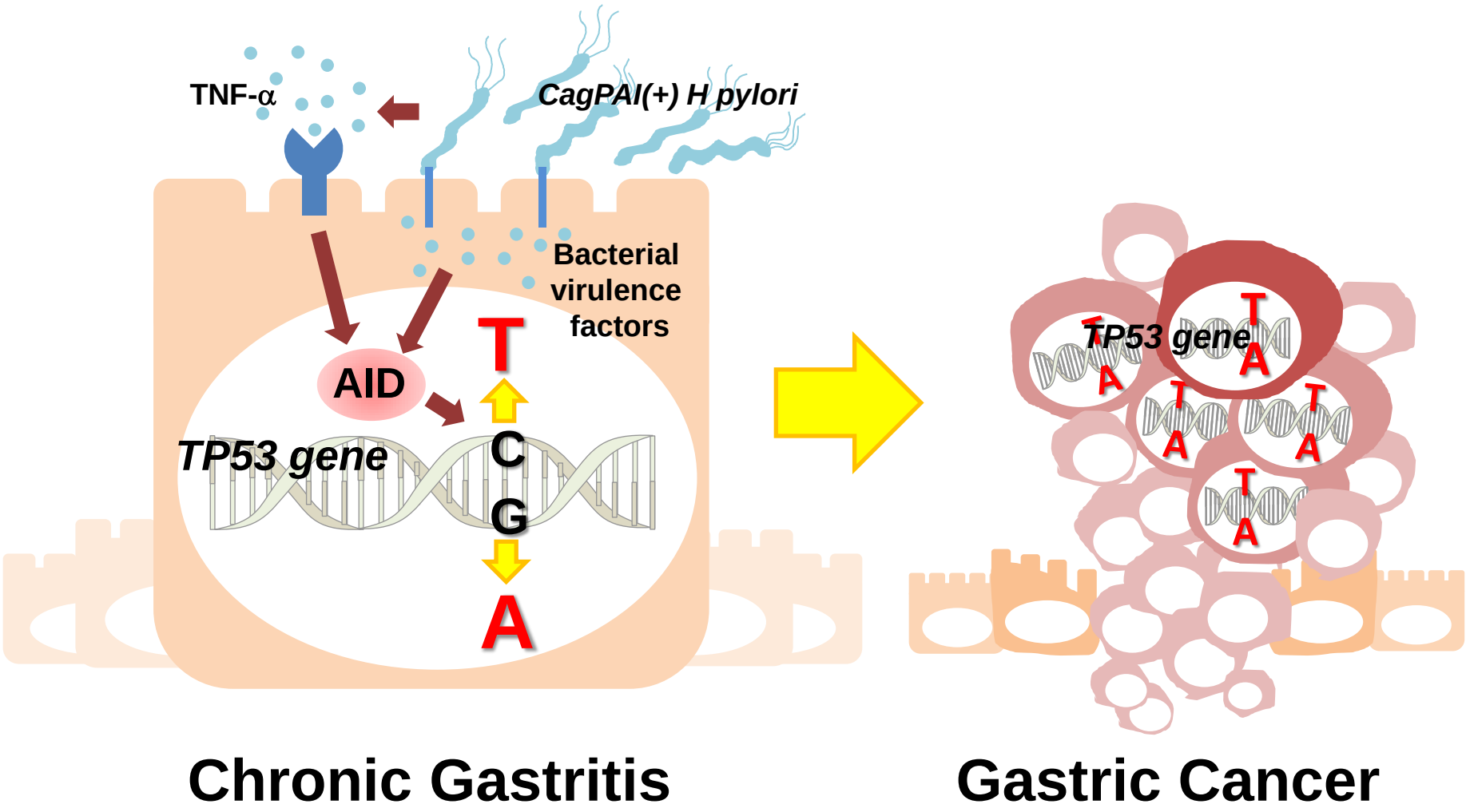


Table 1. List of 26 genes recurrently mutated in two or more tumor tissues of 5 gastric cancer patients.

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 (S. cerevisiae)	4,404	2	2
ROPN1B	rhophilin associated tail protein 1B	639	2	2
SPDYE5	speedy homolog E5 (Xenopus laevis)	1,014	2	2
SUPT6H	suppressor of Ty 6 homolog (S. cerevisiae)	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.

Supplementary Materials and Methods.

Tissue specimens and DNA extraction.

Gastric cancer and non-cancerous gastric tissue specimens were obtained from individuals undergoing surgical or endoscopic resection for primary gastric cancer or upper endoscopy at Kyoto University and Kitano Hospital. All the specimens of non-cancerous gastritis were obtained from the greater curvature of the antrum approximately 5 cm away from the tumors. For surgical samples, two specimens, including mucosa and submucosa, were obtained from the same locations. One was frozen and used for genetic analysis and the other was fixed in formalin and stained with hematoxylin and eosin for evaluation of gastric atrophy and intestinal metaplasia. For biopsy samples, only one specimen was obtained and used for genetic analysis. Gastric mucosal atrophy was evaluated according to the endoscopic-atrophy-border scale described by Kimura and Takemoto, which correlates with the results of histologic evaluation[1, 2], consistent with previous reports[3, 4]. Based on this scale, the grade of atrophy was divided into three types: mild (C1-C2), moderate (C3-O1), and marked (O2-O3). In addition, the histologic evaluation of gastritis was performed according to the updated Sydney system[5]. The degree of mucosal atrophy and intestinal metaplasia in the antrum were classified according to four grades (normal, mild, moderate, and marked). None of the gastritis samples evaluated by histologic examination had evidence of dysplasia. Genomic DNA was extracted from the frozen tissues using DNeasy Tissue kit (Qiagen, Valencia, CA). Tumor samples were assessed for microsatellite instability using an MSI detection kit (Promega, Madison, WI). *H pylori* infection status was defined by the presence of *H pylori* in gastric specimens, rapid urease test (Sankyo, Tokyo, Japan), immunoglobulin G antibody against *H pylori* in the serum, or PCR test for *H pylori* 16S rRNA. Written informed consent for the use of specimens was obtained from all patients in accordance with the Declaration of Helsinki, and the ethics committees of Kyoto University graduate School, Faculty of Medicine and Kitano Hospital approved the study.

Animal Experiments.

The human *TP53* knock-in (Hupki) mice (Trp53$^{tm/Holl}$, homozygous for the knock-in TP53 allele harboring human TP53 sequences in the 129/Sv background) and AID transgenic (Tg) mice were described previously[6-8]. Hupki mice with any type of codon 72 polymorphic variant (codon72, Arg/Arg, Pro/Pro and Arg/Pro) were crossed with AID-Tg mice. We sacrificed 12-month-old male mice and obtained gastric mucosa as described previously[9]. Genomic DNA was extracted from the frozen tissues using DNeasy Tissue kit. All experiments involving mice conformed to the relevant regulatory standards and were reviewed and approved by the Kyoto University School of Medicine Institutional Animal Care and Use Committee.

Whole exome sequencing analysis and variant filtering.

Sequencing data were analyzed with NextGENe 2nd Generation Sequence Analysis Software v2.2.0. We identified somatic mutations by the following variant filtering process (Supplementary Figure 1). First, the obtained reads from the Genome Analyzer IIx were aligned with the reference sequences of the Human Genome Build 37.3. Reads with 96% or more bases matching a particular position of the reference sequences were aligned. Furthermore, reads with a median quality value score of more than 20 and no more than 3 uncalled nucleotides were allowed anywhere in a single read. Aligned sequencing data are shown in Supplementary Table 4. Second, SNVs and short indels with minimum sequence coverage of 11 reads were identified and initially filtered using the NextGENe mutation score, which provided an empirical estimation of the likelihood that a given mutation was real and not an artifact of sequencing or alignment errors, as shown in Supplementary Figure 1. Briefly, the overall mutation scores of SNVs were restricted to more than 7. Also, any 2 or more SNVs located in a 10-bp window were discarded (mismatch score ≥1), and the wrong allele score had to be more than 0.95 to exclude wrong mismatch errors. In addition, the

variant allele balance per mutant position was below 0.2 (1/5< Forward Variant Read/All Forward Read x All Reverse Read/Reverse Variant Read <5). Third, to exclude germline polymorphisms, SNVs present in the lymphocyte sample from the same patient at a frequency of greater than 5% were filtered out. In addition, validated SNPs in dbSNP135 were removed. Fourth, to minimize the number of false positive, the remaining SNVs in gastric cancer and gastric mucosa samples were subsequently filtered with the following criteria. (a) In the nucleotide position whose coverage was ≥11, a predicted mutant allele must be present in at least 20% of the total reads. (b) In the position whose coverage was ≥20, a predicted mutant allele must be present in at least 15% of the total reads. (c) In the position whose coverage was ≥40, a predicted mutant allele must be present in at least 10% of the total reads. Finally, candidate SNVs and indels were validated by standard Sanger sequencing (Applied Biosystems 3500 Geneticanalyzer, Life Technologies) and low frequency SNVs or indels were also tested by TA-cloning (TOPOTA Cloning Kit, Life Technologies) in combination with Sanger sequencing.

Deep sequencing analysis and variant filtering.

In the deep sequencing analysis, we identified low-abundance somatic mutations by the following variant filtering process (Supplementary Figure 2). The obtained reads from the GS Junior System were aligned with the reference sequences using NextGENe v2.2.0. Reads with 90% or more bases matching a particular position of the reference sequences were aligned, and reads with a median quality value score of more than 20 and no more than 3 uncalled nucleotides were allowed anywhere in a single read. The initial filtering process was the same as that of whole exome sequencing analysis except for the following conditions. Indel identification at all nucleotide positions and SNV identification at the positions where over four mononucleotides repeat were ignored because the GS Junior System sequencing data had these kinds of artifacts at low frequency. The predicted mutant alleles must be present in at least 0.1% of the total reads. Finally, candidate mutations were validated by

repeated deep sequencing using independent amplicons derived from the same samples.

Functional prediction analysis and pathway analysis.

Functional prediction analyses for missense mutations detected by sequencing were performed using the Sorting Intolerant From Tolerant (SIFT) algorithm (http://sift.jcvi.org)[10]. Briefly, SIFT prediction is based on the degree of conservation of amino acid residues in sequence alignments derived from closely related sequences, collected through PSI-BLAST. Mutations that were predicted to have a phenotypic effect were "damaging" and those predicted to have no effect were "tolerant". Pathway analyses were performed using the DAVID Bioinformatics Resource 6.7 (http://david.abcc.ncifcrf.gov/)[11]. All genes bearing the functionally altered mutations that were predicted to be "damaging" by SIFT or frame-shift, nonsense, and splice site mutations, detected by whole exome sequencing were used as input. All human genes were used as the population background. Enriched gene sets reflecting Gene Ontology biological processes and Kyoto Encyclopedia of Genes and Genomes were corrected by comparing the classifications of the mutated genes with those of all human genes in the background set. If the P value was less than 0.05, enrichment of a given category was regarded to be statistically significant.

Statistical analysis.

Statistical analysis was performed using the χ^2 test, Fisher's exact test, and ANOVA. Differences were considered to be statistically significant if *P*-values were less than 0.05.

Supplementary References.

1. Satoh K, Kimura K, Taniguchi Y, et al. Distribution of inflammation and atrophy in the stomach of Helicobacter pylori-positive and -negative patients with chronic gastritis. Am J Gastroenterol 1996;91:963-969.
2. Ito S, Azuma T, Murakita H, et al. Profile of Helicobacter pylori cytotoxin derived from two areas of Japan with different prevalence of atrophic gastritis. Gut

1996;39:800-806.

3. 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.

4. Quach DT, Le HM, Nguyen OT, et al. The severity of endoscopic gastric atrophy could help to predict Operative Link on Gastritis Assessment gastritis stage. J Gastroenterol Hepatol 2011;26:281-285.

5. Dixon MF, Genta RM, Yardley JH, et al. Classification and grading of gastritis. The updated Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994. Am J Surg Pathol 1996;20:1161-1181.

6. **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.

7. Okazaki IM, Hiai H, Kakazu N, et al. Constitutive expression of AID leads to tumorigenesis. J Exp Med 2003;197:1173-1181.

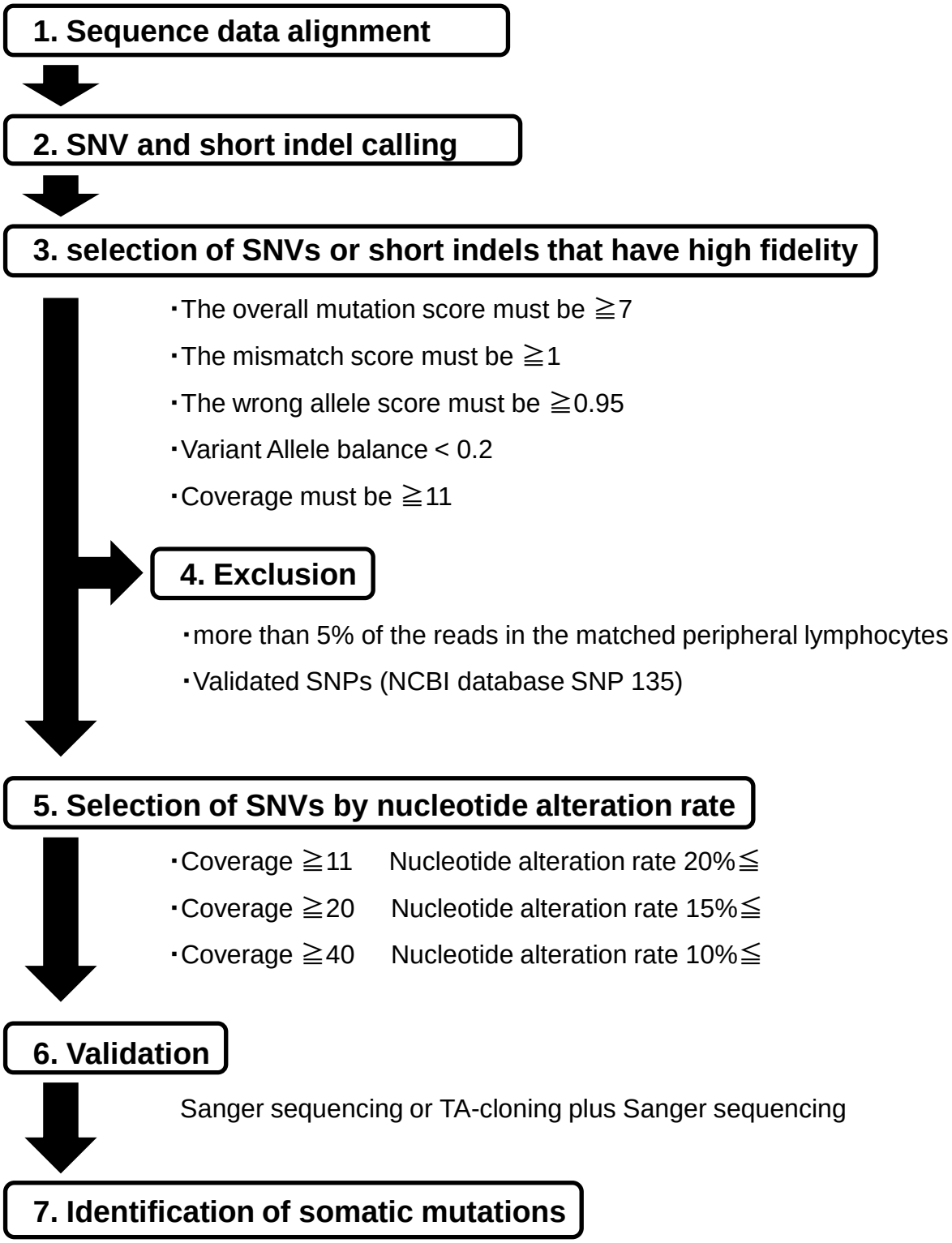
8. Morisawa T, Marusawa H, Ueda Y, et al. Organ-specific profiles of genetic changes in cancers caused by activation-induced cytidine deaminase expression. Int J Cancer 2008;123:2735-2740.

9. 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.

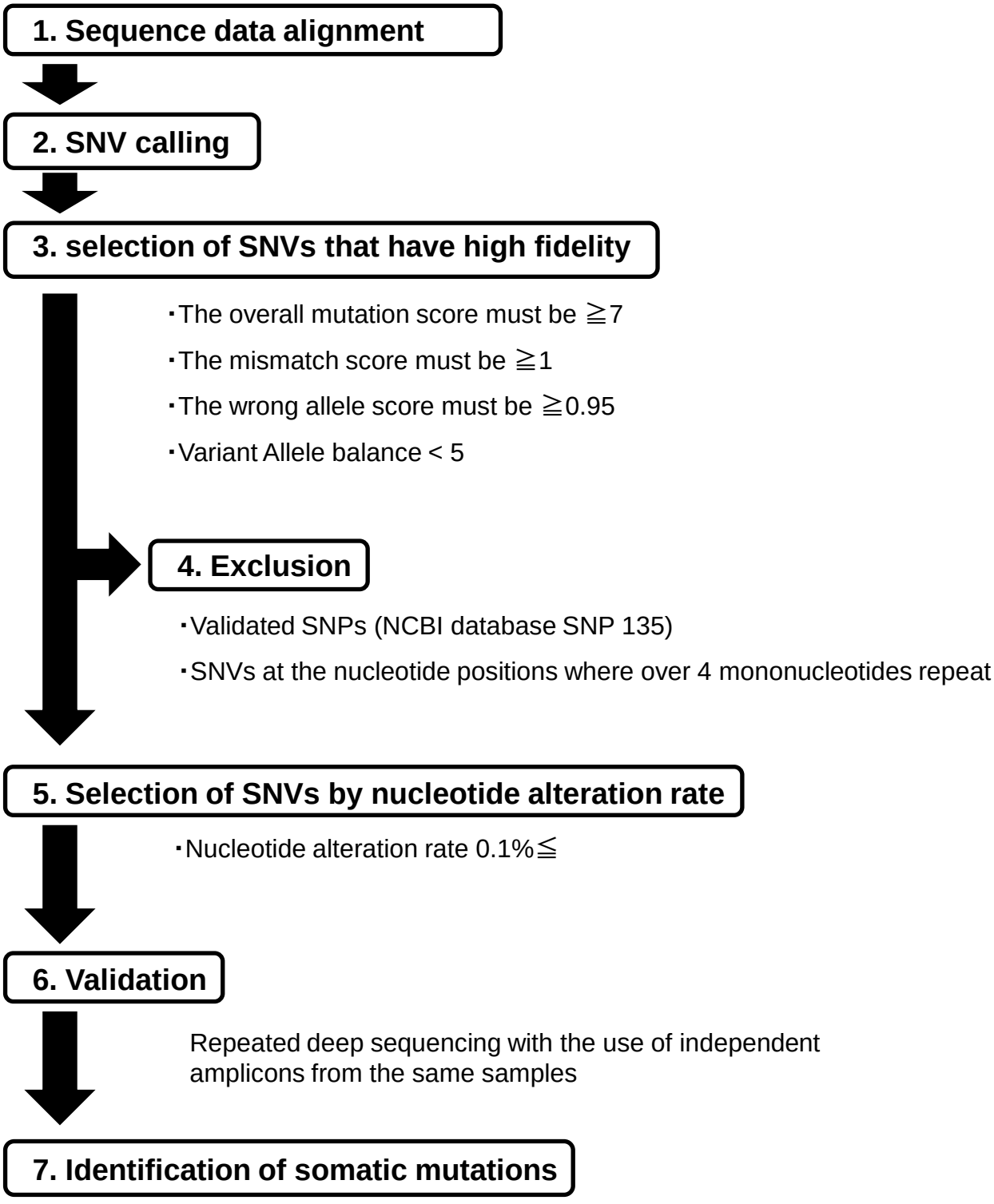
10. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4:1073-1081.

11. **Huang da W**, **Sherman BT**, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44-57.

Author names in bold designate shared co-first authorship.



Supplementary Figure 1. Variant filtering process for whole exome sequencing.
The variant filtering process is described in detail in Supplementary Materials and Methods.



Supplementary Figure 2. Variant filtering process for deep sequencing.

The variant filtering process is described in detail in Supplementary Materials and Methods.

Supplementary Table 1. List of PCR primers used for validated deep sequencing.

Gene	Chr.*	Target position (Chr.* position)	Forward primer sequence	Reverse primer sequence
FAAH	1	46870734	5'-ggaagtgaacaaagggacca-3'	5'-aatgacccaagatgcagagc-3'
NBPF10	1	145301786	5'-tggggagagttttgtccttg-3'	5'-ctgagctcctcagcttgctt-3'
HRNR	1	152187116	5'-tctggccagtctcctagtcg-3'	5'-cctagagccgtgttgttcgt-3'
ASTN1	1	176903446	5'-gcacttccatgtttgggact-3'	5'-caggactcctcaaagccgta-3'
SFXN5	2	73226092	5'-acctgcatccaagttcatcc-3'	5'-ttggaaagcctgtcctgact-3'
IL8RB	2	219000514	5'-attctgggcatccttcacag-3'	5'-tgcacttaggcaggaggtct-3'
MYRIP	3	40286031	5'-actggagacccagctgactg-3'	5'-tggagatgatgggagctgat-3'
RAD54L2	3	51696598	5'-tctcctgacagcccagagat-3'	5'-ccagtcaccaagagggaact-3'
ACSL1	4	185689483	5'-acccctgctgtggtgataaa-3'	5'-gacttcgctgcttggaactt-3'
PGGT1B	5	114598544	5'-cagcacaggagggttctttc-3'	5'-tgactgtgccacgagttacc-3'
PCDHA5	5	140202795	5'-cagccccagtataccgtgtt-3'	5'-agctccacttcctcgtggt-3'
PCDHB5	5	140516419	5'-ccaagccgagtacaacatca-3'	5'-ttggtgcctgagtctctgtct-3'
PCDHB5	5	140516948	5'-gtcgtaccagctgctcaagg-3'	5'-caggtagggctgggagaag-3'
HLA-DRB1	6	32548544	5'-gcacagagcaagatgctgag-3'	5'-gcaggacattagagccgttt-3'
FOXO3	6	108985092	5'-accaattctaacgccagcac-3'	5'-caggtcgtccatgaggtttt-3'
AZGP1	7	99573585	5'-ctgtcctgctgtctctgctg-3'	5'-tgctcctggaagtgtgtgag-3'
MUC12	7	100634718	5'-acaacactgtcccttccag-3'	5'-gtggactggtgtcgattcct-3'
MUC17	7	100681407	5'-catgccaacctcaactcc-3'	5'-agcgtgtggctgacaggtat-3'
TRBV7-8	7	142099512	5'-gatccaatttcgggtcatgt-3'	5'-gctgctggcacagagataca-3'
SLG35G5	8	11189184	5'-gcaaaggttcttccaccgta-3'	5'-agggaaagtgcagagaacga-3'
LOC100289448	8	86727678	5'-ggaatctacgccttggactg-3'	5'-gaaaaagctcagcaggatgg-3'
LOC100288500	8	86757709	5'-ccagctgttcaccgaagagt-3'	5'-ttctccttgagctggtcctg-3'
TRPM3	9	73213484	5'-actggaatgtcacggacctc-3'	5'-ccctgcttaccatttttcca-3'
REXO4	9	136282883	5'-atggggaaggcgaaggtc-3'	5'-cgccttccagttttgagaaa-3'
MAMDC4	9	139753547	5'-gggagctgaaggtactgctg-3'	5'-tggcttcaaacacgatctga-3'
CACNA1B	9	140777319	5'-ctctgggctttgtcttccac-3'	5'-cattcgttcatctccagcag-3'
FAM21A	10	51853101	5'-ccagtttggagacacagcag-3'	5'-cacctcctcgtcctcatcat-3'
SYT13	11	45277241	5'-ggctccagaggtcatcaacta-3'	5'-agacattgacaagccccact-3'
ADRBK1	11	67051835	5'-aacttccccctcaccatctc-3'	5'-gctgtgatccagggtgagtt-3'
A2ML1	12	9027066	5'-ccccacagctcattaagaaca-3'	5'-gtgctacttggcttgggaag-3'
DDN	12	49391616	5'-tgaacagtggtagcgacagc-3'	5'-tgctttttaggggagggaat-3'
ATP4B	13	114304771	5'-cacacaaacgcaacacacag-3'	5'-gaagcgtggaaggaaggaac-3'
ITGA11	15	68643603	5'-gaagggtggaaggaaaggag-3'	5'-cctggaactagtcccttgga-3'
ACAN	15	89398426	5'-agtggagctgttcccctcag-3'	5'-gcccactgaagtcaaggtgt-3'

MEF2A	15	100252754	5’-atcagcatcaagtccgaacc-3’	5’-cccgatcactgccatcatag-3’
COG4	16	70515032	5’-tagaggccagagccctcttt-3’	5’-agtgtgatgagccaggtgtg-3’
CHST5	16	75563446	5’-ctcacctgcgcctgattc-3’	5’-cgtgggtgatgttgtggat-3’
AMAC1L3	17	7385617	5’-ttctgggacctcctgacatc-3’	5’-agtccacaccagtcgtagcc-3’
TP53	17	7578176	5’-gtttctttgctgccgtcttc-3’	5’-ttgcacatctcatggggtta-3’
TBC1D26	17	15641610	5’-cccatgtcctttctcactgg-3’	5’-cctgggctcctgatacagtc-3’
KRTAP1-5	17	39183231	5’-cacgcagctgccagacta-3’	5’-gatacgggtgctcacagctc-3’
VEZF1	17	56060257	5’-agctatgccagtgacccagt-3’	5’-gatgcctccttcatgagacc-3’
VEZF1	17	56060266	5’-agctatgccagtgacccagt-3’	5’-gatgcctccttcatgagacc-3’
HCN2	19	608177	5’-gagagcatgacggacatctg-3’	5’-ctctcagaccccagatccag-3’
STK11	19	1223060	5’-tcctgagtgtgtggcaggta-3’	5’-gtgcaggtcctccaagtacg-3’
CYP4F2	19	16003142	5’-ccgccttccatttcaacat-3’	5’-gaggaaaccaaggctcagtg-3’
FCGBP	19	40376798	5’-ctgtatccccagcgacaagt-3’	5’-tcacgtaggcatggatgttg-3’
SLC32A1	20	37356246	5’-ggctacctggggttgtttct-3’	5’-cactcaccaccacgtacagg-3’
CTSA	20	44520380	5’-ctgctgctgctagtgtcctg-3’	5’-tccctcatccgtcatctctt-3’
CACNA1I	22	40042597	5’-tgcctctctcacccttctgt-3’	5’-ctggaagatctcctcgtagca-3’
PNPLA3	22	44322959	5’-ctcagatcttgtgcggaagg-3’	5’-atccacgacttcgtctttgg-3’
SUV39H1	X	48558770	5’-tctgggacgcatcactgtag-3’	5’-ccctggtcattgtaggcaaa-3’
BEX4	X	102471205	5’-caccaagaattcagcccatt-3’	5’-cccatcgaaaattagggaca-3’
ODZ1	X	123779155	5’-tttgccaaccttcagattcc-3’	5’-ccttcctcgcatatttctcc-3’
ACTRT1	X	127185708	5’-ctatgcctctgcctgtgtca-3’	5’-accacggccttgttgagtat-3’
RBMX2	X	129546608	5’-aatccaggacggcctactct-3’	5’-tgggatttatctcggctcct-3’

*Chr.: Chromosome

Supplementary Table 2. Primer sequences used for deep sequencing analysis.

Primer	Nucleotide sequence
TP53 cDNA-1 forward primer	5'-cccctctgagtcaggaaaca-3'
TP53 cDNA-1 reverse primer	5'-agaatgcaagaagcccagac-3'
TP53 cDNA-2 forward primer	5'-cattctgggacagccaagtc-3'
TP53 cDNA-2 reverse primer	5'-ccagtgtgatgatggtgagg-3'
TP53 exon4 forward primer (human and Hupki)	5'-cctggtcctctgactgctct-3'
TP53 exon4 reverse primer (human and Hupki)	5'-gccaggcattgaagtctcat-3'
TP53 exon5/6 forward primer (human and Hupki)	5'-cacttgtgccctgactttca-3'
TP53 exon5/6 reverse primer (human and Hupki)	5'-cttaacccctcctcccagag-3'
TP53 exon7 forward primer (human and Hupki)	5'-cttgggcctgtgttatctcc-3'
TP53 exon7 reverse primer (human and Hupki)	5'-taagaggtcccaaagccaga-3'
TP53 exon8/9 forward primer (human)	5'-gggagtagatggagcctggt-3'
TP53 exon8/9 reverse primer (human)	5'-ccccaattgcaggtaaaaca-3'
ARID1A exon20-1 forward primer	5'-tagctcccagggttgctaga-3'
ARID1A exon20-1 reverse primer	5'-cctggtgtaccctctgctgt-3'
ARID1A exon20-2 forward primer	5'-caacagcagagggtacacca-3'
ARID1A exon20-2 reverse primer	5'-tccactttgttgcagctcac-3'
CTNNB1 exon3 forward primer	5'-ccctggctatcattctgctt-3'
CTNNB1 exon3 reverse primer	5'-tcaaaactgcattctgactttca-3'
PIK3CA exon21 forward primer	5'-catttgctccaaactgacca-3'
PIK3CA exon21 reverse primer	5'-ggtctttgcctgctgagagt-3'
MLL2 exon52-53 forward primer	5'-cccttcttctgcaacctctg-3'
MLL2 exon52-53 reverse primer	5'-ggaggagctgctttgtcact-3'
SUV39H1 exon4-5 forward primer	5'-ccggctttcctgagacttc-3'
SUV39H1 exon4-5 reverse primer	5'-cccaccttgcatgttgtaatc-3'
TP53 exon8/9-2 forward primer (Hupki)	5'-caagggtggttgggagtaga-3'
TP53 exon8/9-2 reverse primer (Hupki)	5'-ccacttgataagaggtcccaag-3'

Supplementary Table 3. Clinicopathological features of 5 gastric cancer patients for whole exome sequencing.

Case	Age	Sex	Gastric cancer				Gastritis (Endoscopic evaluation)		Gastritis (Histological evaluation)	
			Site	Tumor differentiation	Tumor stage (7th UICC*)	MSI** status	Mucosal Atrophy	IM***	Mucosal Atrophy	IM***
C001	66	M	body	moderate	IIB	MSS	marked	+	mild	mild
C002	79	M	body	well	IV	MSS	marked	+	moderate	marked
C003	69	M	body	poor	IIIC	MSS	marked	+	mild	normal
C004	72	M	pylorus	poor	IV	MSS	marked	+	marked	marked
C005	81	F	antrum	moderate	IB	MSI	marked	+	moderate	marked

*UICC: Union for International Cancer Control
**MSI: microsatellite instability, MSS: microsatellite stability
***IM: intestinal metaplasia

Supplementary Table 4. Summary of whole exome sequencing.

Case	Tissue*	Targeted bases (Mbp)	Total sequence (Gbp)	Bases mapped to reference genome (Gbp)	Bases mapped to targeted regions (Gbp)	Median coverage (target region)	>=1x coverage (%)	>=10x coverage (%)	>=20x coverage (%)	>=30x coverage (%)
C001	T	34.42	3.72	3.16	1.64	47.70	80.83	56.21	49.80	44.11
	NT	34.42	4.41	3.55	1.82	52.99	93.21	68.52	59.65	52.54
	LYM	34.42	4.19	3.56	1.65	47.91	93.53	82.69	73.27	62.06
C002	T	34.42	3.04	2.75	1.04	30.36	82.05	59.86	50.79	41.01
	NT	34.42	2.78	2.57	1.20	34.89	78.35	61.48	54.01	46.07
	LYM	34.42	2.53	2.33	1.05	30.39	94.55	84.74	67.39	43.85
C003	T	34.42	3.06	2.77	1.40	40.63	93.56	78.68	67.26	55.10
	NT	34.42	4.08	3.82	1.84	53.56	94.87	88.13	80.49	71.43
	LYM	34.42	4.40	4.11	1.85	53.74	95.33	89.67	83.00	74.25
C004	T	34.42	3.01	2.58	1.31	38.12	92.83	57.34	46.26	39.23
	NT	34.42	3.10	2.86	1.40	40.74	93.31	81.35	68.04	53.98
	LYM	34.42	3.07	2.71	1.11	32.22	91.20	66.52	52.51	40.37
C005	T	34.42	3.04	2.79	1.42	41.24	81.80	63.02	56.48	49.72
	NT	34.42	3.16	2.94	1.45	42.00	94.66	87.26	77.20	62.38
	LYM	34.42	2.40	2.23	1.07	31.20	94.58	85.02	67.51	44.56
Average		34.42	3.33	2.98	1.41	41.18	90.31	74.03	63.58	52.04

*T: tumor, NT: non-tumorous gastric mucosa, LYM: peripheral lymphocyte

Supplementary Table 5. Summary of somatic mutations in gastric tumor and non-tumorous gastritis tissues of 5 patients detected by whole exome sequencing.

MSI* status	Case	Tissue	Non-synonymous				Indel	Non-silent (Total)	Synonymous	Total	Mutations per Mb DNA
			Missense	Stop gained	Stop lost	Splice site					
MSS	C001	T	67	2	0	1	1	71	20	91	2.64
		NT	16	0	0	0	0	16	9	25	0.73
	C002	T	36	2	0	0	2	40	13	53	1.54
		NT	18	1	0	1	0	20	9	29	0.84
	C003	T	55	3	1	0	0	59	26	85	2.47
		NT	28	0	0	1	0	29	16	45	1.31
	C004	T	50	2	0	3	1	56	23	79	2.30
		NT	9	0	0	0	0	9	3	12	0.35
	Total(Average)	T	208(52)	9(2.25)	1(0.25)	4(1)	4(1)	226(56.5)	82(20.5)	308(77)	(2.24)
		NT	71(17.75)	1(0.25)	0(0)	2(0.5)	0(0)	74(18.5)	37(9.25)	111(27.75)	(0.81)
MSI	C005	T	344	16	0	7	61	428	156	584	16.97
		NT	17	0	0	0	0	17	14	31	0.90
Total(Average)		T	552(110.4)	25(5)	1(0.2)	11(2.2)	65(13)	654(130.8)	238(47.6)	892(178.4)	(5.18)
		NT	88(17.6)	1(0.2)	0(0)	2(0.4)	0(0)	91(18.2)	51(10.2)	142(28.4)	(0.83)

*MSI status was determined from gastric cancer tissue. MSI: microsatellite instability, MSS: microsatellite stability

Supplementary Table 6. List of somatic mutations in 5 gastric cancer tissues detected by whole exome sequencing.

Gene	Chr.*	Chr.* Position	Reference Nucleotide	Altered nucleotide	Coverage	Mutant allele frequency (%)	Case
PLCH2	1	2426991	C	T	112	28.6	C005
ESPN	1	6488310	G	T	65	33.9	C005
MASP2	1	11106960	G	A	34	32.4	C003
CLCN6	1	11888575	G	A	58	22.4	C003
DHRS3	1	12639439	A	C	47	14.9	C005
PRAMEF13	1	13448373	C	T	56	10.7	C001
PRAMEF13	1	13448406	C	T	60	10.0	C001
FBLIM1	1	16101147	T	G	50	14.0	C005
CLCNKA	1	16358311	A	G	144	11.1	C005
ARHGEF19	1	16534644	G	T	40	10.0	C003
SPATA21	1	16731487	G	A	87	31.0	C005
NBPF1	1	16909121	G	T	89	10.1	C005
CROCC	1	17272852	C	T	16	25.0	C003
MFAP2	1	17302194	G	A	129	34.9	C005
PADI2	1	17395587	G	A	61	27.9	C005
ARHGEF10L	1	17958883	G	A	83	33.7	C005
TAS1R2	1	19181327	C	T	112	38.4	C005
UBR4	1	19454224	T	C	96	36.5	C005
EIF4G3	1	21268546	A	C	58	31.0	C003
HSPG2	1	22204720	G	A	54	27.8	C005
CELA3A	1	22336297	T	A	49	10.2	C002
C1orf201	1	24687368	G	C	50	32.0	C005
SRRM1	1	24993374	C	A	23	21.7	C003
SRRM1	1	24995840	T	C	50	10.0	C004
AIM1L	1	26665825	C	T	49	26.5	C003
AIM1L	1	26665843	C	G	45	28.9	C003
AIM1L	1	26669310	C	G	61	27.9	C003
ARID1A	1	27097627	A	del	41	17.1	C005
ARID1A	1	27106252	C	del	52	32.7	C002
MAP3K6	1	27684706	G	A	111	24.3	C005
SPOCD1	1	32265038	G	C	68	47.1	C005
CCDC28B	1	32669643	T	G	70	27.1	C005
KIAA1522	1	33233501	C	del	51	23.5	C005

CSMD2	1	34112338	C	A	88	25.0	C005
C1orf212	1	35320839	T	A	23	17.4	C005
THRAP3	1	36752515	C	T	31	16.1	C005
C1orf102	1	36898152	T	C	59	35.6	C005
CSF3R	1	36935329	G	del	41	19.5	C005
ZC3H12A	1	37948975	G	A	44	31.8	C005
HIVEP3	1	41976927	T	G	158	10.1	C005
HIVEP3	1	42046305	G	A	29	37.9	C005
MUTYH	1	45797851	C	T	51	33.3	C005
FAAH	1	46870734	G	T	40	10.0	C001
CPT2	1	53676303	C	A	46	30.4	C005
ACOT11	1	55072865	C	A	107	36.5	C005
PCSK9	1	55512261	G	A	32	46.9	C003
RPE65	1	68914382	G	A	96	21.9	C005
LOC126987	1	106623728	C	T	40	27.5	C005
FNDC7	1	109260573	C	T	97	21.7	C005
CELSR2	1	109795922	T	C	53	17.0	C005
WNT2B	1	113059784	C	T	54	35.2	C005
PTGFRN	1	117504016	C	T	152	36.2	C005
TBX15	1	119441696	G	A	104	19.2	C005
NOTCH2	1	120539711	G	T	106	13.2	C005
PDE4DIP	1	144868126	G	A	197	15.2	C005
NBPF10	1	145301786	A	T	36	16.7	C001
NBPF10	1	145360696	A	G	33	18.2	C005
ITGA10	1	145537443	G	A	42	38.1	C005
GJA5	1	147231213	C	del	158	26.6	C005
NBPF14	1	148009349	A	G	61	13.1	C002
PLEKHO1	1	150129654	C	T	86	29.1	C005
TARS2	1	150477204	G	del	78	26.9	C005
PI4KB	1	151274417	C	T	58	34.5	C005
TUFT1	1	151547469	C	A	76	27.6	C005
HRNR	1	152187078	G	A	427	13.1	C005
HRNR	1	152187116	C	T	92	14.1	C001
HRNR	1	152187858	G	A	157	21.7	C005
HRNR	1	152190973	G	A	101	37.6	C005
FLG	1	152277698	C	G	65	10.8	C002
FLG	1	152282244	G	A	60	10.0	C002
FLG	1	152283783	G	A	79	15.2	C002

FLG	1	152284798	G	T	141	12.1	C003
SPRR3	1	152975679	C	T	87	10.3	C005
PMF1	1	156182865	G	T	101	30.7	C005
PEAR1	1	156882753	G	A	41	19.5	C003
FCRL5	1	157504612	T	A	41	17.1	C003
CD1E	1	158325818	C	T	118	39.0	C004
IFI16	1	159021773	G	A	46	23.9	C003
NCSTN	1	160326912	C	T	163	25.2	C005
GPR161	1	168066114	G	A	77	20.8	C005
RABGAP1L	1	174363186	G	T	63	36.5	C005
ASTN1	1	176903446	C	T	136	33.1	C001
PRG4	1	186276058	G	A	54	11.1	C005
LAD1	1	201356169	G	del	34	29.4	C005
LGR6	1	202288283	C	T	80	20.0	C005
NFASC	1	204948582	C	T	42	21.4	C003
PCTK3	1	205500491	G	A	88	34.1	C005
FCAMR	1	207134150	C	del	160	26.9	C005
PLXNA2	1	208257725	C	T	61	36.1	C005
HLX	1	221057894	G	A	36	25.0	C005
OBSCN	1	228437718	C	T	96	34.4	C005
OBSCN	1	228461503	C	T	33	15.2	C002
OBSCN	1	228475577	G	A	122	33.6	C005
OBSCN	1	228527746	G	A	68	38.2	C003
OBSCN	1	228538596	A	G	30	26.7	C005
PGBD5	1	230498111	C	T	49	14.3	C005
NID1	1	236187394	G	A	76	39.5	C005
OPN3	1	241803389	G	C	87	41.4	C001
LOC391343	2	905469	G	C	55	20.0	C005
LOC391343	2	905525	T	G	98	11.2	C001
PXDN	2	1670092	G	A	100	31.0	C005
HPCAL1	2	10559908	C	T	79	34.2	C005
DPYSL5	2	27157527	C	T	52	32.7	C003
CCDC121	2	27851284	T	G	23	43.5	C005
PLEKHH2	2	43992633	C	T	63	38.1	C005
TSPYL6	2	54483278	G	A	104	33.7	C005
SPTBN1	2	54856657	T	G	45	22.2	C003
OTX1	2	63280164	C	T	45	31.1	C005
PROKR1	2	68873113	A	T	177	26.0	C005

SFXN5	2	73226092	C	T	135	19.3	C001
DCTN1	2	74593492	G	del	69	21.7	C005
C2orf81	2	74642838	C	T	45	13.3	C005
CTNNA2	2	80085200	C	T	84	21.4	C004
ADRA2B	2	96781817	G	C	64	21.9	C001
FER1L5	2	97357006	C	T	61	24.6	C005
FAHD2B	2	97751577	C	T	71	11.3	C005
FAHD2B	2	97751909	G	A	121	12.4	C001
ZAP70	2	98351763	C	T	89	24.7	C005
TMEM131	2	98373675	T	C	325	24.6	C005
INPP4A	2	99182164	C	A	37	29.7	C005
POLR1B	2	113300209	C	A	41	26.8	C002
STEAP3	2	120005414	C	T	68	32.4	C005
CNTNAP5	2	125175050	G	A	108	37.0	C003
MAP3K2	2	128065299	T	C	27	25.9	C005
ACTBL3	2	132383821	G	A	41	31.7	C005
GPR39	2	133175389	C	T	49	26.5	C005
NAP5	2	133542575	G	A	58	19.0	C003
KCNH7	2	163693133	G	T	40	20.0	C003
SCN2A	2	166245565	C	T	13	23.1	C004
CSRNP3	2	166535902	C	T	65	44.6	C005
RAPGEF4	2	173900913	A	C	13	23.1	C004
C2orf21	2	210707133	T	A	22	18.2	C003
RUFY4	2	218940305	G	A	34	23.5	C003
ALPP	2	233245638	G	T	22	18.2	C001
ECEL1	2	233347309	G	del	63	28.6	C005
ESPNL	2	239009169	G	del	37	21.6	C005
MYEOV2	2	241066289	C	del	116	13.8	C005
SNED1	2	242002223	G	T	49	34.7	C005
LRRN1	3	3886594	A	G	42	31.0	C003
ATP2B2	3	10370709	C	T	100	31.0	C005
TMEM40	3	12778122	G	A	66	34.9	C005
CAND2	3	12869076	C	T	140	27.1	C005
FBLN2	3	13679286	T	C	96	20.8	C001
OSBPL10	3	31725451	G	A	134	35.8	C004
ITGA9	3	37512541	T	C	81	32.1	C005
VILL	3	38043310	C	del	143	30.8	C005
GORASP1	3	39142524	C	T	93	23.7	C005

MYRIP	3	40286031	C	T	60	36.7	C001
SLC6A20	3	45804477	G	A	175	23.4	C001
ALS2CL	3	46727807	G	A	50	30.0	C005
KLHL18	3	47385280	G	A	84	28.6	C005
IFRD2	3	50326277	C	T	97	44.3	C005
CACNA2D2	3	50404024	G	A	180	26.7	C005
DOCK3	3	51417610	C	del	67	31.3	C005
RAD54L2	3	51696598	G	A	49	32.7	C001
RAD54L2	3	51696966	C	A	37	32.4	C005
PCBP4	3	51994598	C	T	177	31.1	C005
TLR9	3	52257685	C	T	37	35.1	C005
DNAH1	3	52378506	G	A	74	28.4	C001
WNT5A	3	55504445	G	A	81	37.0	C005
IL17RD	3	57132127	C	T	126	23.8	C005
MITF	3	70008489	G	A	72	34.7	C005
PDZRN3	3	73433941	G	A	81	12.4	C003
GCET2	3	111849303	C	del	77	19.5	C005
GCET2	3	111849304	T	del	77	19.5	C005
KIAA1407	3	113724559	C	T	176	22.7	C005
NR1I2	3	119531577	G	A	114	34.2	C005
CASR	3	121980435	C	T	71	31.0	C005
CASR	3	122003779	C	T	94	26.6	C005
SEMA5B	3	122647429	C	T	78	25.6	C005
ITGB5	3	124560288	C	T	51	27.5	C005
ROPN1B	3	125694462	G	A	77	11.7	C002
ROPN1B	3	125701197	T	G	81	22.2	C003
ALDH1L1	3	125854472	C	T	242	32.2	C004
LOC100133263	3	131100912	G	A	47	40.4	C005
BFSP2	3	133119018	C	T	117	35.0	C005
PAQR9	3	142681816	C	T	67	11.9	C003
P2RY1	3	152554147	C	T	209	30.6	C005
VEPH1	3	156979073	G	del	78	29.5	C005
TRA2B	3	185643371	G	A	103	22.3	C005
LOC100131635	3	187451313	T	A	62	46.8	C001
MUC4	3	195508630	T	G	27	22.2	C004
MUC4	3	195513308	C	G	151	13.9	C004
MUC4	3	195514163	C	T	66	10.6	C004
KIAA1530	4	1343482	C	T	77	28.6	C005

WHSC1	4	1961236	G	A	26	19.2	C005
WHSC2	4	1985575	C	A	56	26.8	C005
C4orf8	4	2661599	C	del	96	31.3	C005
EVC2	4	5667362	G	A	96	14.6	C003
MAN2B2	4	6580193	C	T	78	28.2	C005
SORCS2	4	7677821	G	A	100	35.0	C005
SORCS2	4	7728553	C	T	198	30.8	C005
AFAP1	4	7770599	C	T	53	35.9	C005
LOC643446	4	11373912	G	A	76	10.5	C004
CPEB2	4	15067864	T	del	181	26.0	C005
CCKAR	4	26483599	G	A	201	24.9	C005
EPHA5	4	66217266	A	G	14	28.6	C005
FRAS1	4	79460492	G	T	12	25.0	C004
DSPP	4	88537393	C	T	301	31.2	C005
PITX2	4	111542380	C	T	82	30.5	C005
TNIP3	4	122078284	G	A	45	31.1	C005
QRFPR	4	122301634	C	T	99	33.3	C001
FAT4	4	126412066	C	A	42	23.8	C004
MAML3	4	140812019	G	A	51	13.7	C003
SH3RF1	4	170028106	C	A	24	16.7	C005
ODZ3	4	183714424	G	T	64	37.5	C005
C4orf41	4	184585151	G	A	40	20.0	C005
ACSL1	4	185689483	C	T	95	14.7	C001
FAT1	4	187539043	G	T	32	18.8	C005
SLC6A3	5	1432800	C	A	89	30.3	C005
CTNND2	5	11082865	C	T	100	34.0	C004
ZFR	5	32404048	G	A	90	34.4	C005
NPR3	5	32712217	C	T	23	21.7	C005
ADAMTS12	5	33643493	C	T	85	18.8	C002
RXFP3	5	33937912	C	T	86	26.7	C005
PPAP2A	5	54763775	C	A	55	12.7	C003
MCC	5	112478953	G	A	101	30.7	C005
PGGT1B	5	114598544	G	T	39	15.4	C001
FSTL4	5	132648398	G	A	142	33.1	C005
CATSPER3	5	134345161	G	A	104	33.7	C005
KDM3B	5	137762959	T	C	48	10.4	C003
PCDHA3	5	140182220	C	T	57	12.3	C003
PCDHA3	5	140182973	G	A	75	38.7	C005

PCDHA5	5	140202795	C	T	66	13.6	C001
PCDHA6	5	140208781	A	G	59	45.8	C004
PCDHA11	5	140249922	C	A	204	28.9	C005
PCDHA11	5	140250887	G	A	55	20.0	C005
PCDHA13	5	140263909	G	A	34	26.5	C001
PCDHB5	5	140516419	A	C	33	15.2	C001
PCDHB5	5	140516622	G	A	30	33.3	C005
PCDHB5	5	140516948	G	T	12	33.3	C001
PCDHB16	5	140562776	G	A	28	32.1	C005
PCDHB12	5	140590609	G	A	43	16.3	C005
PCDHGB2	5	140740377	G	A	74	29.7	C005
PCDHGB7	5	140799251	G	A	62	33.9	C005
PCDHGA12	5	140810464	C	T	55	23.6	C005
PCDHGC3	5	140857572	C	A	61	16.4	C004
FAT2	5	150932787	C	T	118	29.7	C005
FAT2	5	150933985	G	A	85	16.5	C002
SPARC	5	151045951	G	A	90	31.1	C005
CYFIP2	5	156787350	C	T	143	31.5	C005
CCNJL	5	159682655	C	A	62	46.8	C005
WWC1	5	167836926	G	T	61	45.9	C005
RNF44	5	175959091	G	A	31	29.0	C005
NSD1	5	176665375	G	T	43	25.6	C005
B4GALT7	5	177034501	G	A	37	37.8	C005
RMND5B	5	177571070	G	A	37	16.2	C005
TBC1D9B	5	179292334	A	C	48	14.6	C003
FLT4	5	180057082	G	A	73	37.0	C005
FLT4	5	180058720	G	T	108	26.9	C001
LY86	6	6654816	C	T	123	24.4	C005
JARID2	6	15497107	T	G	49	14.3	C005
HIST1H3A	6	26020836	A	C	56	10.7	C004
HFE	6	26091632	A	G	129	34.9	C005
BTN3A2	6	26369014	G	T	46	10.9	C001
HLA-G	6	29797337	C	T	22	18.2	C002
RDBP	6	31922386	G	A	58	15.5	C005
C4A	6	31950148	G	T	22	18.2	C001
RING1	6	33177797	G	T	75	29.3	C005
KIFC1	6	33374374	C	A	147	27.9	C005
SPDEF	6	34512079	C	T	64	35.9	C005

MOCS1	6	39880691	C	A	172	37.2	C005
COL19A1	6	70881875	G	T	24	16.7	C003
SNAP91	6	84417591	C	T	100	31.0	C005
CNR1	6	88853808	T	C	203	33.0	C005
ANKRD6	6	90337333	G	A	126	34.9	C005
SOBP	6	107954989	C	T	42	35.7	C005
HS3ST5	6	114379248	G	A	54	27.8	C005
C6orf174	6	127797286	C	T	164	29.9	C005
LAMA2	6	129371164	C	T	107	13.1	C003
MTHFD1L	6	151265652	G	A	19	42.1	C005
AKAP12	6	151672212	G	A	67	25.4	C005
SYNE1	6	152660445	G	T	12	33.3	C004
CNKSR3	6	154735493	T	G	109	47.7	C005
PNLDC1	6	160239666	C	T	42	26.2	C005
PDE10A	6	165801822	G	A	113	25.7	C005
HGC6.3	6	168377302	T	G	112	10.7	C001
FAM20C	7	208913	G	del	129	20.9	C005
CARD11	7	2959240	C	T	67	38.8	C005
SLC29A4	7	5342503	C	A	39	28.2	C005
THSD7A	7	11676041	G	A	55	32.7	C005
HECW1	7	43506111	C	T	47	34.0	C003
AEBP1	7	44151452	G	T	74	25.7	C005
MYO1G	7	45016560	C	T	79	26.6	C005
RAMP3	7	45216937	G	A	153	26.1	C005
COBL	7	51095704	T	C	74	31.1	C005
DKFZp564N2472	7	53104085	G	T	55	16.4	C005
PHKG1	7	56148818	G	A	35	31.4	C005
PHKG1	7	56149726	C	T	57	35.1	C005
LIMK1	7	73535503	A	C	73	11.0	C001
SPDYE5	7	75130753	G	A	78	10.3	C003
SPDYE5	7	75131015	G	A	112	25.9	C005
DTX2	7	76112183	C	T	35	22.9	C002
PHTF2	7	77572020	C	T	44	13.6	C003
FZD1	7	90895559	C	T	115	36.5	C005
C7orf43	7	99754100	G	A	136	36.0	C005
LOC100287259	7	100275329	G	A	40	37.5	C001
MUC12	7	100634718	T	A	70	32.9	C001
MUC17	7	100681407	C	T	29	17.2	C001

MUC17	7	100684792	T	C	73	16.4	C005
MUC17	7	100696406	C	T	52	17.3	C003
LHFPL3	7	103969671	C	T	99	10.1	C004
CAV1	7	116199088	C	T	116	26.7	C005
WDR91	7	134878026	G	T	78	33.3	C005
KLRG2	7	139164414	A	T	62	25.8	C001
HIPK2	7	139258060	C	T	223	30.5	C005
KIAA1147	7	141362476	C	T	36	22.2	C005
TRBV4-1	7	142013475	C	T	43	11.6	C001
TRBV7-8	7	142099512	T	G	46	19.6	C001
TRBV5-6	7	142131625	G	A	186	26.9	C005
TRBV6-7	7	142144059	A	C	81	12.4	C004
PRSS1	7	142458486	C	T	96	14.6	C003
EPHB6	7	142561808	A	G	31	29.0	C005
FAM131B	7	143053859	C	del	75	16.0	C005
FASTK	7	150776013	G	A	25	32.0	C005
SMARCD3	7	150936192	G	T	98	28.6	C005
CSMD1	8	2886903	C	A	98	40.8	C005
FAM90A18	8	7582817	C	A	27	18.5	C005
PRAGMIN	8	8185372	C	A	76	30.3	C005
RP1L1	8	10465053	T	A	21	19.1	C003
AMAC1L2	8	11189184	C	T	59	13.6	C001
C8orf79	8	12863787	G	A	135	34.1	C005
ASAH1	8	17942298	T	C	79	30.4	C005
SLC18A1	8	20030574	G	A	82	31.7	C005
KIAA1967	8	22472981	C	del	47	14.9	C005
CLU	8	27468014	C	T	94	30.9	C005
PURG	8	30889693	C	T	58	25.9	C005
PURG	8	30889710	G	A	58	25.9	C005
PLEKHA2	8	38809732	C	A	175	29.7	C005
SLC20A2	8	42302216	G	A	189	32.8	C005
POTEA	8	43147917	A	G	37	16.2	C001
C8orf46	8	67425833	C	T	108	29.6	C005
SLCO5A1	8	70744425	G	A	247	15.4	C001
EYA1	8	72234004	G	T	110	28.2	C005
LOC100289448	8	86727125	C	A	191	12.0	C001
LOC100288500	8	86757709	G	T	311	13.2	C001
WDR21C	8	88885151	G	T	164	36.0	C005

ASAP1	8	131193108	C	T	14	21.4	C001
SLC45A4	8	142228711	G	del	20	15.0	C005
BAI1	8	143625704	G	A	59	27.1	C005
LY6H	8	144240250	C	T	87	12.6	C002
FAM83H	8	144812319	C	T	38	36.8	C005
SCRIB	8	144889782	T	C	68	10.3	C003
EPPK1	8	144940649	G	A	45	20.0	C005
PLEC1	8	144992908	C	T	21	47.6	C005
SPATC1	8	145096220	G	A	31	22.6	C003
NFKBIL2	8	145657757	C	T	40	32.5	C005
VLDLR	9	2643532	G	A	68	29.4	C005
FREM1	9	14819410	C	T	20	30.0	C004
LINGO2	9	27949778	G	A	44	31.8	C005
AQP7	9	33385690	G	T	47	12.8	C004
PRSS3	9	33796799	A	T	66	12.1	C003
GBA2	9	35737877	C	A	108	31.5	C005
PAX5	9	36966647	C	A	100	30.0	C002
LOC100289027	9	45363760	G	A	51	21.6	C001
LOC100289027	9	45363775	G	A	47	14.9	C001
TRPM3	9	73213484	C	T	15	20.0	C001
RORB	9	77277447	G	A	42	19.1	C005
RORB	9	77282793	T	A	20	20.0	C004
RORB	9	77282795	A	T	21	19.1	C004
FLJ46321	9	84610009	G	A	36	27.8	C003
DAPK1	9	90272943	G	C	91	31.9	C005
WNK2	9	96055219	G	A	35	31.4	C005
PTCH1	9	98231278	C	T	349	34.1	C005
PTCH1	9	98231437	T	G	137	28.5	C005
NANS	9	100823141	A	G	97	10.3	C003
COL15A1	9	101748136	C	T	128	33.6	C005
COL15A1	9	101818554	G	del	53	26.4	C005
SEC61B	9	101990241	G	del	99	18.2	C005
ASTN2	9	119976838	G	A	136	27.2	C004
CRB2	9	126133707	C	A	76	35.5	C005
LHX2	9	126776345	C	T	72	37.5	C005
ZNF79	9	130198240	G	A	53	24.5	C005
ZNF79	9	130207020	G	A	65	29.2	C005
CRAT	9	131862896	G	A	99	11.1	C002

POMT1	9	134394290	C	T	62	24.2	C005
OBP2B	9	136082717	C	T	81	37.0	C005
SURF6	9	136199149	C	T	54	27.8	C005
REXO4	9	136282883	G	A	66	10.6	C001
COL5A1	9	137622238	G	del	60	31.7	C005
COL5A1	9	137686973	G	T	161	24.8	C004
EGFL7	9	139564377	C	T	29	27.6	C005
MAMDC4	9	139753547	A	G	54	11.1	C001
ABCA2	9	139903068	C	T	44	27.3	C005
ARRDC1	9	140509557	C	T	78	16.7	C001
CACNA1B	9	140777319	G	A	288	10.4	C001
CACNA1B	9	141016368	G	A	21	23.8	C005
IDI2	10	1065522	G	A	75	29.3	C005
PFKP	10	3172122	G	A	39	23.1	C005
DHTKD1	10	12129577	G	A	109	21.1	C005
CUBN	10	16990548	G	A	105	30.5	C005
SLC39A12	10	18276413	G	A	50	12.0	C002
NRP1	10	33491819	C	A	160	35.6	C005
ANKRD30A	10	37430774	G	A	29	17.2	C005
GDF2	10	48414000	C	T	170	25.3	C001
WDFY4	10	49988021	G	T	238	32.8	C004
LRRC18	10	50121609	A	G	140	15.0	C003
LRRC18	10	50121789	T	del	186	23.1	C005
RHOBTB1	10	62671250	A	G	208	31.3	C005
H2AFY2	10	71851671	G	del	123	25.2	C005
ADAMTS14	10	72492114	G	A	87	31.0	C005
OIT3	10	74684254	C	G	40	32.5	C003
SYNPO2L	10	75412937	T	G	48	18.8	C005
BLNK	10	97987185	G	A	74	36.5	C005
PIK3AP1	10	98362071	G	del	72	22.2	C005
HPS1	10	100186994	G	del	69	20.3	C005
CALHM1	10	105218033	A	C	40	10.0	C002
SORCS1	10	108389029	A	G	97	18.6	C002
BUB3	10	124914523	G	A	94	31.9	C005
CPXM2	10	125521483	A	C	82	31.7	C005
MMP21	10	127459002	T	C	20	15.0	C004
MMP21	10	127459004	G	T	20	15.0	C004
C10orf141	10	128974522	C	A	85	38.8	C005

TCERG1L	10	132915194	G	A	97	37.1	C005
PPP2R2D	10	133769244	G	A	47	14.9	C002
STK32C	10	134041533	G	A	61	21.3	C003
STK32C	10	134041540	G	A	110	29.1	C005
C10orf91	10	134261282	G	T	26	15.4	C003
TUBGCP2	10	135098641	C	T	65	35.4	C005
TUBGCP2	10	135099005	G	A	47	14.9	C002
FRG2B	10	135438955	C	T	114	11.4	C004
SIGIRR	11	405954	G	A	22	36.4	C005
MUC5AC	11	1262398	G	A	42	35.7	C005
LOC100287770	11	1780789	C	T	63	28.6	C005
KCNQ1	11	2604728	G	A	258	27.1	C005
FAM160A2	11	6238867	C	T	19	31.6	C003
SMPD1	11	6413274	A	G	87	26.4	C005
DNHD1	11	6592310	C	T	53	45.3	C005
NLRP10	11	7982170	T	A	81	34.6	C005
FIBIN	11	27016359	C	T	56	30.4	C005
TSPAN18	11	44940818	C	T	117	29.9	C005
SYT13	11	45277241	C	A	109	25.7	C001
MADD	11	47317468	G	C	116	26.7	C001
DTX4	11	58962833	C	del	57	22.8	C005
MS4A15	11	60535054	C	T	42	23.8	C005
ZP1	11	60640661	C	T	55	32.7	C005
TMEM132A	11	60701074	G	del	37	27.0	C005
FADS3	11	61645990	G	A	110	37.3	C005
FADS3	11	61646013	C	T	138	26.1	C005
AHNAK	11	62295182	C	T	104	25.0	C005
SLC22A10	11	63057902	C	T	54	20.4	C005
NRXN2	11	64427858	C	T	272	32.4	C005
RASGRP2	11	64509500	G	A	63	30.2	C005
ARL2	11	64786120	A	G	92	38.0	C005
SNX15	11	64802386	G	A	79	13.9	C005
SYVN1	11	64900204	G	C	111	30.6	C001
SPDYC	11	64940669	C	A	73	34.3	C005
KCNK7	11	65361001	G	A	24	41.7	C005
LOC100288054	11	65487687	C	T	95	28.4	C001
OVOL1	11	65554874	G	A	11	27.3	C003
OVOL1	11	65562167	C	T	77	26.0	C005

MUS81	11	65633334	C	T	124	30.7	C005
FIBP	11	65655144	G	C	158	10.8	C004
ADRBK1	11	67051835	C	del	44	20.5	C001
KRTAP5-7	11	71238595	T	C	63	12.7	C001
KRTAP5-7	11	71238639	A	G	94	11.7	C004
CLPB	11	72004466	C	T	223	27.4	C005
C2CD3	11	73814359	T	C	90	35.6	C005
CAPN5	11	76831912	G	A	82	18.3	C005
PRSS23	11	86518919	C	T	57	21.1	C005
NOX4	11	89177310	C	T	98	20.4	C003
FAT3	11	92623942	T	C	39	30.8	C001
MED17	11	93542960	T	C	78	38.5	C005
MMP20	11	102487712	T	C	55	25.5	C005
BUD13	11	116628653	G	T	14	28.6	C001
APOA4	11	116692268	C	T	77	33.8	C005
CEP164	11	117222658	A	del	37	18.9	C005
CEP164	11	117242153	C	G	25	16.0	C001
DSCAML1	11	117306456	T	C	105	32.4	C005
TMPRSS13	11	117789302	C	T	213	38.5	C005
AMICA1	11	118081237	G	A	68	32.4	C005
ABCG4	11	119030980	C	T	66	53.0	C005
GRIK4	11	120833189	C	T	39	35.9	C005
SORL1	11	121384984	G	A	118	32.2	C005
VWA5A	11	124007891	G	A	48	41.7	C003
ADAMTS15	11	130332143	G	A	75	18.7	C005
B3GAT1	11	134253822	G	A	35	45.7	C005
CCND2	12	4398153	C	A	52	36.5	C005
ANO2	12	6031908	G	A	63	27.0	C005
NCAPD2	12	6635479	C	del	81	11.1	C005
ACRBP	12	6756016	C	T	126	30.2	C005
LEPREL2	12	6939661	G	del	35	28.6	C005
LEPREL2	12	6948624	G	A	69	26.1	C005
ATN1	12	7045891	A	G	76	13.2	C001
CD163	12	7653798	G	T	12	25.0	C004
CD163	12	7653836	G	A	53	28.3	C005
AICDA	12	8757454	C	A	23	17.4	C003
A2ML1	12	9027066	G	T	14	21.4	C001
A2M	12	9227320	C	G	191	12.6	C005

PRB1	12	11506749	T	C	156	12.2	C001
PRB1	12	11506774	C	T	131	12.2	C001
PRB1	12	11506774	C	T	51	13.7	C003
ETV6	12	12022707	C	T	86	37.2	C005
HDAC7	12	48189493	T	C	36	19.4	C005
HDAC7	12	48190821	G	A	49	28.6	C005
COL2A1	12	48373318	G	A	82	22.0	C005
DDN	12	49391616	G	T	40	10.0	C001
TUBA1B	12	49522138	C	T	41	12.2	C005
TUBA1B	12	49522146	C	T	34	14.7	C005
KRT81	12	52680246	G	A	30	26.7	C005
KRT6C	12	52865971	G	A	88	12.5	C005
KRT6A	12	52885427	G	A	55	32.7	C005
KRT3	12	53189431	T	A	79	19.0	C004
KRT8	12	53291319	A	C	20	15.0	C005
SOAT2	12	53515124	G	A	103	26.2	C005
ESPL1	12	53662900	C	T	48	25.0	C005
MYO1A	12	57422580	T	del	93	29.0	C005
STAT6	12	57500084	T	C	34	41.2	C005
B4GALNT1	12	58025870	C	A	14	28.6	C003
PLXNC1	12	94543515	C	T	65	38.5	C005
HSP90B1	12	104327983	C	T	48	10.4	C003
ACACB	12	109665250	C	T	43	18.6	C002
MYO1H	12	109863839	A	C	39	28.2	C005
MVK	12	110023903	G	A	94	25.5	C005
CUX2	12	111733155	G	C	62	50.0	C004
RPH3A	12	113307724	C	A	88	33.0	C005
DDX54	12	113614871	G	A	60	30.0	C005
DDX54	12	113618771	G	C	40	12.5	C004
TPCN1	12	113716624	G	A	88	38.6	C005
LHX5	12	113907005	C	T	142	31.0	C005
NOS1	12	117703236	C	del	65	27.7	C005
P2RX7	12	121598721	C	T	115	23.5	C005
TMEM132B	12	126137117	C	T	65	12.3	C005
RIMBP2	12	130907041	G	A	64	37.5	C005
RIMBP2	12	130927127	G	A	231	23.4	C005
GPR133	12	131593323	G	A	224	39.7	C004
ULK1	12	132394535	G	A	61	27.9	C005

NOC4L	12	132635531	G	del	57	22.8	C005
POLE	12	133225617	C	T	72	37.5	C005
CDK8	13	26923245	C	G	31	19.4	C004
CDK8	13	26923247	T	C	31	19.4	C004
STARD13	13	33704350	C	T	53	24.5	C005
LHFP	13	39952604	G	A	30	26.7	C005
KBTBD6	13	41705951	C	T	62	29.0	C005
LECT1	13	53298164	C	T	58	31.0	C005
PCDH20	13	61987890	G	del	40	25.0	C005
LMO7	13	76427460	C	T	50	28.0	C005
SLITRK1	13	84454271	C	T	79	40.5	C005
GPC6	13	93879809	C	T	89	39.3	C005
GPC6	13	94482609	G	T	80	16.3	C003
DZIP1	13	96274710	T	C	19	31.6	C004
DZIP1	13	96274711	G	C	18	33.3	C004
DOCK9	13	99449474	C	T	110	16.4	C003
COL4A1	13	110827680	G	T	120	36.7	C005
ATP11A	13	113513709	C	T	76	14.5	C002
TFDP1	13	114285948	C	T	83	30.1	C005
ATP4B	13	114304771	T	A	113	29.2	C001
UPF3A	13	115064429	G	A	116	11.2	C003
FLJ10357	14	21543144	T	G	89	27.0	C005
MYH7	14	23893131	G	A	193	28.0	C005
MYH7	14	23893979	G	A	86	39.5	C005
RABGGTA	14	24736889	G	A	135	29.6	C005
RABGGTA	14	24738874	C	A	20	15.0	C002
LTB4R2	14	24780676	C	T	29	31.0	C005
FAM179B	14	45433274	T	del	46	23.9	C005
C14orf105	14	57957742	T	C	11	36.4	C003
PRKCH	14	62014511	C	T	71	23.9	C001
SYT16	14	62541988	G	A	85	28.2	C005
ZFYVE26	14	68248094	A	G	57	12.3	C004
PROX2	14	75329377	C	T	126	15.1	C002
NEK9	14	75563893	T	C	36	30.6	C004
TTC7B	14	91113312	G	A	50	20.0	C005
RIN3	14	93107643	G	A	79	27.9	C005
GSC	14	95234857	C	A	69	30.4	C005
C14orf49	14	95906260	G	A	41	26.8	C005

TCL1B	14	96152917	G	A	92	33.7	C005
DYNC1H1	14	102494319	T	G	149	32.9	C005
KLC1	14	104153440	G	A	58	20.7	C005
KIF26A	14	104641501	C	T	42	23.8	C005
KIAA0284	14	105360126	C	T	68	30.9	C005
AHNAK2	14	105411282	C	T	45	20.0	C002
AHNAK2	14	105411302	G	C	60	10.0	C005
APBA2	15	29346195	G	A	113	38.1	C005
APBA2	15	29385365	G	A	80	26.3	C004
APBA2	15	29385398	C	T	75	14.7	C004
DISP2	15	40656020	G	del	36	30.6	C005
BAHD1	15	40756123	C	T	138	33.3	C005
ITPKA	15	41794314	G	A	54	37.0	C005
SPTBN5	15	42147446	G	A	55	25.5	C005
EHD4	15	42211518	G	A	64	21.9	C004
PLA2G4D	15	42373254	C	T	145	27.6	C005
ZSCAN29	15	43658358	G	del	36	30.6	C005
MAP1A	15	43814660	C	T	89	10.1	C005
HISPPD2A	15	43869101	C	T	57	84.2	C005
CKMT1B	15	43888404	G	T	14	21.4	C004
DUOX1	15	45453075	G	A	196	36.2	C005
RAB27A	15	55497786	A	G	107	37.4	C005
ZNF609	15	64966472	C	T	28	17.9	C004
RBPMS2	15	65041342	C	T	124	25.8	C005
CILP	15	65497702	A	G	51	23.5	C004
MEGF11	15	66214766	G	T	40	32.5	C005
FEM1B	15	68583279	T	C	83	36.1	C005
ITGA11	15	68643603	G	A	290	22.4	C001
GOLGA6	15	74363614	T	G	69	27.5	C005
STRA6	15	74473190	C	T	92	22.8	C005
CSPG4	15	75979962	G	A	30	16.7	C004
CSPG4	15	75979966	G	A	31	16.1	C004
KIAA1199	15	81199156	G	A	152	33.6	C005
ZSCAN2	15	85163990	G	A	51	27.5	C001
ACAN	15	89398426	C	A	148	19.6	C001
MFGE8	15	89444938	A	G	72	19.4	C004
MFGE8	15	89444941	C	A	74	13.5	C004
PIGQ	16	629177	C	T	57	31.6	C005

LOC388199	16	855758	G	A	43	34.9	C005
CACNA1H	16	1251719	G	A	69	33.3	C005
TPSD1	16	1306897	A	C	59	13.6	C004
TSC2	16	2103375	G	A	50	36.0	C005
TRAF7	16	2225869	T	C	128	38.3	C005
LOC342346	16	4650180	A	C	28	17.9	C005
PPL	16	4933723	C	T	50	36.0	C005
USP7	16	9012968	C	T	65	33.9	C003
CLEC16A	16	11272353	A	G	77	29.9	C005
MYH11	16	15833980	C	T	55	36.4	C005
C16orf88	16	19725715	T	del	29	37.9	C005
ACSM5	16	20432663	C	A	71	11.3	C005
ACSM5	16	20432668	A	G	75	10.7	C005
DNAH3	16	20975977	G	A	57	28.1	C005
DNAH3	16	21011737	C	T	166	31.9	C005
LOC100132247	16	22545412	G	C	48	12.5	C001
COG7	16	23409386	G	A	164	35.4	C002
AQP8	16	25232880	G	del	58	10.3	C005
GTF3C1	16	27473782	C	T	73	31.5	C005
GTF3C1	16	27481557	G	A	65	35.4	C005
ATXN2L	16	28842344	G	A	23	21.7	C005
SULT1A4	16	29474926	T	C	21	14.3	C005
SRCAP	16	30740332	G	del	50	20.0	C005
HSD3B7	16	30998172	G	del	61	23.0	C005
ZNF668	16	31075681	C	T	54	13.0	C004
ITGAX	16	31393213	G	A	176	28.4	C005
ABCC12	16	48139204	C	T	83	24.1	C003
TOX3	16	52484345	G	T	174	38.5	C005
CCL22	16	57397478	T	A	69	33.3	C005
CNGB1	16	57937783	C	T	214	26.2	C002
CNGB1	16	57949153	C	A	41	26.8	C005
NDRG4	16	58541875	G	T	83	25.3	C005
EXOC3L	16	67218697	G	A	37	27.0	C005
ATP6V0D1	16	67472463	C	T	116	39.7	C005
SMPD3	16	68398755	T	C	54	33.3	C005
HAS3	16	69148752	G	A	46	41.3	C005
COG4	16	70515032	C	T	43	48.8	C001
DHX38	16	72139214	G	T	23	17.4	C003

WDR59	16	74976634	A	G	58	31.0	C005
WWOX	16	78458865	G	A	30	16.7	C005
CDYL2	16	80718797	G	A	63	12.7	C002
PKD1L2	16	81187715	T	del	157	24.2	C005
PKD1L2	16	81187716	C	del	157	24.2	C005
CMIP	16	81697920	C	T	80	35.0	C005
ADAD2	16	84227774	C	T	122	36.1	C005
ADAD2	16	84229289	G	A	38	26.3	C005
GINS2	16	85721133	C	T	56	37.5	C005
FBXO31	16	87364932	C	T	66	22.7	C005
ZFPM1	16	88598619	G	A	28	17.9	C005
ZC3H18	16	88691148	C	del	77	20.8	C005
C16orf84	16	88773602	G	A	117	33.3	C005
SPATA2L	16	89763924	C	T	26	19.2	C005
SERPINF1	17	1680003	G	A	55	21.8	C005
ZZEF1	17	3917764	C	T	53	22.6	C005
FAM64A	17	6348621	G	T	23	17.4	C001
EIF5A	17	7214673	T	A	148	31.8	C005
ACAP1	17	7246835	C	T	80	26.3	C005
C17orf74	17	7329986	C	A	57	28.1	C005
ZBTB4	17	7367129	C	A	116	20.7	C004
AMAC1L3	17	7385440	G	A	100	12.0	C004
AMAC1L3	17	7385463	G	A	66	12.1	C005
TP53	17	7574003	G	A	40	27.5	C003
TP53	17	7578176	C	T	82	35.4	C001
TP53	17	7578190	T	C	123	36.6	C004
PIK3R5	17	8785183	G	A	101	24.8	C002
MYH13	17	10265589	G	A	81	75.3	C005
MYH3	17	10535273	C	T	36	30.6	C005
MYH3	17	10555781	C	T	183	33.3	C005
UBB	17	16285211	A	T	13	23.1	C004
MPRIP	17	17039690	C	T	58	12.1	C002
PEMT	17	17415916	C	T	41	12.2	C002
LGALS9C	17	18396063	C	T	70	27.1	C005
FBXW10	17	18671915	T	C	180	10.6	C004
FLJ25006	17	26939072	C	A	24	25.0	C003
SUPT6H	17	27005608	T	G	18	22.2	C004
SUPT6H	17	27013715	G	T	92	18.5	C002

MYO18A	17	27425932	A	G	172	23.3	C005
BLMH	17	28618424	A	T	64	12.5	C003
ZNF830	17	33289114	A	del	58	19.0	C005
AMAC1	17	33520469	T	C	364	18.7	C004
AMAC1	17	33520744	C	G	94	13.8	C004
TAF15	17	34171513	G	T	43	39.5	C002
SNIP	17	36719655	G	A	99	28.3	C005
KRT222P	17	38816302	G	A	46	23.9	C005
KRT25	17	38907422	C	T	181	55.3	C001
KRTAP1-3	17	39190604	A	G	54	13.0	C002
KRTAP4-9	17	39262015	T	C	29	41.4	C002
KRTAP9-4	17	39406419	T	C	361	13.6	C001
KLHL10	17	40001889	G	A	36	22.2	C005
KCNH4	17	40321565	G	A	70	30.0	C005
HSD17B1	17	40706517	A	C	40	12.5	C002
TUBG1	17	40762136	G	A	102	19.6	C001
EZH1	17	40854956	G	A	41	24.4	C005
AOC2	17	41001333	C	del	64	14.1	C002
RUNDC1	17	41143048	A	G	41	14.6	C002
FMNL1	17	43321166	T	C	43	37.2	C005
CDC27	17	45232084	G	A	20	15.0	C003
HOXB13	17	46804366	C	T	72	27.8	C005
B4GALNT2	17	47246132	A	G	53	34.0	C003
ABCC3	17	48746772	C	T	91	27.5	C005
TRIM25	17	54969327	C	T	61	41.0	C005
AKAP1	17	55183041	C	del	65	27.7	C005
MPO	17	56350865	G	A	199	21.1	C003
TBX2	17	59480520	G	A	69	23.2	C004
FTSJ3	17	61901225	G	A	107	30.8	C005
CPSF4L	17	71257898	C	A	92	29.4	C005
SDK2	17	71412001	C	T	103	33.0	C005
CD300LB	17	72522003	C	T	111	19.8	C003
CDR2L	17	72998268	C	T	98	35.7	C005
KIAA0195	17	73491040	G	T	21	19.1	C003
ENPP7	17	77705039	C	G	54	20.4	C005
AATK	17	79095189	G	C	53	39.6	C005
BAHCC1	17	79410861	G	A	63	34.9	C005
BAHCC1	17	79423472	G	del	24	20.8	C005

SIRT7	17	79872024	C	T	57	36.8	C005
LRRC45	17	79988336	C	T	57	17.5	C005
KIAA0802	18	8819106	C	T	64	20.3	C005
FAM59A	18	29867267	G	A	79	34.2	C005
FHOD3	18	34205521	C	del	73	31.5	C005
BRUNOL4	18	34846517	C	T	117	39.3	C005
SLC14A2	18	43217006	G	A	61	16.4	C002
LOXHD1	18	44143068	C	T	83	25.3	C004
TCEB3B	18	44561188	C	T	29	31.0	C003
ZBTB7C	18	45566525	C	del	49	26.5	C005
MYO5B	18	47369760	A	G	160	24.4	C004
TSHZ1	18	72998565	C	T	171	32.8	C005
MBP	18	74701972	G	T	130	26.2	C005
HCN2	19	608177	G	A	36	16.7	C001
HCN2	19	610278	A	G	28	28.6	C005
ABCA7	19	1051523	C	T	48	29.2	C005
ABCA7	19	1056206	C	A	79	36.7	C005
STK11	19	1223060	C	T	66	33.3	C001
ADAMTSL5	19	1506749	G	A	46	28.3	C005
GNA11	19	3110202	C	T	65	30.8	C005
GNA15	19	3151769	C	T	65	33.9	C005
ZBTB7A	19	4054973	G	T	75	29.3	C005
C19orf10	19	4660726	G	T	79	29.1	C002
PTPRS	19	5220117	G	A	64	39.1	C005
PTPRS	19	5231398	C	T	63	30.2	C005
PTPRS	19	5258061	G	A	118	29.7	C005
HSD11B1L	19	5684874	G	del	73	23.3	C005
C19orf45	19	7570299	A	C	34	17.7	C002
MUC16	19	8994169	T	del	34	23.5	C004
MUC16	19	8994171	C	G	34	23.5	C004
ZNF317	19	9271868	C	T	47	25.5	C005
EIF3G	19	10227832	G	del	59	28.8	C005
DNMT1	19	10259679	G	A	65	20.0	C005
LDLR	19	11224397	C	T	81	29.6	C005
TNPO2	19	12812612	C	T	83	10.8	C002
LPHN1	19	14288366	G	A	49	20.4	C004
SYDE1	19	15223299	G	A	55	29.1	C005
NOTCH3	19	15276841	A	C	48	29.2	C005

F2RL3	19	17001217	T	G	32	34.4	C005
VSTM2B	19	30018171	G	A	55	30.9	C005
ZNF536	19	30936131	C	T	442	47.1	C001
TSHZ3	19	31769744	C	T	126	10.3	C005
RHPN2	19	33493767	G	C	46	15.2	C002
GAPDHS	19	36034261	C	T	67	11.9	C005
ETV2	19	36135664	G	del	64	18.8	C005
ALKBH6	19	36504261	C	T	68	26.5	C005
CLIP3	19	36509915	G	A	78	32.1	C005
PSMD8	19	38872784	G	T	63	28.6	C005
RYR1	19	38951103	C	T	69	30.4	C005
DYRK1B	19	40319210	C	T	89	15.7	C005
FCGBP	19	40376798	G	T	17	23.5	C001
FCGBP	19	40383993	T	G	58	12.1	C005
SPTBN4	19	41009887	C	T	68	32.4	C005
CYP2A6	19	41354198	T	C	64	10.9	C004
CYP2A6	19	41354271	A	C	50	12.0	C002
CYP2A13	19	41601004	C	T	114	14.0	C004
AXL	19	41763488	C	A	67	41.8	C004
BCKDHA	19	41928215	C	T	80	31.3	C005
CEACAM7	19	42181342	A	G	100	33.0	C003
ATP1A3	19	42489525	G	A	72	31.9	C005
MEGF8	19	42862345	C	T	63	30.2	C005
PSG7	19	43430623	T	C	16	37.5	C001
PSG9	19	43772245	G	A	26	15.4	C004
CBLC	19	45293331	C	T	53	13.2	C005
PPP1R13L	19	45901345	A	C	31	12.9	C003
SIGLEC8	19	51958806	C	T	88	33.0	C005
LILRB3	19	54721311	C	T	184	17.4	C005
LILRB3	19	54726833	T	C	78	12.8	C002
LILRA1	19	55106355	T	A	154	10.4	C004
RDH13	19	55568131	G	A	57	26.3	C005
SYT5	19	55689604	A	G	52	26.9	C005
ZSCAN5B	19	56701512	C	T	115	35.7	C005
ZNF324B	19	58967473	G	A	46	41.3	C005
SLC27A5	19	59010016	G	A	159	27.0	C005
TRIB3	20	372049	T	del	32	40.6	C005
FAM110A	20	825621	C	T	22	36.4	C005

SIRPD	20	1532466	T	C	97	13.4	C005
SIRPA	20	1918111	C	T	146	25.3	C005
TMC2	20	2597797	G	A	196	26.0	C005
RRBP1	20	17606191	C	A	58	31.0	C005
ENTPD6	20	25198164	C	A	73	28.8	C005
TTLL9	20	30497635	C	T	103	21.4	C004
CEP250	20	34053596	T	C	140	32.1	C005
C20orf4	20	34828371	G	A	47	34.0	C005
DLGAP4	20	35060706	C	T	92	12.0	C005
SLA2	20	35262904	G	A	80	10.0	C003
SLC32A1	20	37356246	G	A	185	27.0	C001
CHD6	20	40042067	G	A	45	37.8	C004
PTPRT	20	40743924	G	A	194	10.3	C002
MATN4	20	43926950	C	T	63	17.5	C005
ZSWIM3	20	44505590	G	A	69	24.6	C005
ZNFX1	20	47886836	T	C	28	21.4	C005
KCNB1	20	47991241	C	T	67	23.9	C002
B4GALT5	20	48252912	G	A	158	26.6	C005
CYP24A1	20	52789500	G	A	132	37.1	C005
C20orf107	20	55108524	C	T	137	29.2	C005
PMEPA1	20	56234754	C	T	42	35.7	C005
GNAS	20	57429319	C	T	45	26.7	C005
COL20A1	20	61943784	G	A	166	45.2	C001
EEF1A2	20	62124543	C	T	99	15.2	C001
CLDN8	21	31588146	G	A	53	32.1	C005
TIAM1	21	32513646	C	A	146	32.9	C005
KCNE1	21	35821714	G	A	91	29.7	C005
DSCAM	21	41711052	C	T	26	15.4	C003
PRDM15	21	43240501	C	T	107	32.7	C005
CRYAA	21	44589227	G	A	73	23.3	C005
AGPAT3	21	45379668	C	T	58	34.5	C005
TRAPPC10	21	45507649	C	T	98	30.6	C005
KRTAP10-10	21	46057961	C	T	66	10.6	C003
SUMO3	21	46233930	C	T	83	26.5	C005
C21orf58	21	47734704	G	del	18	33.3	C005
CECR1	22	17684488	C	T	51	39.2	C005
IGLL1	22	23915753	G	T	48	29.2	C005
CABIN1	22	24487647	G	T	15	20.0	C003

MYO18B	22	26173561	C	T	30	33.3	C003
TTC28	22	28559246	C	A	100	25.0	C005
SMTN	22	31485927	C	del	75	20.0	C005
TIMP3	22	33255274	C	T	92	28.3	C005
CARD10	22	37914066	G	T	70	30.0	C002
TRIOBP	22	38131311	C	T	59	30.5	C005
CSNK1E	22	38690184	A	C	33	27.3	C005
APOBEC3F	22	39441197	G	A	61	11.5	C004
CACNA1I	22	40042597	C	A	32	21.9	C001
CYP2D6	22	42523558	T	C	71	12.7	C002
PNPLA3	22	44322959	C	G	54	29.6	C001
ZBED4	22	50277892	G	A	35	40.0	C005
ZBED4	22	50280053	G	A	45	26.7	C005
PLXNB2	22	50720051	G	del	34	29.4	C005
LOC100288477	X	2161178	T	C	23	17.4	C001
LOC100288699	X	6452164	G	A	89	60.7	C005
VCX	X	7811259	G	A	189	32.8	C005
VCX3B	X	8434381	C	T	80	11.3	C001
KAL1	X	8503814	G	A	49	36.7	C005
KAL1	X	8522063	C	T	74	50.0	C003
WWC3	X	10092351	G	A	155	30.3	C005
MAGEB4	X	30260501	C	T	70	20.0	C005
MAGEB4	X	30260721	C	T	38	42.1	C003
FAM47A	X	34149162	G	A	72	31.9	C005
RP2	X	46696605	C	A	176	27.8	C005
PHF16	X	46917944	T	G	43	20.9	C005
USP11	X	47103933	C	T	70	34.3	C005
SSX6	X	47979001	G	A	96	27.1	C005
CACNA1F	X	49062144	G	A	120	36.7	C005
CACNA1F	X	49086983	C	T	87	19.5	C005
SHROOM4	X	50377343	C	T	79	24.1	C005
PHF8	X	54020149	G	A	139	28.1	C005
RRAGB	X	55744804	T	C	36	38.9	C005
MSN	X	64956699	G	A	38	15.8	C002
MSN	X	64956699	G	A	67	13.4	C005
OPHN1	X	67273563	G	A	103	33.0	C005
EDA	X	69255245	A	C	110	26.4	C005
EDA	X	69255381	C	T	92	25.0	C005

GPRASP1	X	101911831	A	G	167	35.3	C005
BEX4	X	102471205	G	T	12	33.3	C001
TMEM164	X	109247330	A	T	94	30.9	C005
CAPN6	X	110494792	C	T	51	31.4	C005
HTR2C	X	114141457	G	A	95	24.2	C005
RBMXL3	X	114425015	C	G	74	27.0	C005
RBMXL3	X	114425024	C	T	65	21.5	C005
RHOXF2B	X	119211032	C	T	54	22.2	C003
CT47A4	X	120104337	G	A	142	42.3	C005
ODZ1	X	123779155	C	T	77	35.1	C001
ACTRT1	X	127185708	C	T	76	40.8	C001
SLC25A14	X	129499587	G	A	82	26.8	C003
FLJ30058	X	130222739	G	A	154	21.4	C005
GPC4	X	132439928	G	del	97	27.8	C005
SLITRK4	X	142718004	T	G	16	37.5	C004
HSFX1	X	148857992	A	C	45	11.1	C005
GPR50	X	150349527	C	T	58	20.7	C002
CNGA2	X	150912763	C	T	250	36.0	C005
MAGEA10	X	151303486	G	T	20	40.0	C005
ZNF185	X	152083738	C	T	82	37.8	C005
RPL10	X	153628228	A	C	56	10.7	C005
PCDH11Y	Y	5605693	G	C	77	24.7	C004
LOC100288435	Y	58880230	A	T	58	22.4	C002

*Chr.: Chromosome

Supplementary Table 7. The signaling pathways and biological processes of the mutated genes in 5 gastric cancer tissues.

Term	Count	%	P Value	Genes	Fold Enrichment
cell adhesion (GO:0007155)	36	1.18	1.96E-7	CLDN8, AEBP1, NRP1, PCDH20, ASTN1, ITGA11, LMO7, DSCAML1, PCDHGC3, ROPN1B, KAL1, BAI1, ACAN, CSF3R, PCDHA11, PCDHB5, CNTNAP5, PCDH11Y, CNKSR3, COL15A1, HSPG2, NFASC, PCDHGB7, PTPRS, MFGE8, PTPRT, NID1, SIRPA, COL5A1, LAMA2, ITGA9, COL19A1, FREM1, CYFIP2, FCGBP, MUC16	2.67
steroid metabolic process (GO:0008202)	12	0.39	1.84E-3	GBA2, SOAT2, HSD3B7, HSD17B1, SORL1, MVK, ABCA2, IDI2, SULT1A4, WWOX, SLC27A5, VLDLR	3.08
endocytosis (GO:0006897)	12	0.39	3.57E-3	RAMP3, ABCA7, CAV1, P2RX7, ULK1, SORL1, HFE, OPHN1, MFGE8, PI4KB, VLDLR, EHD4	2.83
extracellular structure organization (GO:0043062)	9	0.29	1.35E-2	COL19A1, ADAMTS14, PCDHB5, POMT1, NFASC, ACAN, HSPG2, NID1, COL5A1	2.86
G1 DNA damage checkpoint (GO:0031571)	3	0.10	1.81E-2	PCBP4, TP53, FBXO31	14.1
skeletal system development (GO:0001501)	13	0.42	2.02E-2	WNT5A, AEBP1, CASR, TBX15, HSPG2, DSCAML1, NPR3, GJA5, P2RX7, EYA1, COL19A1, ACAN, WWOX	2.11
cell morphogenesis (GO:0000902)	14	0.46	2.52E-2	NRP1, SHROOM4, SPTBN4, DPYSL5, NFASC, DSCAML1, P2RX7, ULK1, LHX2, KAL1, BAI1, OPHN1, UBB, KLHL10	2.04
transmembrane transport (GO:0055085)	19	0.62	2.56E-2	TRPM3, HCN2, SLC45A4, CPT2, RRBP1, CACNA1I, SCN2A, EIF5A, AQP7, CNGB1, SLC22A10, SFXN5, SLC25A14, SEC61B, RYR1, KCNH7, KCNQ1, KCNH4, CACNA1B	1.73
response to extracellular stimulus (GO:0009991)	10	0.33	2.66E-2	BCKDHA, SOAT2, CAV1, P2RX7, ACSL1, AXL, TP53, HFE, PTCH1, VLDLR	2.36
muscle contraction (GO:0006936)	8	0.26	2.84E-2	HCN2, SMTN, MYH3, RYR1, MYH7, ADRBK1, GJA5, KCNQ1	2.71
vesicle-mediated transport (GO:0016192)	19	0.62	2.95E-2	RAMP3, ABCA7, FAM160A2, CAV1, SNAP91, SORL1, SPTBN4, HFE, MFGE8, PI4KB, COG4, P2RX7, COG7, ULK1, ROPN1B	1.71

Supplementary Table 8. List of somatic mutations in 5 non-cancerous gastritis mucosa detected by whole exome sequencing.

Gene	Chr.*	Chr.* Position	Reference Nucleotide	Altered nucleotide	Coverage	Mutant allele frequency (%)	Case
CPSF3L	1	1249723	A	G	20	15.0	C005
ATAD3B	1	1431048	A	G	51	13.7	C002
PRAMEF4	1	12943042	C	T	57	10.5	C003
PRAMEF4	1	12943171	T	C	48	12.5	C005
GRIK3	1	37285436	A	C	40	10.0	C005
DEM1	1	40980748	G	T	26	15.4	C004
C1orf173	1	75037860	T	C	20	20.0	C003
NBPF10	1	145301740	C	G	69	10.1	C005
GJA8	1	147380513	A	G	48	10.4	C005
HRNR	1	152187116	C	T	73	12.3	C001
HRNR	1	152187510	C	T	109	11.9	C002
OBSCN	1	228437711	T	G	52	15.4	C003
HIST3H2A	1	228645201	A	G	59	13.6	C005
LOC391343	2	905469	G	C	31	16.1	C005
LOC391343	2	905525	T	G	104	11.5	C001
MYT1L	2	1895892	T	G	23	21.7	C003
NAGK	2	71305511	C	T	68	14.7	C002
POTEF	2	130832519	C	G	29	17.2	C005
SP3	2	174820398	C	T	53	18.9	C003
IL8RB	2	219000514	T	C	94	10.6	C001
ABCB6	2	220080882	T	G	40	10.0	C003
SPEG	2	220333982	T	G	23	17.4	C003
ZNF860	3	32031992	C	G	45	17.8	C003
PFKFB4	3	48573717	T	C	48	10.4	C005
MST1R	3	49940214	G	T	31	16.1	C005
MUC4	3	195506354	C	G	55	18.2	C004
MUC4	3	195513831	A	G	25	16.0	C002
DRD5	4	9784802	G	A	47	10.6	C003
UGT2B11	4	70079940	C	G	125	12.8	C003
UGT2B11	4	70079941	C	G	127	11.0	C003
UGT2B4	4	70346458	G	A	152	11.8	C004
FAM160A1	4	152583777	G	T	40	10.0	C005
PCDHA5	5	140202795	C	T	61	13.1	C001

PCDHB7	5	140553676	C	T	81	12.4	C005
PCDHB12	5	140590609	G	A	11	27.3	C005
PCDHGB4	5	140767471	A	G	13	23.1	C005
NOTCH4	6	32187939	T	G	74	20.3	C002
HLA-DRB5	6	32497975	A	G	15	20.0	C003
HLA-DRB1	6	32548544	T	G	56	10.7	C001
TFEB	6	41658537	G	T	50	10.0	C003
C6orf132	6	42074648	A	G	20	15.0	C003
LRRC1	6	53660082	T	G	22	18.2	C005
FOXO3	6	108985092	C	G	78	14.1	C001
FOXK1	7	4800785	T	C	20	15.0	C003
DTX2	7	76112183	C	T	44	36.4	C002
AZGP1	7	99573585	G	A	110	12.7	C001
MUC12	7	100636257	A	G	58	10.3	C005
MUC17	7	100681513	A	G	69	11.6	C003
TRBV7-8	7	142099512	T	G	43	20.9	C001
TRBV7-7	7	142119838	C	A	55	12.7	C004
FAM90A18	8	7582817	C	A	20	20.0	C005
AMAC1L2	8	11189132	C	G	58	12.1	C003
AMAC1L2	8	11189184	C	T	79	12.7	C001
FAM160B2	8	21959744	A	C	47	10.6	C002
LOC100289448	8	86727678	G	T	40	15.0	C001
TOP1MT	8	144399984	G	T	21	19.1	C004
RANBP6	9	6013625	G	A	56	14.3	C003
FLJ46321	9	84607492	C	T	68	11.8	C003
EXD3	9	140245488	A	C	15	20.0	C003
PNPLA7	9	140389597	A	C	24	16.7	C003
FAM23A	10	17794347	C	T	14	21.4	C003
ANXA8L1	10	47161265	T	C	15	20.0	C004
FAM21A	10	51853101	C	A	58	10.3	C001
LZTS2	10	102763732	C	G	20	25.0	C002
DMBT1	10	124377777	T	C	70	10.0	C002
MUC2	11	1092966	C	A	79	10.1	C003
MUC2	11	1093136	C	T	71	15.5	C002
MUC5AC	11	1267452	C	G	46	13.0	C002
MRGPRX3	11	18158780	G	A	66	13.6	C002
PGA3	11	60975854	T	G	20	15.0	C004
RELT	11	73103289	T	G	18	22.2	C003

LOC642446	11	89608966	A	C	48	12.5	C005
MFRP	11	119213313	T	G	55	10.9	C005
CACNA1C	12	2788856	C	A	43	14.0	C003
PRB1	12	11506749	T	C	73	12.3	C001
PRB1	12	11506774	C	T	62	11.3	C001
MLL2	12	49444466	G	T	21	19.1	C004
TUBA1B	12	49522146	C	T	23	17.4	C005
R3HDM2	12	57648757	A	C	43	18.6	C005
AHNAK2	14	105411282	C	T	47	25.5	C002
AHNAK2	14	105413058	A	G	52	13.5	C003
AHNAK2	14	105416223	G	A	47	14.9	C003
AHNAK2	14	105417821	C	T	37	21.6	C003
GOLGA6	15	74368322	G	A	48	10.4	C003
MFGE8	15	89444938	A	G	33	15.2	C004
MEF2A	15	100252754	A	C	28	17.9	C001
LOC100132247	16	22545410	G	C	45	17.8	C002
ATXN2L	16	28844630	T	G	21	19.1	C003
THAP11	16	67876620	A	C	12	33.3	C003
COG4	16	70515032	C	T	28	57.1	C001
CHST5	16	75563446	T	G	20	15.0	C001
AMAC1L3	17	7385617	A	G	158	10.1	C001
TBC1D26	17	15641610	A	G	143	16.8	C001
AMAC1	17	33520478	G	A	203	11.8	C002
DHRS11	17	34956590	G	T	107	10.3	C002
TBC1D3D	17	36288674	G	A	274	10.2	C002
GPR179	17	36487356	G	T	16	25.0	C004
KRTAP1-5	17	39183231	A	G	78	12.8	C001
KRTAP4-9	17	39262015	T	C	29	17.2	C002
KRTAP9-9	17	39412093	C	A	50	12.0	C005
KRT15	17	39674999	T	C	42	11.9	C003
KRT17	17	39780537	A	C	31	16.1	C002
DHX58	17	40254239	C	A	41	14.6	C003
MEOX1	17	41738532	A	C	20	15.0	C003
GRN	17	42429478	A	G	40	10.0	C003
VEZF1	17	56060257	C	T	39	15.4	C001
VEZF1	17	56060266	A	C	46	13.0	C001
VEZF1	17	56060647	T	C	44	20.5	C002
ENPP7	17	77705039	C	G	28	17.9	C005

ARHGDIA	17	79826861	A	C	13	23.1	C003
FAM108A1	19	1881263	G	A	36	16.7	C002
TMPRSS9	19	2422101	A	C	11	27.3	C003
UHRF1	19	4910866	T	G	20	15.0	C004
VAV1	19	6822512	G	T	16	25.0	C003
CYP4F2	19	16003142	T	C	111	16.2	C001
RHPN2	19	33493767	G	C	45	13.3	C002
CYP2A6	19	41354271	A	C	57	14.0	C002
CYP2A13	19	41596322	T	G	109	11.0	C002
PSG4	19	43708365	G	C	84	11.9	C005
ZNF541	19	48047674	A	C	24	16.7	C005
SYT3	19	51135822	T	G	13	23.1	C005
LILRB3	19	54726833	T	C	60	20.0	C002
LOC284297	19	56029912	A	C	47	10.6	C003
ZNF264	19	57716795	G	A	73	11.0	C003
ASXL1	20	31024687	G	T	40	10.0	C005
DLGAP4	20	35064749	C	A	40	10.0	C002
CTSA	20	44520380	G	T	26	15.4	C001
LOC100132288	21	9909089	T	C	22	18.2	C003
RUNX1	21	36164790	G	T	46	10.9	C002
RIMBP3	22	20460303	C	G	15	26.7	C005
AIFM3	22	21331990	T	G	23	17.4	C003
ASCC2	22	30198175	A	C	28	17.9	C003
CYP2D6	22	42523558	T	C	96	13.5	C002
LOC100288699	X	6451898	A	C	27	18.5	C005
SUV39H1	X	48558770	G	T	47	12.8	C001
XAGE1B	X	52240541	T	G	44	11.4	C003
ITM2A	X	78622682	C	G	20	15.0	C003
RBMX2	X	129546608	A	G	43	11.6	C001
HSFX1	X	148858282	T	C	20	15.0	C005
OPN1MW	X	153455668	T	G	14	21.4	C005
IL9R	X	155227430	A	G	15	20.0	C004
LOC100288435	Y	58880230	A	T	15	20.0	C002

*Chr.: Chromosome

Supplementary Table 9. The signaling pathways and biological processes of the mutated genes in 5 non-cancerous gastritis tissues.

Term	Count	%	P Value	Genes	Fold Enrichment
transcription (GO:0006350)	10	2.87	5.94E-02	MYT1L, UHRF1, HSFX1, SP3, ASXL1, SUV39H1, VEZF1, RUNX1, THAP11, MLL2	1.89
regulation of Ras protein signal transduction (GO:0046578)	3	0.86	9.26E-02	OBSCN, TBC1D26, VAV1	5.68

Supplementary Table 10. Genes commonly mutated in both the gastric cancer and the matched background gastritis mucosa of the same patients.

Gene symbol	Name	Chr.*	Chr.* Position	Reference Nucleotide	Altered Nucleotide	Amino Acid Change**	Case
HRNR	hornerin	1	152187116	C	T	R2330Q	C001
C2orf90	chromosome 2 open reading frame 90	2	905469	G	C	S	C005
C2orf90	chromosome 2 open reading frame 90	2	905525	T	G	M141L	C001
PCDHA5	protocadherin alpha 5	5	140202795	C	T	R479W	C001
PCDHB12	protocadherin beta 12	5	140590609	G	A	S	C005
DTX2	deltex homolog 2 (Drosophila)	7	76112183	C	T	S	C002
TRBV7-8	T cell receptor beta variable 7-8	7	142099512	T	G	K97T	C001
FAM90A18	family with sequence similarity 90, member A18	8	7582817	C	A	L197I	C005
SLC35G5	solute carrier family 35, member G5	8	11189184	C	T	T190I	C001
MRGPRX3	MAS-related GPR, member X3	11	18158780	G	A	E11K	C002
PRB1	proline-rich protein BstNI subfamily 1	12	11506749	T	C	S	C001
PRB1	proline-rich protein BstNI subfamily 1	12	11506774	C	T	R88Q	C001
TUBA1B	tubulin, alpha 1b	12	49522146	C	T	S	C005
AHNAK2	AHNAK nucleoprotein 2	14	105411282	C	T	S	C002
MFGE8	milk fat globule-EGF factor 8 protein	15	89444938	A	G	S	C004
COG4	component of oligomeric golgi complex 4	16	70515032	C	T	D751N	C001
KRTAP4-9	keratin associated protein 4-9	17	39262015	T	C	V119A	C002
VEZF1	vascular endothelial zinc finger 1	17	56060647	T	C	S	C002
ENPP7	ectonucleotide pyrophosphatase/phosphodiesterase 7	17	77705039	C	G	D46E	C005
RHPN2	rhophilin, Rho GTPase binding protein 2	19	33493767	G	C	I300M	C002
CYP2A6	cytochrome P450, family 2, subfamily A, polypeptide 6	19	41354271	A	C	D169E	C002
LILRB3	leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 3	19	54726833	T	C	T6A	C002
TEKT4P2	tektin 4 pseudogene 2	21	9909089	T	C	S	C003
CYP2D6	cytochrome P450, family 2, subfamily D, polypeptide 6	22	42523558	T	C	Y355C	C002
DYZ1L18	DYZ1 repeat sequence	Y	58880230	A	T	S3T	C002

*Chr.: Chromosome

**S: Synonymous mutation

Supplementary Table 11. The coverages and mutant allele frequencies of the non-synonymous mutations that were commonly mutated in both the tumor and the matched background gastritis mucosa of the same patients.

Gene	Chr.*	Chr.* Position	Reference Nucleotide	Altered Nucleotide	Gastric cancer		Gastritis		Lymphocyte Coverage	Case
					Coverage	Mutant allele frequency (%)	Coverage	Mutant allele frequency (%)		
HRNR	1	152187116	C	T	92	14.1	73	12.3	33	C001
C2orf90	2	905525	T	G	98	11.2	104	11.5	48	C001
PCDHA5	5	140202795	C	T	66	13.6	61	13.1	39	C001
TRBV7-8	7	142099512	T	G	46	19.6	43	20.9	28	C001
FAM90A18	8	7582817	C	A	27	18.5	20	20.0	11	C005
SLC35G5	8	11189184	C	T	59	13.6	79	12.7	39	C001
MRGPRX3	11	18158780	G	A	56	21.4	66	13.6	39	C002
PRB1	12	11506774	C	T	131	12.2	62	11.3	33	C001
COG4	16	70515032	C	T	43	48.8	28	57.1	11	C001
KRTAP4-9	17	39262015	T	C	29	41.4	29	17.2	11	C002
ENPP7	17	77705039	C	G	54	20.4	28	17.9	25	C005
RHPN2	19	33493767	G	C	46	15.2	45	13.3	33	C002
CYP2A6	19	41354271	A	C	50	12.0	57	14.0	33	C002
LILRB3	19	54726833	T	C	78	12.8	60	20.0	21	C002
CYP2D6	22	42523558	T	C	71	12.7	96	13.5	48	C002
DYZ1L18	Y	58880230	A	T	58	22.4	15	20.0	69	C002

*Chr.: Chromosome

Supplementary Table 12. The accuracy of detecting the low-abundance mutations by deep sequencing.

Expected mutant allele frequency (%)	First run			Second run		
	Coverage	# of Mutant alleles	% of Mutant alleles	Coverage	# of Mutant alleles	% of Mutant alleles
1%	2381	37	1.55	9453	72	0.76
0.10%	4265	5	0.12	12264	13	0.11
0.01%	2628	0	0	12921	3	0.02

Supplementary Table 13. Clinicopathological features of individuals with or without gastric cancer for deep-sequencing.

Case*	Age	Sex	Gastric cancer					Gastritis (Endoscopic evaluation)		Gastritis (Histological evaluation)	
			Site	Tumor differentiation	Lauren Classification	Tumor stage (7th UICC)	MSI	Mucosal Atrophy	IM	Mucosal Atrophy	IM
C001	66	M	body	moderate	intestinal	IIB	MSS	marked	+	mild	mild
C002	79	M	body	well	intestinal	IV	MSS	marked	+	moderate	marked
C003	69	M	body	poor	intestinal	IIIC	MSS	marked	+	mild	normal
C004	72	M	pylorus	poor	intestinal	IV	MSS	marked	+	marked	marked
C005	81	F	antrum	moderate	intestinal	IB	MSI	marked	+	moderate	marked
C006	80	M	body, antrum (two lesions)	moderate	intestinal	IV	MSS	marked	+	mild	moderate
C007	79	M	antrum (two lesions)	poor	intestinal	IIB	MSS	marked	+	moderate	moderate
C008	59	M	antrum	moderate	intestinal	IA	MSS	marked	+	marked	moderate
C009	86	F	antrum	well	intestinal	IA	MSI	marked	+	marked	moderate
C010	76	M	antrum	poor	intestinal	IIIA	MSS	moderate	+	mild	moderate
C011	75	M	antrum (two lesions)	poor	intestinal	IB	MSS	marked	+	marked	moderate
C012	84	M	body	poor	intestinal	IV	MSS	marked	+	moderate	marked
C013	59	M	body	poor	intestinal	IV	MSS	marked	+	mild	mild
C014	59	F	body	poor	intestinal	IV	MSS	moderate	+	marked	normal
C015	77	M	body	poor	intestinal	IIIC	MSS	marked	+	moderate	marked
C016	85	M	antrum	poor	intestinal	IB	MSS	marked	+	mild	marked
C017	91	M	antrum	mucinous	intestinal	IV	MSS	marked	+	moderate	marked
C018	51	F	antrum~body	poor	diffuse	IV	MSS	moderate	-	moderate	normal
C019	73	F	antrum	poor	intestinal	IIB	MSS	moderate	+	mild	marked

C020	71	M	antrum	moderate	intestinal	IA	MSI	moderate	+	mild	marked
C021	81	M	body	moderate	intestinal	IIIB	MSS	marked	+	marked	marked
C022	62	M	antrum	poor	intestinal	IA	MSS	marked	+	mild	marked
C023	73	M	antrum	well	intestinal	IB	MSS	moderate	+	mild	marked
C024	79	F	antrum	poor	intestinal	IIIA	MSS	marked	+	mild	marked
C025	72	M	body	poor	intestinal	IIIC	MSS	marked	+	mild	marked
C026	76	M	antrum	well	intestinal	IA	MSS	marked	+	mild	marked
C027	72	M	antrum	well	intestinal	IA	MSS	marked	+	moderate	marked
C028	76	M	body	well	intestinal	IA	MSS	marked	+	marked	marked
A029	65	F	none					marked	+	N/A	N/A
A030	74	F	none					marked	+	N/A	N/A
A031	62	F	none					marked	+	N/A	N/A
A032	41	M	none					mild	-	N/A	N/A
A033	72	M	none					moderate	-	N/A	N/A
A034	59	F	none					moderate	-	N/A	N/A
N035	48	M	none					none	-	N/A	N/A
N036	71	F	none					none	-	N/A	N/A
N037	43	M	none					none	-	N/A	N/A
N038	62	M	none					none	-	N/A	N/A
N039	60	M	none					none	-	N/A	N/A

*C001-C028 were *H. pylori*-related gastritis patients with gastric cancer. A029-A034 were *H. pylori*-related gastritis patients without gastric cancer. N035-N039 were individuals without *H. pylori* infection and gastric cancer.
Abbreviations: UICC, Union for International Cancer Control; MSI, microsatellite instability; MSS, microsatellite stability; IM, intestinal metaplasia; N/A, not available

Supplementary Table 14. Validation of low-abundance mutations detected by deep sequencing.

Case	Gene	Mutation	First run		Second run	
			Coverage	Mutant allele frequency (%)	Coverage	Mutant allele frequency (%)
C002	*TP53*	c.365T>C p.V122A	3271	0.15	5249	0.10
	TP53	c.734G>A p.G245D	5590	0.16	6673	0.15
C011	*TP53*	c.734G>A p.G245D	3482	0.23	7237	0.10
C012	*TP53*	c.329G>T p.R110L	3887	0.18	6439	0.11
C017	*TP53*	c.734G>A p.G245D	4174	0.14	8868	0.11
C023	*TP53*	c.817C>T p.R273C	3834	0.26	6677	0.19
A024	*TP53*	c.734G>A p.G245D	7501	0.13	9475	0.15
	TP53	c.827C>T p.A276V	4262	0.12	9580	0.16
A025	*TP53*	c.677G>A p.G226D	3994	0.15	1816	0.17

Representative validated mutations were showed in this table.