



OPEN

Somatic alterations and mutational burden are potential predictive factors for metachronous development of early gastric cancer

Kazuhiro Sakuta¹, Yu Sasaki¹✉, Yasuhiko Abe², Hidenori Sato³, Masakuni Shoji¹, Takao Yaoita¹, Makoto Yagi², Naoko Mizumoto¹, Yusuke Onozato¹, Takashi Kon¹, Ayumi Koseki¹, Sonoko Sato¹, Ryoko Murakami³, Yuki Miyano³ & Yoshiyuki Ueno¹

The risk of developing metachronous gastric cancer (MGC) following curative endoscopic submucosal dissection (ESD) of early gastric cancer (EGC) remains even after eradicating *Helicobacter pylori* (HP) successfully. We screened initial EGC and adjacent non-cancerous mucosa ESD-resected specimens for somatic variants of 409 cancer-related genes, assessing their mutational burden (MB) to predict molecular markers for metachronous post-ESD development. We compared variants between ten patients diagnosed with MGC more than 3 years after ESD and ten age-matched patients who did not have MGC developments after successful HP eradication. We found no significant background differences between the two groups. In adjacent non-cancerous mucosa, the MB tended to be higher in the patients with metachronous developments than in the others. Somatic genomic alterations of *RECQL4*, *JAK3*, *ARID1A*, and *MAGI1* genes were significantly associated with MGC development. The criteria including both the MB and their variants, which had potential significant values for predicting MGC. In conclusion, combined of assessing specific somatic variants and MB may be useful for predicting MGC development. This study included a limited number of subjects; however, our novel findings may encourage further exploration of the significance of the molecular features of EGC that predict MGC development, thereby promoting focused follow-up strategies and helping elucidate the mechanisms.

Gastric cancer is the fifth most frequently diagnosed cancer and the third leading cause of death worldwide¹. Endoscopic submucosal dissection (ESD) is a well-established procedure, showing good prognoses after treatment of early stage neoplasms in the gastrointestinal tract, including early gastric cancer (EGC)^{2,3}. However, the risk of metachronous gastric cancers (MGCs) development following curative ESD of EGCs needs to be addressed properly⁴. The MGC incidence rate is approximately 2.7–15.6% after endoscopic resection⁵. Although successfully eradicating *Helicobacter pylori* (HP) has a significant preventive impact on MGC development^{6–8}, the risk remains present. Therefore, patients who undergo curative ESDs receive standard follow-up EGDs, at least annually, throughout their lives (even after successful HP eradication) to detect any developing MGCs at an early stage^{9,10}. A predictive marker for identifying patients with an increased risk of MGC is needed to establish a proper, efficient follow-up strategy.

Reports have shown that older people, men, current smokers, patients with multiple initial EGCs, or those with severe gastric mucosal atrophy have an increased risk of MGC⁵. Expression of CD44 variant 9 (CD44v9), a major adhesion molecule for the extracellular matrix, and a functional cancer stem cell marker in primary EGC lesions leads to a higher MGC rate than it would in CD44v9-negative tumors¹¹. In addition, the non-cancerous gastric mucosa's methylation level in gastric cancer (GC) patients, especially in the tumor-suppressor *miR-124a-3*, was significantly associated with an increased risk of developing MGC by a multicenter, prospective cohort study^{12,13}. These findings suggest that altering the normal gastric mucosa or initial EGCs may be markers for MGC occurrence.

¹Department of Gastroenterology, Faculty of Medicine, Yamagata University, 2-2-2 Iida-Nishi, Yamagata 990-9585, Japan. ²Division of Endoscopy, Yamagata University Hospital, 2-2-2 Iida-Nishi, Yamagata 990-9585, Japan. ³Genomic Information Analysis Unit, Department of Genomic Cohort Research, Yamagata University Faculty of Medicine, 2-2-2 Iida-Nishi, Yamagata 990-9585, Japan. ✉email: y-sasaki@med.id.yamagata-u.ac.jp

Like most solid epithelial cancers, GCs are often characterized by somatically-acquired mutations in various genes¹⁴. In fact, the Cancer Genome Atlas (TCGA) project provides a landmark in the molecular characterization of mainly advanced GC and identification of four subtypes¹⁵, and is considered a potential roadmap for patient stratification and trials of targeted therapies. However, studies on EGCs' genetic features, including those of the normal gastric mucosa, are scarce, and whether these genetic features may predict MGC development remains unclear. Therefore, we surveyed the somatic variants present in ESD-resected specimens obtained from patients with metachronous developments. We used next-generation sequencing (NGS) panel analyses on both the initial EGC and the non-cancerous mucosa specimens. Our results suggested that alterations of specific somatic variants and mutational burden may be linked to the risk of MGC development. The further analysis of these alterations, using a large number of patients, may help to estimate the risk of MGC and elucidate the mechanisms.

Results

Clinical characteristics of the patients. The study design was shown in Fig. 1. We found no significant differences between the non-MGC development (N) group and the metachronous development (M) group in terms of men/women ratios, alcohol consumption, current smoker rates, tumor locations, macroscopic types, gastric mucosal atrophy degrees, the presence of intestinal metaplasia, and age, body mass index, or maximum size of tumor means (Table 1, Suppl. Table 1). All initial EGCs were of the intestinal type and were confined to the mucosa.

Characteristics of somatic alterations. We identified 71,100 variants in the NGS panel (Suppl. Table 2 shows a complete list of all somatic variants), of which 3645 somatic variants were predicted to be protein altering (2103 missense, 1313 frame shift deletions, 162 nonsense, and 67 frame shift insertions; Fig. 2A,B). The variant load in our study stood almost in the middle of 33 landmark cohorts in the TCGA (Suppl. Fig. 1). The most common type of variant was a single nucleotide polymorphism (SNP; 58.0%, Fig. 2C), and most were C-to-T (30.2%) and G-to-A (29.1%) transitions (Fig. 2D). The median transition/transversion (Ti/Tv) ratio was 2.25 [interquartile range (IQR), 1.11–3.00, Suppl. Fig. 2]. We restricted the next analyses to those 3645 somatic variants.

Comparison of somatic variants between the N and the M groups considering both initial EGC and adjacent non-cancerous mucosa. The median number of somatic variants was 63.5 (IQR, 52.5–95.5) in all cases. We found no significant differences in the number of somatic variants between the N group and the M groups (Fig. 3A). In both the N and the M groups, the numbers of missense variants ($p=0.01$) and nonsense variants ($p=0.04$) were significantly higher in the adjacent non-cancerous mucosa specimens than in the initial EGC specimens, whereas the numbers of frame shift deletions ($p<0.01$) or insertions ($p<0.01$) were significantly higher in the initial EGC specimens than in the adjacent non-cancerous mucosa specimens (Fig. 3B). We found no significant differences between the N and the M groups in terms of the classifications or types of variants (Fig. 3B,C). Moreover, G-to-A and C-to-T transitions were more frequent in the adjacent non-cancerous mucosa than in the initial EGC in both groups, but not significantly so (Fig. 3D).

Mutational burden between the N and the M groups in initial EGC and adjacent non-cancerous mucosa specimens. In initial EGC specimens, the MBs were similar between the N and M groups, whereas in the adjacent non-cancerous mucosa, the MBs tended to be higher in the M group than in the N group (Fig. 4A,B). Based on our receiver operating characteristic (ROC) analysis (Fig. 4C), the area under the curve (AUC) value for MB in the adjacent non-cancerous mucosa (0.695) was higher than that of the initial EGC (0.425) for predicting metachronous developments. The calculated cutoff point was 7.75 (Suppl. Table 4).

Specific genes in the N and M groups. We compared the genes harboring somatic variants between the initial EGC and the adjacent non-cancerous mucosa specimens in the N and M groups (Fig. 5) with a false discovery rate <0.3 . In the N group (Fig. 5A), the initial EGCs had more variants in the *RecQ like helicase 4* (*RECQL4*), *Transcription factor 3* (*TCF3*), *Serine/threonine kinase 36* (*STK36*), *Insulin like growth factor 2 receptor* (*IGF2R*), and the *Adenomatous polyposis coli* (*APC*) genes, which were absent from the adjacent non-cancerous mucosa specimens, whereas the *Janus kinase 3* (*JAK3*) gene was highly altered in the adjacent non-cancerous mucosa. In the M group (Fig. 5B), the *AT-rich interactive domain containing protein 1A* (*ARID1A*) and the *membrane associated guanylate kinase, WW and PDZ domain containing 1* (*MAG11*) genes were highly altered in the initial EGCs, whereas the adjacent non-cancerous mucosa tended to have more variants in the *thrombospondin 1* (*THBS1*) and the *prostaglandin-endoperoxide synthase 2* (*PTGS2*) genes.

Predictive criteria for metachronous development. We sought to identify predictive criteria for metachronous development of gastric cancer by combining MB with specific gene variants (Fig. 6). The criteria (Table 2) for predicting non-MGC development included cases with initial EGC harboring *RECQL4* variants with *JAK3* variants or a low MB in the adjacent non-cancerous mucosa. In the N group, 70% of the patients met the criteria ($p=0.01$). The MGC prediction criteria were cases with initial EGC harboring variants in both the *ARID1A* and *MAG11* with high MBs in the adjacent non-cancerous mucosa. In the M group, 70% of the patients met the MGC prediction criteria, whereas none in N group fulfilled the MGC prediction criteria ($p<0.01$).

The frequencies of the specific gene variants in TCGA cohort. We analyzed these identified markers using the variant data sets of gastric adenocarcinomas ($n=440$) in the TCGA cohort¹⁵, although there were no

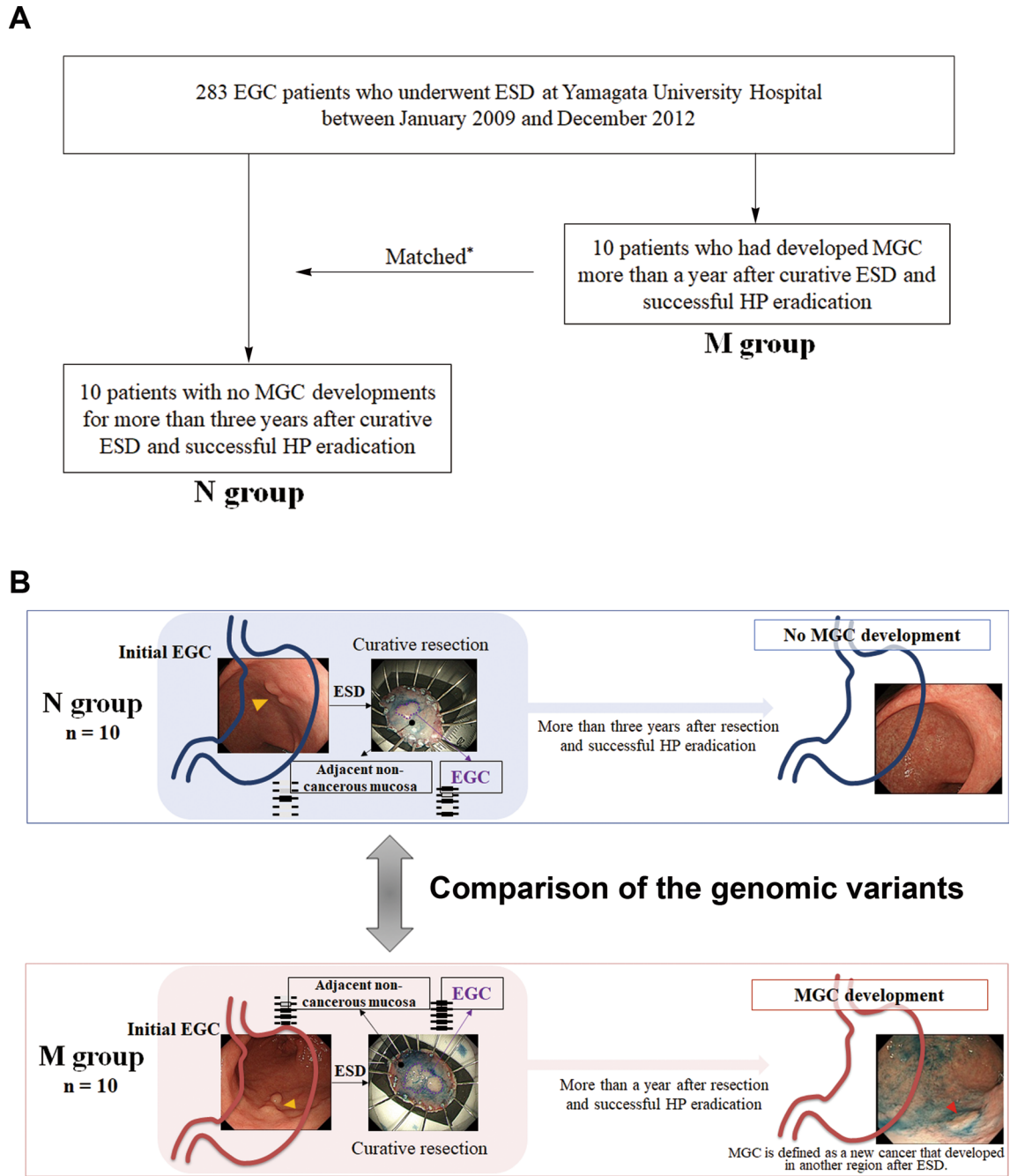


Figure 1. Study design. Among 283 EGC patients who underwent ESD at the Yamagata University Hospital between January 2009 and December 2012 (A), we selected ten patients who had developed MGC more than a year after curative ESD and successful HP eradication (the M group). We then selected ten patients with no MGC developments for more than 3 years after ESD and successful HP eradication (the N group). The somatic variants in initial EGC and adjacent non-cancerous mucosa were compared between the samples obtained from the M group and those obtained from the N group (B). We used the term “initial EGC” to differentiate the condition from MGC, which is also called “secondary cancer”. *The patients in the N group matched the ten patients in the M group in terms of age, gender, location, size, and histological type of the initial EGC. EGC early gastric cancer, ESD endoscopic submucosal dissection, MGC metachronous gastric cancer, HP *Helicobacter pylori*.

available variant data of non-cancerous mucosa with information on metachronous development. In the TCGA cohort, gastric cancer (GC) harboring variants in *RECQL4*, *ARID1A*, *MAG11*, or *JAK3* were 4.5% (n = 20), 27.0% (n = 119), 7.5% (n = 33), or 3.8% (n = 17), respectively. Of *RECQL4*-variants positive GCs, which is the criteria of the initial EGC for predicting non-MGC development in the present study, 70% of the GCs were negative for the variants in both *ARID1A* and *MAG11*. Of both *ARID1A*- and *MAG11*-variants positive GCs (n = 24, 5.4%), which is the criteria of initial EGC for predicting MGC development in the present study, 75% of the GCs were

	N group N = 10	M group N = 10	<i>p</i>
Age, years	66.0 ± 6.0	66.5 ± 6.7	Matched
Male	8 (80)	8 (80)	Matched
BMI, kg/m ²	23.8 ± 3.1	23.4 ± 3.7	0.90
Alcohol consumption rate	3 (30)	6 (60)	0.36
Current smoking rate	6 (60)	7 (70)	1.00
Gastric mucosal atrophy			Matched
Open type	10 (100)	10 (100)	
Closed type	0 (0)	0 (0)	
Intestinal metaplasia	10 (100)	10 (100)	1.00
Tumor location			Matched
Antrum	6 (60)	6 (60)	
Body	4 (40)	4 (40)	
Tumor size, mm	14.5 ± 6.6	17.7 ± 14.7	0.84
Macroscopic type			1.00
Depressed	7 (70)	7 (70)	
Elevated	3 (30)	3 (30)	
Pathological type			Matched
Intestinal	10 (100)	10 (100)	
Diffuse	0 (0)	0 (0)	
Depth of invasion			Matched
Mucosal lesion	10 (100)	10 (100)	
After successful eradication*, months	65.5 ± 17.6	38.5 ± 16.9	<0.01

Table 1. Clinical characteristics at the time of initial endoscopic treatment. Values are expressed as means ± standard deviations or as n (%). *We selected the patients in the M group > 1 year after they had undergone curative resection and successful HP eradication; we selected the patients in the N group > 3 years after they had undergone ESD and successful HP eradication.

negative for the variants in *RECQL4*. These criteria were seldom met in the same GCs in the TCGA cohort (1.3%, $p < 0.01$). These findings suggest that each criteria of the initial EGC in the present study seems to indicate a different genomic characteristic group of GC development.

Discussion

Molecular analyses from large-scale, international, collaborative studies have elucidated the genomic landscape of gastric cancer, highlighting its enormous heterogeneity and the complex interplay between genetic and epigenetic alterations^{14–16}. This has allowed identification of biomarkers with predictive, therapeutic, and prognostic values that may help physicians offer more personalized medicine than possible before. A recent study showed the genomic characteristics of both primary advanced GC and lymph nodes metastasis by using whole-exome sequencing¹⁷. However, EGC's molecular features remain unclear. In this present study, we evaluated somatic variants of 409 cancer-related genes and MBs in both EGC and adjacent non-cancerous mucosal specimens obtained from patients who later developed metachronous cancers and from those who remained in remission. We found that somatic genomic alterations of the *RECQL4*, *JAK3*, *ARID1A*, and *MAG11* genes, combined with MBs, were associated with the development of metachronous cancers after curative ESD and successful HP eradication. These findings provide new evidence of molecular EGC stratification to predict MGC and should help develop a specific follow-up strategy as well as elucidate the mechanisms.

HP has been accepted as the most important risk factor for gastric carcinogenesis^{18,19}, and its eradication should have prevented metachronous developments¹⁰; however, the risk of MGC cannot be removed completely. In this study, we revealed different genomic features of the initial EGCs and their adjacent non-cancerous mucosae in specimens from MGC patients (M group) and in those without MGC developments (N group), although both groups had similar clinical and endoscopic characteristics. These findings suggest that HP-infected individuals harbor different genomic alterations during early cancer stages. The timing of HP eradication affects not only the eradication rate, but also the MGC incidence rate^{10,20}. Our findings may be associated with a point-of-no-return in gastric carcinogenesis. Further studies are required to clarify the factors leading to these somatic alterations (such as host genomic factors, HP strains, and its virulence factors).

A high tumor mutational burden (TMB) is an emerging biomarker of sensitivity to immune checkpoint inhibitors²¹. Target capture sequencing may identify actionable genes and further determine tumor or mucosal MBs in a more cost-effective, efficient manner that will achieve higher sequencing depths than whole exome sequencing. A study using a hybridization, capture-based NGS panel, Biotecan PanCancer Panoramic Detection, revealed that the tumor mutational burden is associated with DNA damage repair genes, Lauren classification, differentiation, and HER2 status²². In that study, more than 80% of subjects had advanced GCs, and data

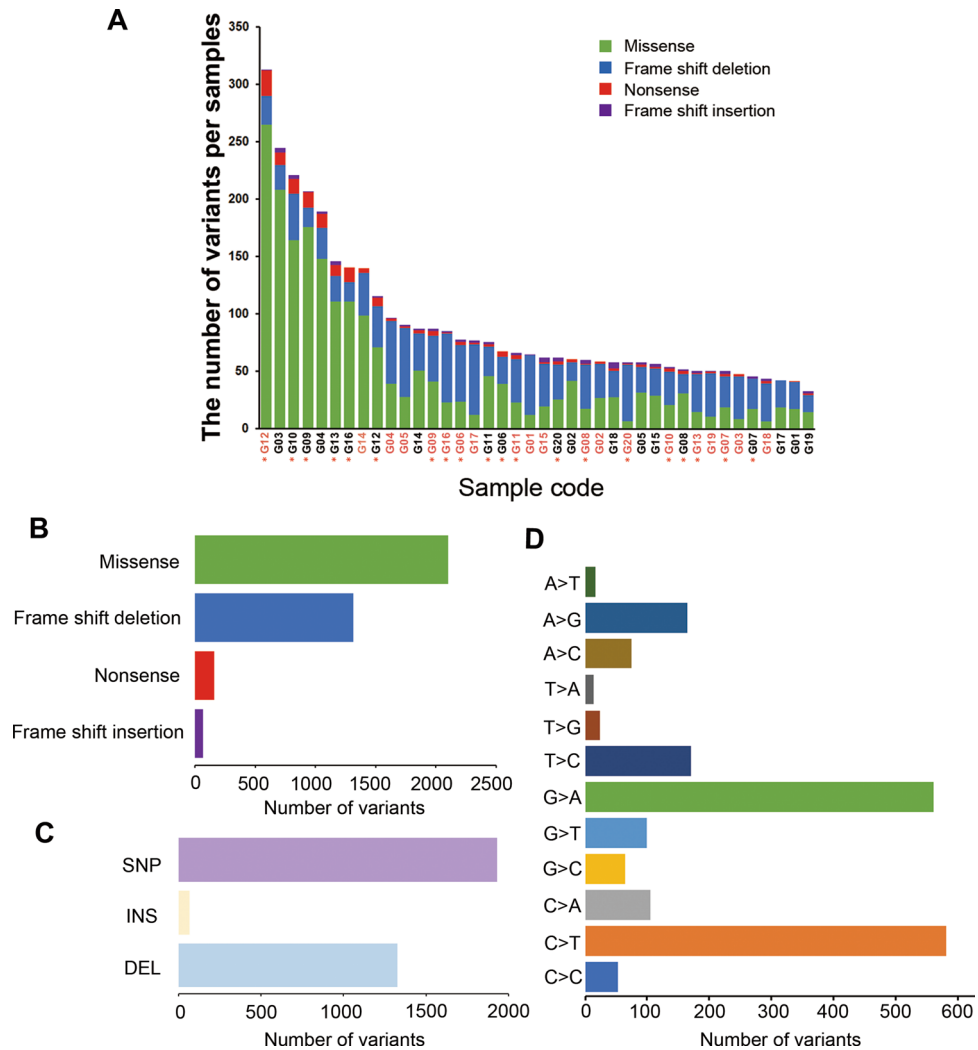


Figure 2. Summary of identified somatic variants. The total somatic variants in each sample, which are predicted to be protein altering, is displayed in a stacked bar plot (A: orange names belong to cancerous tissue, black ones belong to adjacent non-cancerous specimens; the asterisk indicates the metachronous development group). The median numbers of total somatic variants, missenses, frame shift deletions, nonsense, and frame shift insertions are 63.5 (IQR, 52.5–95.5), 28 (IQR, 18–61), 30 (IQR, 23.5–39), 2 (IQR, 1–4.5), and 2 (IQR, 1–2.5), respectively. Variant types in a boxplot summarized by variant classification (B) and type (C), and by SNP class (D). *SNP* single nucleotide polymorphism, *INS* insertion, *DEL* deletion.

of non-cancerous mucosa were not shown. In our study, we found high MBs in the adjacent non-cancerous mucosa in the specimens from the M group, despite the fact that the TMBs in the initial EGC specimens were similar in both groups. We speculate that a highly mutated gastric mucosa may reflect a cancerous background that may lead to development of MGCs. MB is a potential predictive marker for MGC and plays an important role in its pathogenesis.

In this study, somatic alterations of two cancer-related genes, including the *RECQL4* and the *JAK3*, had significant associations with non-MGC development, and the *ARID1A* and the *MAGI1* genes had significant associations with metachronous development. *RECQL4* belongs to the RecQ helicase family, a group of DNA unwinding enzymes that ensure proper repair of DNA damage to maintain genome stability, replication, recombination, and transcription²³. Mutational loss of the *RECQL4* is associated with increased risk for osteosarcoma development in Rothmund-Thomson syndrome²⁴. Studies have demonstrated that *RECQL4* acts as a tumor promoter in some cancers, such as prostate²⁵, breast²⁶, and colorectal cancers²⁷. Only two other studies have investigated associations between the *RECQL4* and GC. One reported that *RECQL4* drives cisplatin resistance by activating an AKT-YB1-MDR1 signaling pathway²⁸. The other study showed that increased *RECQL4* expression may predict poor prognosis in GC patients²⁹. In GC patients, *RECQL4* seems to be associated with treatment resistance, and EGC harboring the *RECQL4* variant was associated with non-MGC development. *JAK3* is a protein-tyrosine kinase of the JAK family composed of four members, including JAK1, JAK2, JAK3 and tyrosine kinase 2 (TYK2). JAK1, JAK2, and TYK2 are thought to be expressed ubiquitously, whereas JAK3 is confined to hematopoietic, myeloid, and lymphoid cells³⁰. *JAK3*, a potential tumorigenic-regulator, was shown to be upregulated directly

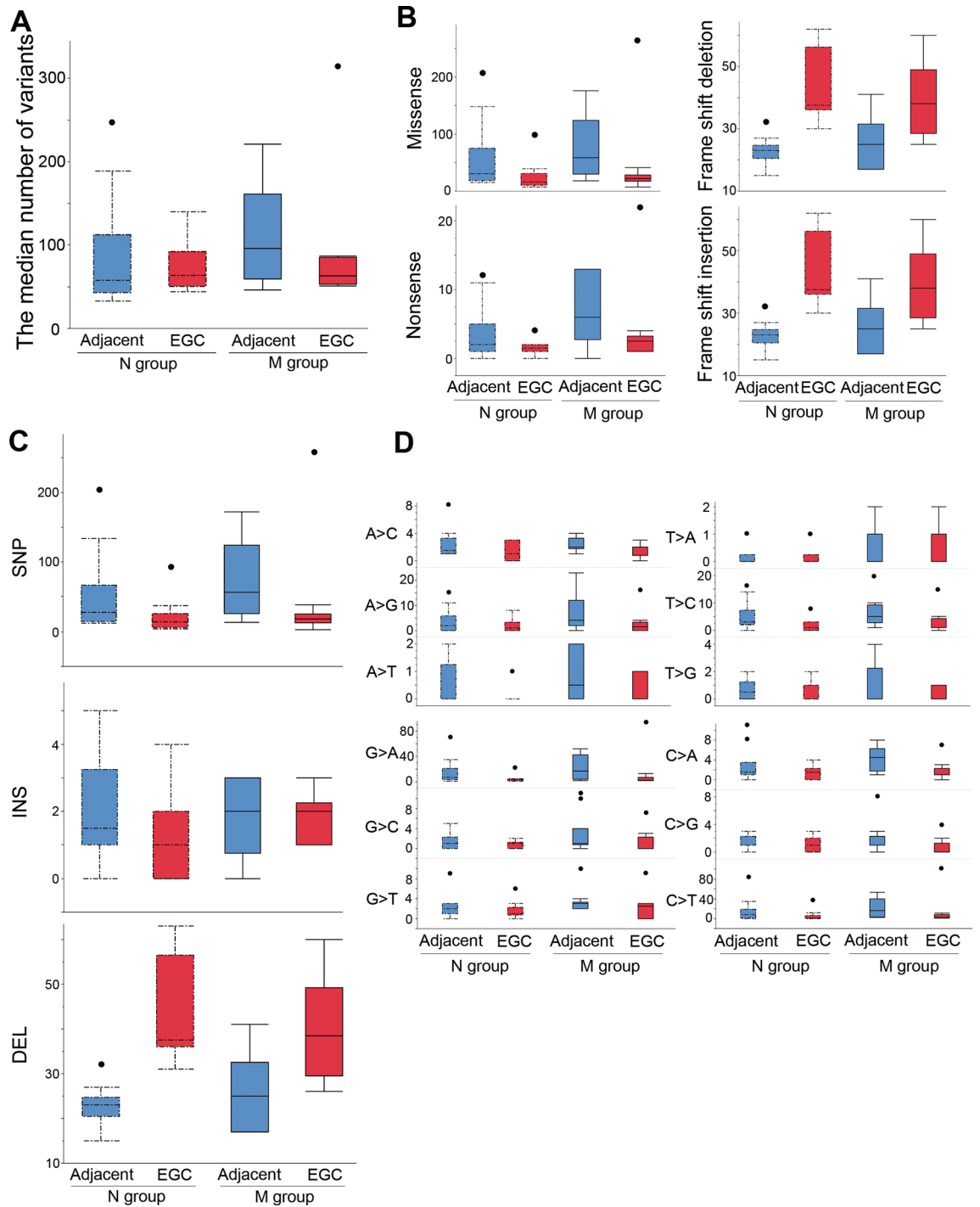


Figure 3. Comparison of somatic variants between the N and M groups. The median numbers of somatic variants were 58.0 (IQR, 42.7–112.5) in the adjacent non-cancerous mucosa of the N group, 63.5 (IQR, 50.2–92.5) in the initial EGC of the N group, 96.0 (IQR, 59.5–161.2) in the adjacent non-cancerous mucosa of the M group, and 63.0 (IQR, 53.2–85.5) in the initial EGC of the M group ($p=0.36$, **A**). The classification (**B**) or type (**C**) of variants and SNP classes (**D**) were not significantly different between the groups. Box plot: the bottom and top of each box represent the 25th and 75th percentiles, respectively, and the band in the box is the median. Whiskers: the lowest datum is within the minimum, and the highest datum is still within the 1.5 IQR of the upper quartile. Values of outliers are drawn as dots. We used the Kruskal–Wallis test, with post hoc Steel–Dwass test, to evaluate the statistical differences between groups. Actual values are provided in Suppl. Table 3. *SNP* single nucleotide polymorphism, *INS* insertion, *DEL* deletion.

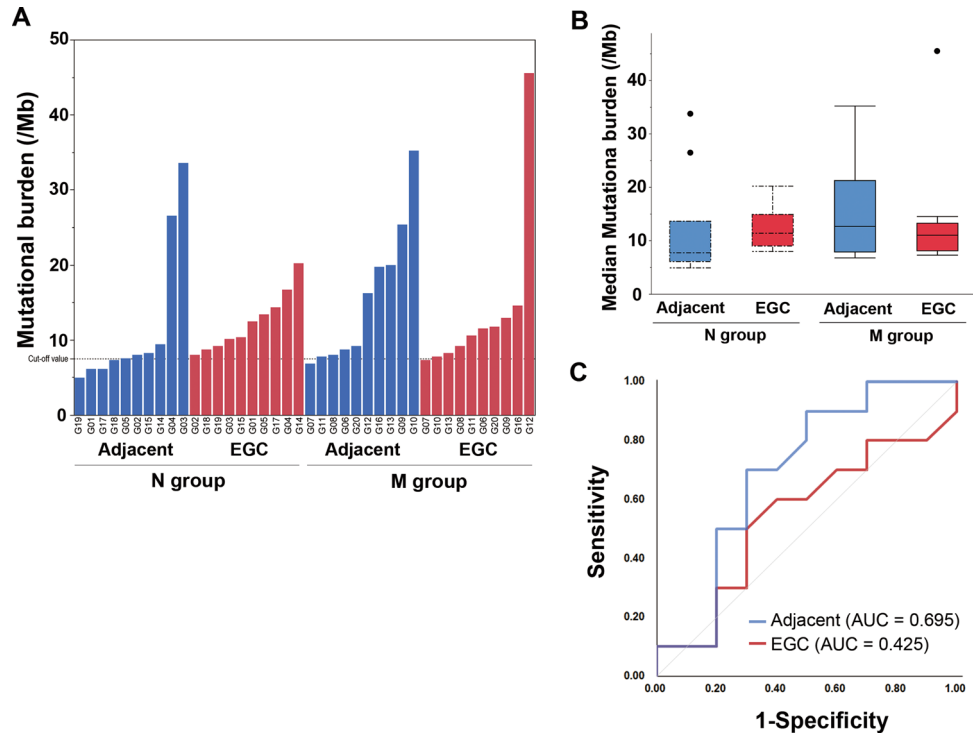


Figure 4. Mutational burden. MBs (A,B) tended to be higher in the M group [12.6 (7.9–21.3), in the adjacent non-cancerous mucosa, and 11.0 (8.1–13.3) in the initial EGC] than in the N group [7.7 (6.1–13.6) in the adjacent non-cancerous mucosa, 11.3 (9.0–14.9) in the initial EGC]. ROC analysis (C) showing that the AUC value of the MB in the adjacent, non-cancerous mucosa is higher than that in the initial EGC, and confirms 7.75 as the cutoff value for predicting metachronous developments. Box plot: the bottom and top of each box represent the 25th and 75th percentiles, respectively, and the band in the box is the median. Whiskers: the lowest datum is within the minimum, and the highest datum is still within the 1.5 IQR of the upper quartile. Values of outliers are drawn as dots. MB mutational burden, EGC early gastric cancer, ROC receiver operating characteristic, AUC area under the curve, IQR interquartile range.

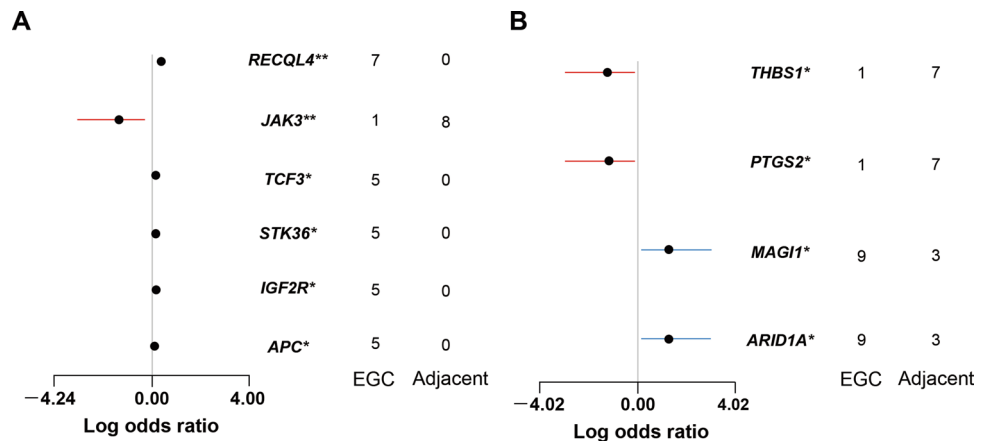


Figure 5. Specific genes harboring somatic variants associated with the initial EGC and the adjacent non-cancerous mucosa in the N group (A) or M group (B). We evaluated log odds ratios based on a false discovery rate <math>< 0.3</math>. * $p < 0.05$, ** $p < 0.01$. *RECQL4* RecQ like helicase 4, *TCF3* transcription factor 3, *STK36* serine/threonine kinase 36, *IGF2R* insulin like growth factor 2 receptor, *APC* adenomatous polyposis coli, *JAK3* Janus kinase 3, *ARID1A* AT-rich domain containing protein 1A, *MAG1* membrane associated guanylate kinase WW and PDZ domain containing 1, *THBS1* thrombospondin 1, *PTGS2* prostaglandin-endoperoxide synthase 2.

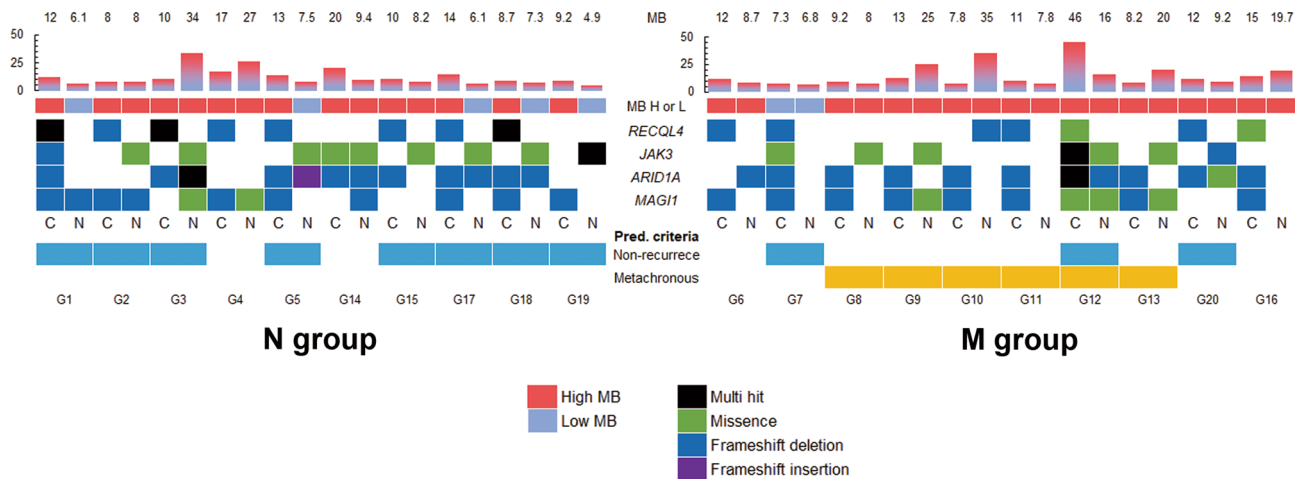


Figure 6. Prediction of metachronous development or non-development based on MB and specific gene variants. MB, mutational burden; H, high; L, low; *RECQL4*, *RecQ like helicase 4*; *JAK3*, *Janus kinase 3*; *ARID1A*, *AT-rich interactive domain containing protein 1A*; *MAGI1*, *membrane associated guanylate kinase WW and PDZ domain containing 1*; N, adjacent non-cancerous lesion; C, initial EGC; Pred. criteria, predictive criteria. The patients met the predictive criteria of non-development as blue whereas metachronous development as orange.

Prediction	Criteria			N group	M group	p
	Initial EGC		Adjacent non-cancerous mucosa			
Non development	<i>RECQL4</i> -variant positive	and	<i>JAK3</i> positive or low MB	7 (70)	1 (10)	0.019
Metachronous development	<i>ARID1A</i> - and <i>MAGI1</i> -variant positive	and	High MB	0 (0)	7 (70)	0.003

Table 2. Predictive criteria for MGC development or non-development. Values are expressed as n (%).

by activation of signal transducer and activator of transcription 3 (STAT3) following HP infection³¹. Further studies will need to clarify *RECQL4* and *JAK3* variants’ roles in GC as well as in infiltrating inflammatory cells.

The *ARID1A* gene encodes a key component of the adenosine triphosphate-dependent chromatin-modeling complex switch/sucrose-nonfermentable (SWI/SNF) chromatin-remodeling complex discovered to be a driver for ovarian clear cell carcinoma (as a tumor-suppressor)^{32,33}. Exome sequencing analyses in GC identified frequent inactivating mutations or protein deficiency of *ARID1A* in 83% of cases with microsatellite instability (MSI)³⁴ and an association with concurrent *PIK3CA* mutations and MSI³⁵. *ARID1A* inactivation leads to an altered mismatch repair, and increased mutagenesis, and its deficiency correlates with MSI³⁶. In our study, we revealed that an EGC harboring a variant of the *ARID1A*, with a high MB in the adjacent non-cancerous mucosa, was associated with an increased rate of developing MGC after HP eradication. Moreover, the methylation accumulation in non-cancerous gastric mucosa of GC patients has been suggested to be a promising biomarker for risk of MGC development^{12,13}. These findings suggest that disordered chromatin-remodeling in a genomic unstable gastric mucosa may lead to gastric carcinogenesis and may be associated with difficulties in reducing MGC risk after HP eradication. In addition to *ARID1A* alterations, the environment inducing *MAGI1* alterations during the EGC may be a MGC risk factor. *MAGI1* is a member of the *MAGI* subfamily of membrane associated guanylate kinases, which is emerging as important in coupling the extracellular environment to intracellular signaling pathways and the cytoskeleton at synapses and tight junctions³⁷. The interaction between HP CagA and the partitioning-defective 1 (PAR1) disrupts tight junctions and causes loss of epithelial cell polarity³⁸. Reports have demonstrated that *MAGI1* acts as a tumor suppressor by inhibiting the mitogen-activated protein kinase/extracellular signal-regulated kinase signaling pathway in GC³⁹. The *MAGI1* modulation, based on a disorganized tight junction, may underlie HP-infection induced gastric carcinogenesis.

We are aware of the limitations of our study. First, a selection bias may exist due to retrospective sample collection. Second, the limited sample size could affect the conclusion. Future studies with larger populations are required to verify our results. It is important to evaluate the quality of our findings in this discovery cohort using the new setting validation cohort. Third, further investigations need to determine the molecular mechanisms behind the association between somatic alterations or MB and development of MGC. Different gastric tumors possess multiple clonal origins⁴⁰. We did not compare the genomic characteristics of MGC and the initial EGC, which would have been interesting.

In conclusion, we presented that the combined assessment of specific somatic genomic alterations and the MB of the non-cancerous mucosa may differ in the M and N groups. These differences may help to identify the predictive biomarkers for developing MGC after curative ESD and successful HP eradication. This study included a limited number of subjects and needs to be validated in larger studies; however, our findings might provide new

insights into understanding the genomic landscape and mutational profile underlying EGC development, and have potential significant implications for improving future screening or treatment strategies for gastric cancer.

Methods

Participants. We retrospectively screened 283 EGC patients who underwent ESD at Yamagata University Hospital between January 2009 and December 2012. Among them, we selected 10 patients who had developed MGC more than a year after undergoing curative resection and successful HP eradication (M group). We then selected ten patients with no MGC developments for more than 3 years after ESD and successful HP eradication (N group), ensuring they matched the other 10 patients in M group in terms of age, gender, and location, size, and histological type of the initial EGC (Fig. 1A). We considered resections as curative when all the following pathological conditions were fulfilled: en bloc resection of the intestinal histological type, mucosal lesion, negative horizontal margin, negative vertical margin, and no lympho-vascular infiltration⁴¹. This assessment enabled us to distinguish between the part of cancer and adjacent non-cancerous tissue in the tissue specimen. We obtained clinical information on height, weight, alcohol consumption, smoking, and endoscopic findings of participants from their medical records.

The Ethics Committee of Yamagata University Faculty of Medicine approved this study (#2017-490), and we conducted it in accordance with the tenets of the Declaration of Helsinki. Written informed consent was obtained from all subjects.

Cancer-related gene panel sequencing. We extracted DNA separately, using the GeneRead DNA FFPE tissue kit (QIAGEN, Hilden, Germany), from the initial 10% formalin-fixed paraffin-embedded ESD-resected EGC and adjacent non-cancerous mucosa specimens. We used AmpliSeq Library Kit v2.0 (Thermo Fisher Scientific, Waltham, MA, USA) to construct libraries according to the manufacturer's instructions. Quantification of the libraries was performed on a 2200 TapeStation System using High Sensitivity D1000 Reagents and High Sensitivity D1000 ScreenTape (Agilent Technologies, Santa Clara, CA, USA). We submitted amplified libraries to emulsion PCR, using an Ion OneTouch 2 System with an Ion PI Template OT2 200 Kit v3 (Thermo Fisher Scientific, Waltham, MA, USA). Ion sphere particles were enriched, using Ion OneTouch ES, and were loaded onto an Ion PI Chip v2. NGS was performed by Ion Proton with the Ion Proton and Ion AmpliSeq Comprehensive Cancer Panel (Thermo Fisher Scientific, Waltham, MA, USA), which targeted 409 genes (Suppl. Table 5). Read sequence files were run through the assembly programs constructed with Bowtie2 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>) and BWA (<http://bio-bwa.sourceforge.net/bwa.shtml>), while consecutively carrying out a FreeBayes program (<https://github.com/ekg/freebayes>) to obtain variant call format files. Concordant genetic variants, detected by the two pipelines, were annotated using ANNOVAR (<http://annovar.openbioinformatics.org>).

Somatic variant analysis. After excluding synonymous mutations, we extracted the variants with a minor allele frequency not reported in the following genome databases: 1000 Genomes Project (<https://www.internationalgenome.org/home>), ExAC (<http://exac.broadinstitute.org/>), and our own healthy general population (N=176) cohort databases⁴². We defined the detected variants as somatic variants in this study. The somatic variants in the initial EGC and adjacent non-cancerous mucosa were compared between primary ESD-resected specimens obtained from the M and N groups (Fig. 1B).

Mutational burden assessments. The number of somatic variants detected on NGS, using the Ion AmpliSeq Comprehensive Cancer Panel (Thermo Fisher Scientific, Waltham, MA, USA), interrogated 1.29 Mb of the genome. We calculated the MB in variants per megabase (Mb) using maftools⁴³. Based on the ROC analysis to determine the MB cutoff point to predict MGC developments, we divided the MB levels into one of two groups: low (less than 7.75 variants/Mb) and high (≥ 7.75 variants/Mb).

Statistical analyses. We analyzed continuous variables and categorical variables using the two-tailed Wilcoxon test and the Fisher exact test, respectively. We used the Kruskal–Wallis test to compare continuous variables following a post-hoc Steel–Dwass test. We computed the ORs and 95% confidence intervals using logistic regression model analyses. We considered differences with $P < 0.05$ to be statistically significant. We carried out statistical calculations using JMP 14 (SAS Institute) and R programming language version 3.6.1 software.

Data availability

All data generated or analyzed during this study are included in this article. All other data generated and/or analyzed to support our findings are available from the corresponding author upon reasonable request.

Received: 24 October 2019; Accepted: 7 December 2020

Published online: 16 December 2020

References

1. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**, 394–424 (2018).
2. Hasuike, N. *et al.* A non-randomized confirmatory trial of an expanded indication for endoscopic submucosal dissection for intestinal-type gastric cancer (cT1a): The Japan Clinical Oncology Group study (JCOG0607). *Gastric Cancer* **21**, 114–123 (2018).
3. Ono, H. *et al.* Guidelines for endoscopic submucosal dissection and endoscopic mucosal resection for early gastric cancer. *Dig. Endosc.* **28**, 3–15 (2016).

4. Kato, M. Metachronous gastric cancer risk after endoscopic resection of early gastric cancer and *H. pylori* status. *J. Gastroenterol.* **51**, 19–20 (2019).
5. Abe, S. *et al.* Metachronous gastric cancer following curative endoscopic resection of early gastric cancer. *Clin. Endosc.* **51**, 253–259 (2018).
6. Fukase, K. *et al.* Effect of eradication of *Helicobacter pylori* on incidence of metachronous gastric carcinoma after endoscopic resection of early gastric cancer: An open-label, randomised controlled trial. *Lancet* **372**, 392–397 (2008).
7. Choi, J. M. *et al.* Effects of *Helicobacter pylori* eradication for metachronous gastric cancer prevention: A randomized controlled trial. *Gastrointest. Endosc.* **88**, 475–485.e2 (2018).
8. Choi, I. J. *et al.* *Helicobacter pylori* therapy for the prevention of metachronous gastric cancer. *N. Engl. J. Med.* **378**, 1085–1095 (2018).
9. Mori, G. *et al.* Incidence of and risk factors for metachronous gastric cancer after endoscopic resection and successful *Helicobacter pylori* eradication: results of a large-scale, multicenter cohort study in Japan. *Gastric Cancer* **19**, 911–918 (2016).
10. Xiao, S., Li, S., Zhou, L., Jiang, W. & Liu, J. *Helicobacter pylori* status and risks of metachronous recurrence after endoscopic resection of early gastric cancer: A systematic review and meta-analysis. *J. Gastroenterol.* **54**, 226–237 (2019).
11. Hirata, K. *et al.* CD44 variant 9 expression in primary early gastric cancer as a predictive marker for recurrence. *Br. J. Cancer* **109**, 379–386 (2013).
12. Asada, K. *et al.* Demonstration of the usefulness of epigenetic cancer risk prediction by a multicentre prospective cohort study. *Gut* **64**, 388–396 (2015).
13. Maeda, M. *et al.* High impact of methylation accumulation on metachronous gastric cancer: 5-year follow-up of a multicentre prospective cohort study. *Gut* **66**, 1721–1723 (2017).
14. Tan, P. & Yeoh, K.-G. Genetics and molecular pathogenesis of gastric adenocarcinoma. *Gastroenterology* **149**, 1153–1162.e3 (2015).
15. Cancer, T. *et al.* Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202–209 (2014).
16. Jácome, A. A., Coutinho, A. K., Lima, E. M., Andrade, A. C. & Dos Santos, J. S. Personalized medicine in gastric cancer: Where are we and where are we going?. *World J. Gastroenterol.* **22**, 1160–1171 (2016).
17. Zhang, J. *et al.* Genomic alterations in gastric cancers discovered via whole-exome sequencing. *BMC Cancer* **18**, 1–12 (2018).
18. Amieva, M. & Peek, R. M. Pathobiology of *Helicobacter pylori*-induced gastric cancer. *Gastroenterology* **150**, 64–78 (2016).
19. Uemura, N. *et al.* *Helicobacter pylori* infection and the development of gastric cancer. *N. Engl. J. Med.* **345**, 784–789 (2001).
20. Sokic-Milutinovic, A., Alempijevic, T. & Milosavljevic, T. Role of *Helicobacter pylori* infection in gastric carcinogenesis: Current knowledge and future directions. *World J. Gastroenterol.* **21**, 11654–11672 (2015).
21. Chalmers, Z. R. *et al.* Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med.* **9**, 1–14 (2017).
22. Cai, H. *et al.* Mutational landscape of gastric cancer and clinical application of genomic profiling based on target next-generation sequencing. *J. Transl. Med.* **17**, 1–12 (2019).
23. Bernstein, K. A., Gangloff, S. & Rothstein, R. The RecQ DNA helicases in DNA repair. *Annu. Rev. Genet.* **44**, 393–417 (2010).
24. Wang, L. L. *et al.* Association between osteosarcoma and deleterious mutations in the RECQL4 gene in Rothmund–Thomson syndrome. *J. Natl. Cancer Inst.* **95**, 669–674 (2003).
25. Su, Y. *et al.* Human RecQL4 helicase plays critical roles in prostate carcinogenesis. *Cancer Res.* **70**, 9207–9217 (2010).
26. Fang, H. *et al.* RecQL4 helicase amplification is involved in human breast tumorigenesis. *PLoS ONE* **8**, 1–9 (2013).
27. Lao, V. V. *et al.* Altered RECQ helicase expression in sporadic primary colorectal cancers. *Transl. Oncol.* **6**, 458–469 (2013).
28. Mo, D. *et al.* Human helicase RECQL4 drives cisplatin resistance in gastric cancer by activating an AKT-YB1-MDR1 signaling pathway. *Cancer Res.* **76**, 3057–3066 (2016).
29. Chen, H. *et al.* Overexpression of RECQL4 is associated with poor prognosis in patients with gastric cancer. *Oncol. Lett.* **16**, 5419–5425 (2018).
30. Roskoski, R. Janus kinase (JAK) inhibitors in the treatment of inflammatory and neoplastic diseases. *Pharmacol. Res.* **111**, 784–803 (2016).
31. Zhao, J. *et al.* *Helicobacter pylori*-induced STAT3 activation and signalling network in gastric cancer. *Oncoscience* **1**, 468–475 (2014).
32. Jones, S. *et al.* Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science* **330**, 228–231 (2010).
33. Wiegand, K. C. *et al.* ARID1A mutations in endometriosis-associated ovarian carcinomas. *N. Engl. J. Med.* **363**, 1532–1543 (2010).
34. Wang, K. *et al.* Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nat. Genet.* **43**, 1219–1223 (2011).
35. Zang, Z. J. *et al.* Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. *Nat. Genet.* **44**, 570–574 (2012).
36. Shen, J. *et al.* ARID1A deficiency promotes mutability and potentiates therapeutic antitumor immunity unleashed by immune checkpoint blockade. *Nat. Med.* **24**, 556–562 (2018).
37. Feng, X., Jia, S., Martin, T. A. & Jiang, W. G. Regulation and involvement in cancer and pathological conditions of MAGI1, a tight junction protein. *Anticancer Res.* **34**, 3251–3256 (2014).
38. Saadat, I. *et al.* *Helicobacter pylori* CagA targets PARI/MARK kinase to disrupt epithelial cell polarity. *Nature* **447**, 330–333 (2007).
39. Jia, S. *et al.* MAGI1 inhibits migration and invasion via blocking MAPK/ERK signaling pathway in gastric cancer. *Chin. J. Cancer Res.* **29**, 25–35 (2017).
40. Takaoka, S. *et al.* Molecular subtype switching in early-stage gastric cancers with multiple occurrences. *J. Gastroenterol.* **54**, 674–686 (2019).
41. Sano, T. & Kodera, Y. Japanese gastric cancer treatment guidelines 2014 (ver. 4). *Gastric Cancer* **20**, 1–19 (2017).
42. Koyama, S. *et al.* Clinical and radiological diversity in genetically confirmed primary familial brain calcification. *Sci. Rep.* **7**, 1–9 (2017).
43. Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C. & Koeffler, H. P. Maftools: Efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* **28**, 1747–1756 (2018).

Acknowledgements

This study was supported in part by the JSPS KAKENHI Grant-in-Aid for Young Scientists (B) (Grant numbers JP17K15922, JP17K15920, and JP19K17389) and by AMED (Grant number JP18ck0106277).

Author contributions

K.S. conducted experiments acquired data, analyzed results, and drafted the manuscript. Y.S. analyzed and interpreted results, obtained funding, wrote, reviewed, and edited the original draft, and supervised the study. Y.A. interpreted results, obtained funding, and supervised the study. H.S. conducted experiments, acquired data, analyzed and interpreted results, and revised the manuscript. M.S., T.Y., N.M., Y.O., and T.K. acquired data. M.Y. acquired data, and obtained funding. A.K. and S.S. conducted experiments. R.M. and Y.M. conducted

experiments, acquired data, and analyzed results. Y.U. conceptualized and supervised the study, obtained funding, and revised the manuscript. All authors agreed with the final version of the manuscript for submission.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-020-79195-0>.

Correspondence and requests for materials should be addressed to Y.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020