Expansion of gastric intestinal metaplasia with copy number aberrations contributes to field cancerization

Ken Kumagai[1], Takahiro Shimizu[1], Atsushi Takai[1], Nobuyuki Kakiuchi[2], Yasuhide Takeuchi[2,3,4], Tomonori Hirano[2], Haruhiko Takeda[1], Aya Mizuguchi[1], Mari Teramura[1], Takahiko Ito[1], Eriko Iguchi[1], Mitsuhiro Nikaido[1], Yuji Eso[1], Ken Takahashi[1], Yoshihide Ueda[1,5], Shin'ichi Miyamoto[1,6], Kazutaka Obama[7], Seishi Ogawa[2], Hiroyuki Marusawa[1,8], Hiroshi Seno[1]

[1]Department of Gastroenterology and Hepatology, Graduate School of Medicine, Kyoto University, Kyoto, Japan
[2]Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan
[3]Department of Diagnostic Pathology, Kyoto University Hospital, Kyoto, Japan
[4]Clinical Bio Resource Center, Kyoto University Hospital, Kyoto, Japan
[5]Department of Gastroenterology and Hepatology, Graduate School of Medicine, Kobe University, Hyogo, Japan
[6]Department of Gastroenterology and Hepatology, National Hospital Organization Kyoto Medical Center, Kyoto, Japan
[7]Department of Surgery, Graduate School of Medicine, Kyoto University, Kyoto, Japan
[8]Department of Gastroenterology and Hepatology, Osaka Red Cross Hospital, Osaka, Japan

Running title: Genetic analysis of gastric IM

Key words: gastric cancer; atrophic gastritis; *Helicobacter pylori*; single gland isolation; whole-exome sequencing

Corresponding author:
Takahiro Shimizu

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; shimy@kuhp.kyoto-u.ac.jp

Conflicts of interest:

The authors declare no potential conflicts of interest.

Abstract

Intestinal metaplasia (IM) is a risk factor for gastric cancer following infection with *Helicobacter pylori*. To explore the susceptibility of pure gastric IM to cancer development, we investigated genetic alterations in single IM gastric glands. We isolated 50 single IM or non-IM glands from the inflamed gastric mucosa of 11 patients with intramucosal gastric carcinoma (IGC) and 4 patients without IGC; nineteen single glands in the non-inflamed gastric mucosa of 11 individuals from our cohort and previous dataset were also included as controls. Whole exome sequencing of single glands revealed significantly higher accumulation of somatic mutations in various genes within IM glands compared with non-IM glands. Clonal ordering analysis showed that IM glands expanded to form clusters with shared mutations. Additionally, targeted-capture deep sequencing and copy number (CN) analyses were performed in 96 clustered IM or non-IM gastric glands from 26 patients with IGC. CN analyses were also performed on 41 IGC samples and the Cancer Genome Atlas-Stomach Adenocarcinoma datasets. These analyses revealed that polyclonally expanded IM commonly acquired copy number aberrations (CNA), including amplification of chromosomes 8, 20, and 2. A large portion of clustered IM glands typically consisted of common CNAs rather than other cancer-related mutations. Moreover, the CNA patterns of clustered IM glands were similar to those of IGC, indicative of precancerous conditions. Taken together, these findings suggest that, in the gastric mucosa inflamed with *H. pylori* infection, IM glands expand via acquisition of CNAs comparable to those of IGC, contributing to field cancerization.

Statement of Significance

This study contributes to our understanding of gastric intestinal metaplasia as a risk factor for gastric adenocarcinoma via their multifocal expansion and acquisition of copy number aberrations and somatic mutations.

Introduction

Gastric cancer is one of the leading causes of cancer-associated mortality worldwide(1). Gastric adenocarcinoma is classified into two histological subtypes: intestinal-type and diffuse-type(2). The more prevalent form of intestinal-type gastric adenocarcinoma is thought to develop as a stepwise progression of histologic stages initiated by chronic gastritis, leading to atrophic gastritis, intestinal metaplasia (IM), and dysplasia(3). *Helicobacter pylori* infection is an important inducer of this progression(4). Indeed, 0.21–0.49% of patients with IM are at risk of developing gastric cancer each year(5). Even after eradication therapy for *H. pylori*, IM remains a risk factor for gastric cancer development(6). However, some reports have shown that metaplasia is simply a lesion that forms as an adaptation to environmental stimuli, such as chronic inflammation(7,8). Thus, whether IM represents a precursor lesion, or simply a confounding factor, for gastric cancer remains unknown.

Cancer cells are generated from normal epithelial cells through stepwise accumulation of genetic alterations(9–11). Recent studies have demonstrated that genetic alterations spontaneously accumulate in normal cells in association with ageing(12). During inflammation-associated carcinogenesis, the generation of genetic alterations can be accelerated by various factors, including reactive oxygen species and the activity of endogenous DNA mutator enzymes, including activation-induced cytidine deaminase (13–16). Indeed, we have previously reported that somatic mutations accumulate in

various genes within gastric tissues inflamed with *H. pylori* infection(17). Further, another group reported the genomic and epigenomic profiles of gastric mucosa with IM and stratified individuals as high-risk for development of gastric cancer according to the genomic and epigenomic alterations in their gastric mucosa(18). These data suggest that the inflamed gastric mucosa with IM contains more genetic alterations than uninflamed mucosa. However, genetic analyses using mass tissues, such as biopsy samples, could not accurately clarify the true genetic aberrations in non-cancerous gastric glands. Considering that cancer cells are derived from epithelial cells, investigation of genetic alterations in pure gastric epithelial cells is important for determining the mechanism of gastric carcinogenesis. Gutierrez-Gonzalez et al. used laser capture microdissection and Sanger sequencing to demonstrate that multiple cells in metaplastic glands share a common mitochondrial DNA mutation, and metaplastic and dysplastic glands can be genetically related, suggesting that gastric cancer can arise from metaplasia(19,20). However, there has been no comprehensive genetic analysis of gastric glands consisting solely of epithelial cells, including IM. Therefore, in the present study, we clarified the genetic alterations in their entirety in gastric single IM glands by single gland isolation from fresh samples. Furthermore, to clarify how IM contributes to field cancerization, we investigated genetic aberrations in clustered lesions consisting purely of IM and compared them with those in early-stage gastric cancer.

Materials and Methods

Study population

In total, we enrolled 67 patients who were admitted to Kyoto University hospital from 2007 through 2021 (Fig. 1a, Supplementary Fig. S1). Among them, 41 patients were

treated with endoscopic submucosal dissection for differentiated intramucosal gastric carcinoma (IGC), classified using the Vienna classification(21), and had inflamed gastric mucosa due to current, or previous, *H. pylori* infection. Seven patients without IGC, including four with *H. pylori*-related inflamed mucosa and three with non-inflamed mucosa were treated for other diseases, such as gastric submucosal tumors. 19 IGC patients enrolled in a previous study(22) were included in the current study based on an opt-out method of consent.

For whole-exome sequencing (WES) of gastric single glands, 36 single glands in the inflamed mucosa from 11 patients with IGC, 14 single glands in the inflamed mucosa from 4 patients without IGC, and 6 single glands in the non-inflamed mucosa from 3 patients with gastric submucosal specimens were obtained as described below. For WES of IGC, cancer components were obtained from formalin-fixed paraffin-embedded (FFPE) tissue specimens collected from the same 11 patients used for WES of gastric single glands. For targeted-capture deep sequencing and copy number (CN) analysis, 72 clustered IM glands and 24 clustered non-IM glands in the inflamed mucosa from 26 patients with IGC were obtained as described below; of these, 2 patients were included in the above-mentioned WES analysis. We also re-analyzed previously reported CN microarray data for 41 IGC samples from 19 patients. For transcriptome analysis, five clustered IM glands and three clustered non-IM glands in the inflamed mucosa from six patients with IGC were obtained as described below. The characteristics of the participants are summarized in Supplementary Table S1.

All protocols were approved by the ethics committee of Kyoto University (G0463, G1084, G1315) and complied with all provisions of the Declaration of Helsinki. Clinical samples were obtained with written informed consent or based on an opt-out method of

consent from all participants.

Isolation of gastric single IM glands and clustered IM glands

We endoscopically and microscopically confirmed the non-cancerous regions in specimens endoscopically resected for IGC and obtained non-cancerous tissue samples using a punch biopsy with a 2 mm diameter (BP-20F, Kai Industries, Tokyo, Japan) (Fig. 1b, c). We confirmed that the accompanying cancerous tissues were not included in the collected samples based on histological analysis of the remaining tissue. Gastric mucosa samples from patients without IGC were obtained using the surgically resected specimens or biopsy specimens. After incubating the samples in 20 mM EDTA phosphate-buffered saline, the gastric epithelium was manually dissociated from non-epithelial components as described previously(23). We then stained the samples with Alcian blue dissolved in 3% acetic acid at 4 ℃ for 20 min and distinguished IM glands from non-IM glands based on the presence or absence of goblet cells (Fig. 1b, d). For WES of gastric single glands, we isolated “single IM glands” and “single non-IM glands” from inflamed gastric mucosa, as well as “single normal glands” from non-inflamed gastric mucosa. To analyze the expansion of gastric glands, we also isolated clustered lesions consisting of pure IM glands or pure non-IM glands, referred to as “clustered IM glands” or “clustered non-IM glands,” respectively.

Whole-genome amplification (WGA) and WES of gastric single glands

For WES of gastric single glands, genomic DNA (gDNA) was extracted from each single gland and split into two aliquots; WGA was performed for each aliquot independently using the REPLI-g Single Cell Kit (Qiagen, Hilden, Germany) as described

previously(23) (Supplementary Fig. S2). The details of WES are described in Supplementary Methods.

Targeted-capture deep sequencing of gastric clustered glands

gDNA was extracted from gastric-clustered glands using the QIAamp DNA Micro Kit (Qiagen) or phenol-chloroform methods. Libraries were constructed using xGen custom-designed Gene Capture Pools (Integrated DNA Technologies, Coralville, IA, USA) to cover the 98 relevant targeted genes (Supplementary Table S2). We also used the xGen CNV Backbone Panel as a spike-in panel to analyze the CN of gastric clustered glands(24). The details of gene selection, sample preparation, and sequencing are described in the Supplementary Methods.

CN analysis

Sequenced reads of single glands and clustered glands were re-aligned to the human reference genome GRCh37 for CN analysis. We calculated the log2 ratio of each sample using CNACS as described previously(23). We obtained the copy number aberration (CNA) score for CN gain and loss separately using the web application tool CNApp (25). The score was categorized as follows: broad copy number score (BCS), which represents the chromosome or arm level aberrations, and focal copy number score (FCS), which represents the focal CNAs. We also used Copy Number Explorer(26) to calculate the segment gain or loss (SGOL) score representing the total number of samples crossing the thresholds of segmented log2 ratio. The default cutoff value of the log2 ratio was <-0.2 and >0.2. High absolute SGOL scores indicate a high frequency and/or magnitude of alterations across samples(27,28). To evaluate similarity in the CNAs of non-IM glands,

IM glands, and IGC, we re-analyzed previously reported CN microarray data from 41 IGC samples using the OncoScan CNV FFPE assay kit (Affymetrix, Santa Clara, CA, USA)(22). We obtained the CNA scores and SGOL scores based on the log2 ratio in a range longer than 0.34 Mb, as the xGen Backbone Panel is designed across every 0.34 Mb of the genome. We also calculated the CNA scores and SGOL scores of cases from The Cancer Genome Atlas (TCGA)-Stomach Adenocarcinoma (STAD) and plotted the SGOL scores using Copy Number Explorer.

External data

For genetic analysis of early-stage gastric cancer and non-inflamed gastric mucosa, we re-analyzed the gastric adenocarcinoma dataset in TCGA-STAD (https://portal.gdc.cancer.gov/projects/TCGA-STAD) (Fig. 1a, Supplementary Table S3, S4) and a previous study (accession number JGAS000313)(29)(Fig. 1a, Supplementary Table S1), respectively. Sample selection is detailed in the Supplementary Methods.

Data availability statement

Sequence data have been deposited in the Japanese Genotype–phenotype Archive (https://trace.ddbj.nig.ac.jp/jga), which is hosted by the DDBJ, under accession number JGAS000269.

Results

Genetic analysis of gastric single IM glands

To investigate the glands that are genetically susceptible to cancer development, we first sought to examine the genetic alterations in single IM glands or non-IM glands in

gastric mucosa. We collected gastric mucosa from endoscopic submucosal resection samples, identified IM glands or non-IM glands by Alcian blue staining, and isolated single IM glands or non-IM glands (Fig. 1b-d). After WGA of single-gland DNA, we conducted WES to detect genetic alterations (Supplementary Fig. S2). The variant allele frequency (VAF) distribution of mutations in single glands showed a peak of approximately 0.5, and genetic alterations with VAF ≥ 0.25 were matched at 97.7% in two independent analyses of the same samples (Supplementary Fig. S3a-c). To investigate the precise genetic alterations existing clonally in single glands and to minimize WGA and sequencing errors, we applied VAF ≥ 0.25 as a cutoff value for detecting genetic alterations by WES of single glands, which was the same parameter used for colon crypt analysis in a previous study(23).

Gastric single IM glands highly accumulate mutations and CNAs

We isolated 50 single IM glands or non-IM glands in inflamed gastric mucosa from 11 patients with IGC and 4 patients without IGC, and performed WES after WGA (Fig. 1a, Supplementary Fig. S1). As controls, we performed WES on 6 single normal glands in non-inflamed gastric mucosa and re-analyzed WES data of 13 single normal glands from 8 individuals, which were performed by the same methods as in the present study(29). The mean depth of targeted regions in single glands was 88.3× (54.5×–142.1×), while that in leukocytes was 89.5× (62.6×–141.3×) (Supplementary Table S5). The mean number of mutations in single normal glands was 17.8, whereas those in single non-IM glands and single IM glands were 29.8 and 80.8, respectively. Simple linear regression analysis showed that single IM glands acquired 1.105 mutations per year (R^2 = 0.934), which was 2.7- and 3.4-fold greater than those in single non-IM glands (0.407 mutations

per year, R^2 =0.825) and single normal glands (0.325 mutations per year, R^2 = 0.894), respectively (Fig. 2a, b, Supplementary Table S6, S7). The same trends were observed following analysis of only synonymous mutations (Supplementary Fig. S4a, Supplementary Table S6, S7). There was no difference in number of mutations of single IM glands or non-IM glands between patients with IGC and without IGC (Supplementary Fig. S4b) and no relationship between histological atrophic degrees and genetic alterations (Supplementary Fig. S4c). Somatic mutations in single normal glands and single non-IM glands were within the range predicted to be acquired spontaneously in normal cells of various tissues(12), whereas single IM glands had more mutations than those predicted in normal cells. We also investigated CNAs in single IM glands and non-IM glands by analyzing WES data with the CNACS pipeline(23,30). WGA samples of single normal glands were used as controls to reduce the artifacts of WGA and sequencing. Seven of the 29 single IM glands showed CNAs, including amplification of chromosomes 2, 7, 8, 13, and 20. On the other hand, 4 of 21 single non-IM glands showed CNAs, including amplification of chromosomes 13 and 20 (Fig. 2c, Supplementary Fig. S5 and Supplementary Table S8).

To elucidate the reason gastric IM glands are susceptible to genetic alterations, we first analyzed the mutation signatures of gastric single glands, which represent the total mutagenesis from their birth to the present time. C to T substitutions at the CpG site, or signature 1, were clearly dominant in single IM glands (Fig. 3a, b, and Supplementary Fig. S6). This result suggests that IM glands had an increased cell turnover till that time. Furthermore, we performed transcriptome analysis of eight gastric clustered IM or non-IM glands. Interestingly, the expression of gene sets related to DNA repair, G2/M checkpoint, and mitotic spindle was reduced in IM glands compared to that in non-IM

glands (Supplementary Fig. S7 and Supplementary Table S9, S10). The reduction in the functions of DNA repair, G2/M checkpoint and mitotic spindle might be associated with the increase in somatic mutations and CNAs. Although these expression profiles were biased by the timing of sampling, they might be the possible cause of the increased rate of genetic alterations in gastric IM glands.

To compare somatic mutations in gastric single glands with those in gastric cancer, we performed WES of IGC using FFPE tissues from the same 11 cases used for gastric single-gland analysis. The mean targeted depth was 97.9× (66.8×–123.4×) (Supplementary Table S5), and averages of 89.65 (30–175) and 1591 mutations (1425–1777) were found in IGC with microsatellite stability (MSS) and instability (MSI), respectively (Supplementary Table S7, S11). Furthermore, we re-analyzed the WES data of submucosal gastric carcinoma at stage IA and gastric adenocarcinoma at stage IB, excluding MSI and hypermutated-single-nucleotide variant predominant (HM-SNV) cancer, using TCGA database. Averages of 81.4 (47–109) and 203.4 (38–671) mutations were observed in submucosal gastric carcinoma at stage IA and gastric adenocarcinoma at stage IB, respectively (Fig. 2b). Taken together, these results suggest that single IM glands in patients with gastric cancer had similar numbers of somatic mutations as those in intramucosal and submucosal gastric carcinoma, although statistical analysis was not performed due to the differences in research protocols used.

Several cancer-related mutations exist in gastric single IM glands

Among the mutated genes identified in gastric single glands, 34 genes were recurrently mutated as non-synonymous mutations in ≥ 3 cases (Fig. 4a and Supplementary Fig. S8). Mutations in *MUC6*, *TTN*, *MUC19*, *RYR2*, *THSD7B*, *MUC16*,

SYNE1, and *ADGRL3* were observed frequently in single IM glands or single non-IM glands. Twenty-five of these 34 genes were likely susceptible to mutations due to their relatively large size, longer than 50,000 bases. Meanwhile, no mutated genes or pathways were specifically enriched in single IM glands or single non-IM glands.

The detected mutations were classified as established cancer-related genes (Tier1), suspected cancer-related genes (Tier2), or unregistered genes based on the Catalogue of Somatic Mutations in Cancer (COSMIC) cancer gene census(31). We defined both Tier1 and Tier2 as cancer-related genes in the present study. Limited to cancer-related genes, single IM glands showed 3.4 non-synonymous mutations on average, including those in *KMT2C*, *KMT2D*, *FAT4*, and *ATM*, which were reported to be frequently observed in gastric cancer, whereas single non-IM glands and normal glands showed an average of 1.4 and 0.9 non-synonymous mutations, respectively (Fig. 4b and Supplementary Fig. S9). Mutations in *KMT2C*, *KMT2D*, *FAT3*, *FAT4*, *ATM*, *LRP1B*, *ATRX*, *ARID2*, *GNAS*, *TNC,* and *IRF4* were identified in the IM glands of two or more cases. Among them, several mutations were predicted as functionally damaged (Supplementary Table S6), but no mutations were detected in the driver genes crucial for gastric cancer development, such as *TP53*, *ARID1A*, *APC*, *CDH1*, and *RHOA*. It was predicted that single glands with incomplete IM had more mutations than those with complete IM (Supplementary Fig. S10a,b), whereas no relationship was found between the number of cancer-related mutations and IM subtypes (Supplementary Fig. S10c,d).

In contrast, an average of 6.0 and 79.7 non-synonymous mutations in cancer-related genes were identified in IGC with MSS and MSI, respectively. *TP53*, *ARID1A*, and *APC* mutations were observed in 4, 3, and 2 of the 11 IGC cases, respectively (Supplementary Table S11). Furthermore, an average of 5.2 (0–8) and 10.1 (3–30) cancer-related

mutations were observed in submucosal gastric carcinoma at stage IA and gastric adenocarcinoma at stage IB of the TCGA dataset, respectively. These findings suggest that single IM glands have similar number of mutations, but fewer number of cancer-related mutations compared to IGC.

Gastric IM glands expand to form aggregates

In inflamed gastric mucosa, aggregates of IM are often observed by endoscopic and histologic examination. To investigate how gastric glands expand in inflamed gastric mucosa, we first performed clonal ordering analyses using the genomic data of multiple single glands from each 2 mm punch biopsy (Fig. 5 and Supplementary Fig. S11). Among the pairs of single IM glands in the 10 IM aggregates, five pairs located less than 0.5 mm from each other showed common genetic alterations that accounted for over half of those detected in each single gland. Four of the five pairs of single IM glands located more than 0.5 mm from each other showed relatively fewer common genetic alterations (Supplementary Fig. S12). Pairs of single non-IM glands in punched-out lesions also showed common genetic alterations in four of six cases; however, their proportion was < 50%. These findings suggest that gastric glands expand during chronic inflammation and that IM glands from multiple origins expand clonally, resulting in the formation of IM aggregates.

Next, to clarify the genetic alterations involved in the expansion of gastric glands, we investigated genetic alterations in clustered lesions consisting purely of IM glands or non-IM glands by deep sequencing. We obtained 72 clustered IM glands and 24 clustered non-IM glands from the inflamed gastric mucosa of 26 patients with IGC (Fig. 1a, Supplementary Fig. S1). Approximately 150–300 glands were included per clustered

gland. The exons of the total 98 genes in 95 clustered glands, with the exception of one clustered IM gland, were captured and deep-sequenced. The details of targeted genes are described in the Supplementary Methods and Supplementary Table S2. Further, we applied a VAF cutoff value ≥ 0.02 to detect mutations prevailing in more than 4% of the clustered glands. The mean targeted depth was 585.4× (282.4–989.3×) (Supplementary Table S12), and 176 mutations in 25 genes were detected in 95 clustered glands (Fig. 6a). We then randomly selected 48 mutations that were successfully validated via PCR-based deep sequencing analysis.

The number of somatic mutations in targeted genes was similar between clustered IM glands and clustered non-IM glands (1.92 vs 1.67 on average, respectively, $p = 0.60$). The most frequently mutated genes were *MUC6*, *CSMD1*, *CCDC168,* and *RYR2*, none of which have been reported as cancer-related genes classified as Tier1 or Tier2 (Fig. 6a, Supplementary Table S13). Among them, *CSMD1* was frequently mutated in clustered IM glands ($p < 0.05$) according to Fisher's exact test; however, the dN/dS ratio of *CSMD1* was approximately 1 (Supplementary Table S14), indicating the presence of passenger mutations. On the other hand, *MUC6* was frequently mutated in clustered non-IM glands ($p < 0.05$), and the dN/dS ratio of *MUC6* was high (Supplementary Table S14), suggesting an involvement in the positive selection. For cancer-related genes, clustered IM glands showed more mutations compared to clustered non-IM glands; however, the difference was not significant (0.27 vs 0.13 per sample, respectively, $p = 0.39$) and most of these mutations were predicted as functionally non-damaged (Supplementary Table S13). No mutations were detected in crucial driver genes, including *TP53*, *ARID1A*, *APC*, *CDH1*, and *RHOA*. The VAF in 136 of 176 mutations (77.3%) was < 0.1 (Fig. 6b). Five of the 71 clustered IM glands had mutations at VAF > 0.3, whereas none of the 24 clustered non-

IM glands had mutations at VAF > 0.3. There was no shared mutation between clustered IM or non-IM glands and IGC, although these clustered glands were located far from IGC (Supplementary Table S15). These results suggest that gastric glands expanded, to some extent, in inflamed gastric mucosa, and that IM glands were more highly expanded than non-IM glands, forming aggregates.

IM expands with accumulation of CNAs before acquiring driver mutations

Targeted-capture deep sequencing analysis of clustered glands revealed that somatic mutations did not significantly contribute to the expansion of inflamed gastric glands. We, therefore, performed autosomal CN analyses for 96 clustered glands of inflamed gastric mucosa. Interestingly, CNAs were identified in 33 of the 72 clustered IM glands, whereas only 2 glands among the 24 clustered non-IM glands had CNAs (Supplementary Fig. S13 and Supplementary Table S16). There was no significant relationship between IM subtypes and CNAs (Supplementary Fig. S14). The CNA score revealed that clustered IM glands had significantly broader CN gain compared to clustered non-IM glands (BCS 1.62 vs 0.17, $p = 0.0121$; Fig. 7a), whereas no difference was observed between clustered IM and non-IM glands in terms of CN loss (BCS 0.11 vs 0, $p = 0.61$. FCS 0.51 vs 0, $p = 0.757$) or focal CN gain (FCS 0.56 vs 0, $p = 0.535$; Fig. 7a). To compare the CNAs of IM or non-IM glands with those of gastric cancer, we calculated the autosomal CNA score for 41 IGC from our previous study(22), as well as 10 submucosal gastric carcinoma samples at stage IA and 34 gastric adenocarcinoma samples at stage IB from TCGA database. Broad CNAs showed similar trends between IM and IGC and were significantly increased in submucosal gastric carcinoma as well as in stages that were more progressive. In contrast, focal CNAs were increased in IGC compared to those in IM glands (Fig. 7a).

Next, to evaluate the chromosomal changes that may contribute to expansion of the gastric glands, we calculated and plotted SGOL scores of clustered IM glands and clustered non-IM glands (Fig. 7b). CNAs frequently observed in clustered IM glands showed broad gains in chromosomes 2, 8, 13, and 20; whereas broad gains in chromosomes 13q and 20 were also detected in some clustered non-IM glands. Chromosome 8 contains *MYC*, chromosome 2 contains *MYCN* and *ERBB4*, and chromosome 20 contains *GNAS*, *SALL4*, and *SRC*, suggesting that their overexpression is involved in gastric gland proliferation. Among these genes, MYC and SRC proteins were highly expressed in IM glands compared to non-IM glands (Supplementary Fig. S15). To evaluate the similarity of the CNA distribution between IM and gastric cancer, we calculated and plotted the SGOL scores for gastric cancer (Fig. 7b). The distribution of CN gain in IM glands was similar to that in IGC and more progressive stages. CN loss was frequently observed in stage IA and IB gastric carcinoma, whereas it was rarely observed in gastric IM glands and non-IM glands. Furthermore, fluorescence in situ hybridization (FISH) analyses confirmed that amplification of chromosome 8 and *MYC* were shared in IGC and their adjacent IM glands (Supplementary Fig. S16). These results suggest that broad CN gains, including gain of chromosomes 8 and 20, were acquired by IM before malignant transformation, and that focal CNAs and broad CN loss emerged from the early stage of gastric cancer.

To investigate whether CNAs are related to clonal expansion of IM glands, we evaluated the proportion of each CNA in clustered IM glands. Considering the result of single-gland CN analyses, we assumed that the CN became 1 or 3 under the condition that whole cells acquired CN gain or CN loss, respectively. Interestingly, 42/86 (49%) CNAs in clustered IM glands had spread to more than 50% of the glands (Fig. 7c,

Supplementary Table S16). Particularly, 19/36 (53%) CNAs with gain of chromosome 8 or 20 spread to more than 50% of the glands in IM. However, estimation of mutation VAFs in cancer-related genes in clustered IM glands revealed no mutated cancer-related genes with functional damage and VAF > 0.25 (Fig. 6b). These findings suggest that CNAs, rather than somatic mutations, could be involved in expansion of gastric IM glands.

Discussion

Gastric mucosa with IM is thought to represent a high-risk factor for gastric cancer. However, whether gastric IM serves as a precursor lesion or simply a confounding factor for gastric cancer is unclear. In this study, we clarified the landscape of genetic aberrations in pure IM glands and demonstrated that IM is susceptible to gastric cancer development. Currently, two methods are used to investigate the genomic data of pure gastric epithelial cells, the organoid culture method and analysis of single glands or single-cells(23,32,33). The former provides data on single stem cell-derived populations amplified by culture; in this method, selection bias due to only successfully cultured cells being analyzed is inevitable. The latter provides data on pure gastric epithelial cells; however, it involves WGA, which may increase contamination by artefacts. Previous studies have successfully minimized the artefacts of WGA using a specific platform(23,30), which was also applied in the present study. We also established original methods for isolating gastric single IM glands consisting of a single origin. Using these approaches, we obtained a complete view of genetic aberrations in gastric single IM glands and clustered IM glands and determined how IM may contribute to field cancerization in the stomach.

In the current study, we demonstrated that single IM glands have similar numbers of somatic mutations as those in intramucosal and submucosal gastric carcinoma, which are

2.7- and 3.4-fold more than those in single non-IM and normal glands, respectively. However, IM glands lacked the crucial mutations for cancer development that contribute to gland expansion. Previously, we observed low-abundance *TP53* mutations in inflamed gastric mucosa(17); Huang et al. also identified specific driver mutations, including those in *TP53* and *FBXW7*, in gastric mucosa with IM(18). The present results suggest that driver mutations do not contribute to the expansion of most IM, although rare gastric glands or a small proportion of cells in gastric mucosa may possess these driver mutations. These results differ from those stating that *TP53* mutations are frequently observed in vast regions of other organs with inflammation-associated carcinogenesis, such as ulcerative colitis and Barrett's esophagus(34–37).

CNAs were frequently observed in clustered IM glands, suggesting that CNAs contribute to the expansion of IM glands. Unexpectedly, the CNA pattern in IM glands was similar to that in IGC. Focal CNAs and CN loss were not frequently observed in IM glands and increased with adenocarcinoma progression from early to advanced stages. Among the CNAs commonly detected in IM glands and gastric cancer in the early stage, gain of chromosomes 8 and 20 was frequently observed. These chromosomes contain the oncogenes *MYC*, *KAT6*, and *FGFR1* and *SRC*, *SALL4*, and *GNAS*, respectively, indicating their advantages to form a precancerous field, as proposed with *KRAS* mutations in the intestine(38,39). Gain of chromosome 2 was also frequently observed in IM glands, whereas this CNA was not common in gastric cancer. All clustered IM glands with gain of chromosome 2 also showed gain of chromosome 20 in this study, suggesting that chromosome 2 gain is a passenger alteration, or might have tumor suppressor functions. We showed that MYC and SRC were highly expressed in IM as representative molecules whose genes are located at these amplified chromosomes. However, further

comprehensive analyses, such as those of transcriptome or proteome may help us find essential pathways associated with positive selection and cancer development. These findings suggest that IM glands expand with the acquisition of CNAs and form a precancerous field multi-focally, possibly followed by malignant transformation through acquisition of additional genetic aberrations, such as cancer driver mutations and CN loss (Fig. 8).

Whether gastric IM is the origin of gastric cancer has been argued. Several studies showed a possible genetic link of gastric IM and cancer tissues(40,41), and Gutierrez-Gonzalez et al. demonstrated clonal relationship between IM glands and cancerous glands by laser capture microdissection of gastric glands adjacent to the gastric cancer using fixed samples and Sanger sequencing (19). In this study, we did not directly show their clonal relationship by comprehensive genetic analyses, because we could not obtain IM glands just adjacent to gastric cancer from fresh samples without the contamination of cancerous glands. Nevertheless, the present results that CNA patterns and the number of mutations in gastric IM and early-stage gastric cancer are similar could support the hypothesis that most intestinal-type gastric cancers originate from IM glands. To further clarify the origin of gastric cancer, additional studies for cancer formation dynamics, such as organoid research may be required.

The present results indicate that IM glands have a higher risk for gastric cancer development than severe atrophic mucosa without IM. Various studies have demonstrated that the eradication of *H. pylori* reduces the incidence of gastric cancer (42,43). After eradication, the atrophic mucosa gradually recover, but most gastric IM remain(44), and gastric cancer still emerges with a high probability. Gastric cancer risk remains in IM glands in the gastric mucosa even after the eradication of *H. pylori*. Thus, to prevent

gastric cancer development, *H. pylori* eradication should be performed before the development of IM. For individuals with gastric IM, the affected region should be evaluated carefully using short-interval endoscopy as a surveillance for cancer development. Furthermore, our results suggest that IM glands with CNAs might be at the highest risk of gastric cancer development. For the detection of them in the clinical practice, the development of an assay that can analyze CNAs more easily and cost-effectively is important.

This study had certain limitations. First, we did not show clonal relationship between gastric IM and cancer directly as described above. Second, we focused on genetic aberrations existing clonally in each gastric single gland and did not address low-abundance mutations in some cells of gastric single glands with very low VAF(17,45). Third, we investigated genetic data with inflamed stomach tissue from only 4 patients who had not developed gastric cancer. Although we obtained interesting data from the non-cancerous gastric mucosa of patients with gastric cancer to clarify the high-risk glands, these issues should be addressed in future studies.

In summary, we demonstrated that gastric IM is susceptible to somatic mutations and CNAs, and that it expands multi-focally with the acquisition of CNAs commonly found in gastric cancer, resulting in field cancerization. The existence of IM in gastric atrophic mucosa inflamed by *H. pylori* infection is important for the surveillance of gastric cancer development. It is also important to determine the mechanism by which IM develops during chronic gastric inflammation in order to develop therapeutic and preventive approaches for gastric cancer specifically targeting IM.

Authors' contributions

KK: study concept and design; acquisition of data; analysis and interpretation of data; drafting of the manuscript; critical revision of the manuscript for important intellectual content, TS: study concept and design; analysis and interpretation of data; drafting of the manuscript; critical revision of the manuscript for important intellectual content, AT: study concept and design; analysis and interpretation of data, NK: analysis and interpretation of data; YT: analysis and interpretation of data, TH: analysis and interpretation of data, HT: analysis and interpretation of data, AM: acquisition of data, MT: acquisition of data, TI: analysis and interpretation of data, EI: analysis and interpretation of data, MN: acquisition, analysis and interpretation of data, YE: analysis and interpretation of data, KT: analysis and interpretation of data, YU: analysis and interpretation of data, SM: acquisition of data, KO: acquisition of data, SO: critical revision of the manuscript for important intellectual content, HM: study concept and design; study supervision, HS: analysis and interpretation of data, study supervision.

Acknowledgements

We thank Drs Soichi Arasawa, Masako Mishima, Shigeharu Nakano, and Fumiyasu Nakamura for interpretation of data and helpful advice. We also thank Drs Akira Yokoyama, and Atsushi Yamada for material support. This work was supported by grants from the Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research KAKENHI (grant number 18K15781), Takeda Science Foundation (grant number 203190700042), and The NOVARTIS Foundation for the Promotion of Science (grant number 203200700003).

References

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

2. Lauren P. The two histological main types of gastric carcinoma: Diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. *Acta Pathol Microbiol Scand* **1965**;64:31–49.

3. Correa P. A human model of gastric carcinogenesis. *Cancer Res* 1988;48:3554–60.

4. Polk DB, Peek RM Jr. Helicobacter pylori: gastric cancer and beyond. *Nat Rev Cancer* **2010**;10:403–14.

5. Akbari M, Tabrizi R, Kardeh S, Lankarani KB. Gastric cancer in patients with gastric atrophy and intestinal metaplasia: A systematic review and meta-analysis. *PLoS One* **2019**;14:e0219865.

6. Chen H-N, Wang Z, Li X, Zhou Z-G. Helicobacter pylori eradication cannot reduce the risk of gastric cancer in patients with intestinal metaplasia and dysplasia: evidence from a meta-analysis. *Gastric Cancer* **2016**;19:166–75.

7. Slack JMW. Metaplasia and transdifferentiation: from pure biology to the clinic. *Nat Rev Mol Cell Biol* **2007**;8:369–78.

8. Giroux V, Rustgi AK. Metaplasia: tissue injury adaptation and a precursor to the dysplasia-cancer sequence. *Nat Rev Cancer* **2017**;17:594–604.

9. Curtius K, Wright NA, Graham TA. An evolutionary perspective on field cancerization. *Nat Rev Cancer* **2018**;18:19–32.

10. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, et al. Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **2018**;173:1823.

11. Nguyen LH, Goel A, Chung DC. Pathways of colorectal carcinogenesis. *Gastroenterology* **2020**;158:291–302.

12. Blokzijl F, de Ligt J, Jager M, Sasselli V, Roerink S, Sasaki N, et al. Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* **2016**;538:260–4.

13. Chiba T, Marusawa H, Ushijima T. Inflammation-associated cancer development in digestive organs: mechanisms and roles for genetic and epigenetic modulation. *Gastroenterology* **2012**;143:550–63.

14. Matsumoto Y, Marusawa H, Kinoshita K, Endo Y, Kou T, Morisawa T, et al. Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. *Nat Med* **2007**;13:470–6.

15. Matsumoto Y, Marusawa H, Kinoshita K, Niwa Y, Sakai Y, Chiba T. Up-regulation of activation-induced cytidine deaminase causes genetic aberrations at the CDKN2b-CDKN2a in gastric cancer. *Gastroenterology* **2010**;139:1984–94.

16. Nagata N, Akiyama J, Marusawa H, Shimbo T, Liu Y, Igari T, 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**;49:427–35.

17. Shimizu T, Marusawa H, Matsumoto Y, Inuzuka T, Ikeda A, Fujii Y, et al. Accumulation of somatic mutations in TP53 in gastric epithelium with Helicobacter pylori infection. *Gastroenterology* **2014**;147:407-17.e3.

18. Huang KK, Ramnarayanan K, Zhu F, Srivastava S, Xu C, Tan ALK, et al. Genomic and Epigenomic Profiling of High-Risk Intestinal Metaplasia Reveals Molecular Determinants of Progression to Gastric Cancer. *Cancer Cell* **2018**;33:137-150.e5.

19. Gutierrez–Gonzalez L, Graham TA, Rodriguez–Justo M, Leedham SJ, Novelli MR, Gay LJ, et al. The clonal origins of dysplasia from intestinal metaplasia in the human stomach. *Gastroenterology* **2011**;140:1251-1260.e6.

20. McDonald SAC, Greaves LC, Gutierrez-Gonzalez L, Rodriguez-Justo M, Deheragoda M, Leedham SJ, et al. Mechanisms of field cancerization in the human stomach: the expansion and spread of mutated gastric stem cells. *Gastroenterology* **2008**;134:500–10.

21. Schlemper RJ, Riddell RH, Kato Y, Borchard F, Cooper HS, Dawsey SM, et al. The Vienna classification of gastrointestinal epithelial neoplasia. *Gut* **2000**;47:251–5.

22. Mizuguchi A, Takai A, Shimizu T, Matsumoto T, Kumagai K, Miyamoto S, et al. Genetic features of multicentric/multifocal intramucosal gastric carcinoma. *Int J Cancer* **2018**.1923–34.

23. Kakiuchi N, Yoshida K, Uchino M, Kihara T, Akaki K, Inoue Y, et al. Frequent mutations that converge on the NFKBIZ pathway in ulcerative colitis. *Nature* **2020**;577:260–5.

24. Halper-Stromberg E, Frelin L, Ruczinski I, Scharpf R, Jie C, Carvalho B, et al. Performance assessment of copy number microarray platforms using a spike-in experiment. *Bioinformatics* **2011**;27:1052–60.

25. Franch-Expósito S, Bassaganyas L, Vila-Casadesús M, Hernández-Illán E, Esteban-Fabró R, Díaz-Gay M, et al. CNApp, a tool for the quantification of copy number alterations and integrative analysis revealing clinical implications. *Elife* **2020**;9.

26. Newman S. Interactive analysis of large cancer copy number studies with Copy Number Explorer. *Bioinformatics* **2015**;31:2874–6.

27. Hu J, Hwang SS, Liesa M, Gan B, Sahin E, Jaskelioff M, et al. Antitelomerase therapy provokes ALT and mitochondrial adaptive mechanisms in cancer. *Cell* **2012**;148:651–63.

28. Wiedemeyer WR, Dunn IF, Quayle SN, Zhang J, Chheda MG, Dunn GP, et al. Pattern of retinoblastoma pathway inactivation dictates response to CDK4/6 inhibition in GBM. *Proc Natl Acad Sci U S A* **2010**;107:11501–6.

29. Nikaido M, Kakiuchi N, Miyamoto S, Hirano T, Takeuchi Y, Funakoshi T, et al. Indolent feature of Helicobacter pylori-uninfected intramucosal signet ring cell carcinomas with CDH1 mutations. *Gastric Cancer* **2021**;24:1102–14.

30. Yokoyama A, Kakiuchi N, Yoshizato T, Nannya Y, Suzuki H, Takeuchi Y, et al. Age-related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature* **2019**. 312–7.

31. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer* **2018**;18:696–705.

32. Nanki K, Fujii M, Shimokawa M, Matano M, Nishikori S, Date S, et al. Somatic inflammatory gene mutations in human ulcerative colitis epithelium. *Nature* **2020**;577:254–9.

33. Lee-Six H, Olafsson S, Ellis P, Osborne RJ, Sanders MA, Moore L, et al. The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* **2019**;574:532–7.

34. Leedham SJ, Graham TA, Oukrif D, McDonald SAC, Rodriguez-Justo M, Harrison RF, et al. Clonality, founder mutations, and field cancerization in human ulcerative colitis-associated neoplasia. *Gastroenterology* **2009**;136:542-50.e6.

35. Stachler MD, Camarda ND, Deitrick C, Kim A, Agoston AT, Odze RD, et al. Detection of mutations in Barrett's esophagus before progression to high-grade dysplasia or adenocarcinoma. *Gastroenterology* **2018**;155:156–67.

36. Bian YS, Osterheld MC, Bosman FT, Benhattar J, Fontolliet C. p53 gene mutation and protein accumulation during neoplastic progression in Barrett's esophagus. *Mod Pathol* **2001**;14:397–403.

37. Dolan K, Walker SJ, Gosney J, Field JK, Sutton R. TP53 mutations in malignant and premalignant Barrett's esophagus. *Dis Esophagus* **2003**;16:83–9.

38. Snippert HJ, Schepers AG, van Es JH, Simons BD, Clevers H. Biased competition between Lgr5 intestinal stem cells driven by oncogenic mutation induces clonal expansion. *EMBO Rep* **2014**;15:62–9.

39. Vermeulen L, Snippert HJ. Stem cell dynamics in homeostasis and cancer of the intestine. *Nat Rev Cancer* **2014**;14:468–80.

40. Correa P, Shiao YH. Phenotypic and genotypic events in gastric carcinogenesis. *Cancer Res* **1994**;54:1941s–3s.

41. Fenoglio-Preiser CM, Wang J, Stemmermann GN, Noffsinger A. TP53 and gastric carcinoma: a review. *Hum Mutat* **2003**;21:258–70.

42. Choi IJ, Kook M-C, Kim Y-I, Cho S-J, Lee JY, Kim CG, et al. Helicobacter pylori therapy for the prevention of metachronous gastric cancer. *N Engl J Med* **2018**;378:1085–95.

43. Fukase K, Kato M, Kikuchi S, Inoue K, Uemura N, Okamoto S, 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* **2008**;372:392–7.

44. Kodama M, Okimoto T, Mizukami K, Hirashita Y, Wada Y, Fukuda M, et al. Gastric mucosal changes, and sex differences therein, after Helicobacter pylori eradication: A long-term prospective follow-up study. *J Gastroenterol Hepatol* **2021**;36:2210–6.
45. Yamashita S, Kishino T, Takahashi T, Shimazu T, Charvat H, Kakugawa Y, et al. Genetic and epigenetic alterations in normal tissues have differential impacts on cancer risk among tissues. *Proc Natl Acad Sci U S A* **2018**;115:1328–33.

Figure 1. Sample collection and isolation of gastric single- and clustered glands

(a) Sample lists. “C” indicates patients with gastric cancer, whereas “WO” indicates patients without gastric cancer. All these patients had inflamed gastric mucosa with atrophic gastritis and intestinal metaplasia. “N” and “KGC” indicates cases with non-inflamed gastric mucosa. WES+CN analyses were performed for single glands following WGA. Targeted-capture deep sequencing and CN analyses were performed for clustered glands. WES of FFPE samples was performed for IGC. Previous dataset and TCGA-STAD dataset were re-analyzed*, **, ***. AG, atrophic gastritis; CN, copy number; GC, gastric cancer; IGC, intramucosal gastric cancer; WES, whole-exome sequencing. (b) Schematic for the gland isolation procedure. (c) Non-cancerous gastric tissue samples were collected from specimens endoscopically resected for IGC using a punch biopsy with a 2 mm diameter. (d) Macroscopic images to identify IM. Gastric epithelium was manually dissociated from non-epithelial components. After staining with Alcian blue, gastric single or clustered glands with the sole presence or absence of IM, were isolated. Contrast was enhanced by dehydration with ethanol. WGA, whole genome amplification.

Figure 2. Somatic mutations and copy number aberrations in gastric single glands

(a) Number of somatic mutations detected by WES of single IM glands (orange), non-IM glands (green), and normal glands (blue). Closed- and open circles represent samples from the patients with and without IGC, respectively. Linear regression lines are shown with slope and R^2. (b) Number of somatic mutations in gastric single glands (n = 71) and IGC with the MSS phenotype (n = 8) detected by WES in the present study, and those in gastric cancer at stage IA (n = 5) and stage IB (n = 20). (c) Representatives of copy number aberrations in single IM glands. The upper panel of each case represents the copy number and the lower one represents allelic ratio.

Figure 3. Mutational signature analysis in gastric single glands

(a) Sequence context of somatic mutations in IGC, single IM glands, non-IM glands, and normal glands in the stomach. (b) Mutation patterns determined in single normal glands, non-IM glands, IM glands, and IGC in the stomach. IGC, intramucosal gastric cancer.

Figure 4. Mutations frequently detected in gastric single glands

(a) Frequently mutated genes detected in more than three cases via WES of gastric single glands. Gene symbols for which mutations were detected in more than four cases are listed. Details are shown in Supplementary Fig. S8. (b) Mutated cancer-related genes detected in gastric single glands by WES. Genes were classified as Tier 1 (established cancer-related genes) or Tier 2 (suspected cancer-related genes) based on the COSMIC cancer gene census. Statistical analysis was performed using Mann–Whitney *U*-test. *$p < 0.05$. High-resolution image shown in Supplementary Fig. S9.

Figure 5. Clonal ordering analysis using somatic mutations detected in gastric single glands and accompanying IGC

Three representative cases (C01, C03, C09) are shown. Each case represents narrow-band images obtained by endoscopy, macroscopic and Alcian blue-stained stereomicroscopic images of the resected specimen, scheme of sampling locations, Venn-diagram, and visualization of clonal ordering analysis of somatic mutations detected by WES of gastric single glands. Mutations in Tier 1 or Tier 2 genes are shown in clonal ordering analysis. Tissue samples including the border between the IM and non-IM region in the non-cancerous area were obtained from the orange circle. Tumors are shown by the red dotted line in the endoscopic images.

Figure 6. Somatic mutations detected in gastric clustered IM glands or non-IM glands

(a) Mutated genes detected in gastric clustered glands by targeted-capture deep sequencing. The mutational frequencies of each gene detected in gastric clustered glands and in gastric cancer from the TCGA-STAD database are indicated using color bars. Genes were classified into Tier 1 (established cancer-related gene), Tier 2 (suspected cancer-related gene), and unregistered based on the COSMIC cancer gene census. FS, frameshift; NF, non-frameshift; NS, non-synonymous; S, synonymous; SNV, single-nucleotide variant. Fisher's exact test *$p < 0.05$. (b) VAF distribution of mutated genes in gastric clustered glands. Each dot represents the mutation detected in clustered glands. Black dots indicate Tier 1 or Tier 2 genes.

Figure 7. CN analysis of gastric clustered glands and gastric cancers

(a,b) CNAs of gastric clustered glands in the present study, IGC (n = 41) from our previous study, and gastric cancer at stage IA (n = 10) and IB (n = 34) from TCGA-STAD. All gastric cancers at stage IA were submucosal carcinomas. (a) Distribution of broad and focal CN scores. CN scores for gain and loss were calculated separately. (b) Distribution of SGOL scores across the genome. Blue and red bars indicate gain and loss, respectively. (c) Proportion of CNAs in clustered IM glands. Each dot represents the CNA detected in clustered IM glands. The proportion of CNAs was calculated assuming that CN becomes 1 or 3 under the condition that whole cells acquire CN gain or CN loss, respectively. BCS, broad copy number score; CNA, copy number aberration; FCS, focal copy number score; SGOL, segment gain or loss.

Figure 8. Schematic for accumulation of genetic alterations and expansion of gastric glands during gastric carcinogenesis

Schematic images of clonality, number of mutations, broad CN gain and loss, and focal CN gain and loss of the glands are shown in gastric normal glands, non-IM and IM glands of gastric mucosa inflamed by *H. pylori* infection, IGC, Stage IA gastric cancer (submucosal gastric carcinoma), and Stage IB gastric cancer. In the schematic image of clonality, the colored circle represents a gland, and blue open circles depict clonally expanded glands. Somatic mutations are accumulated in IM glands to the same degree as in IGC and submucosal gastric carcinoma (Stage IA). Broad CN gains are also accumulated in the IM glands to an extent similar to that in IGC, and broad CN loss and focal CNAs are increased with adenocarcinoma progression from early to advanced stages. Gastric IM glands expand multi-focally with the acquisition of CNAs, resulting in field cancerization, and are possibly followed by malignant transformation through

the acquisition of additional genetic aberrations, such as cancer driver mutations and further CNAs. CN, copy number; CNA, copy number aberration; IGC, intramucosal gastric cancer.

Figure 1

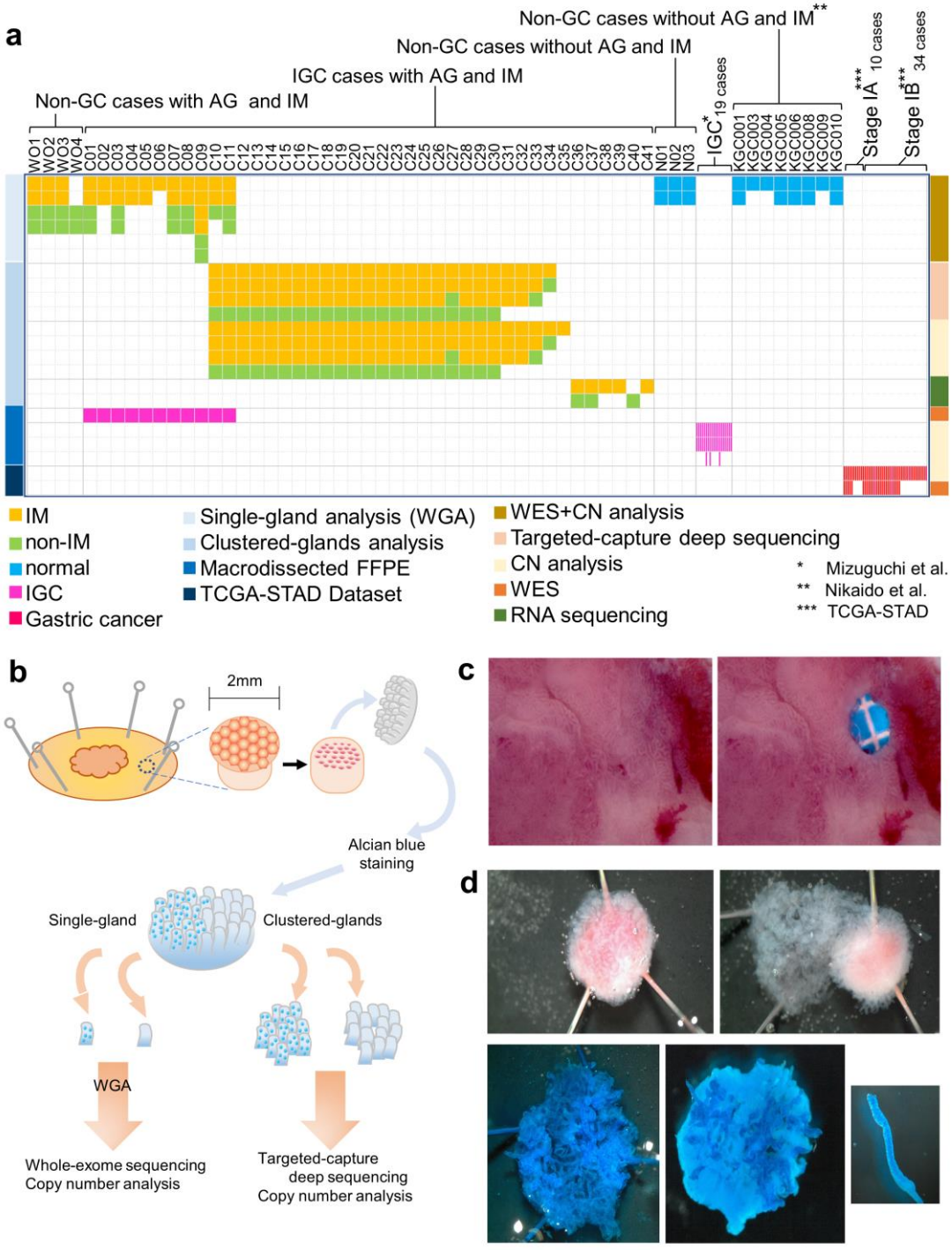


Figure 2

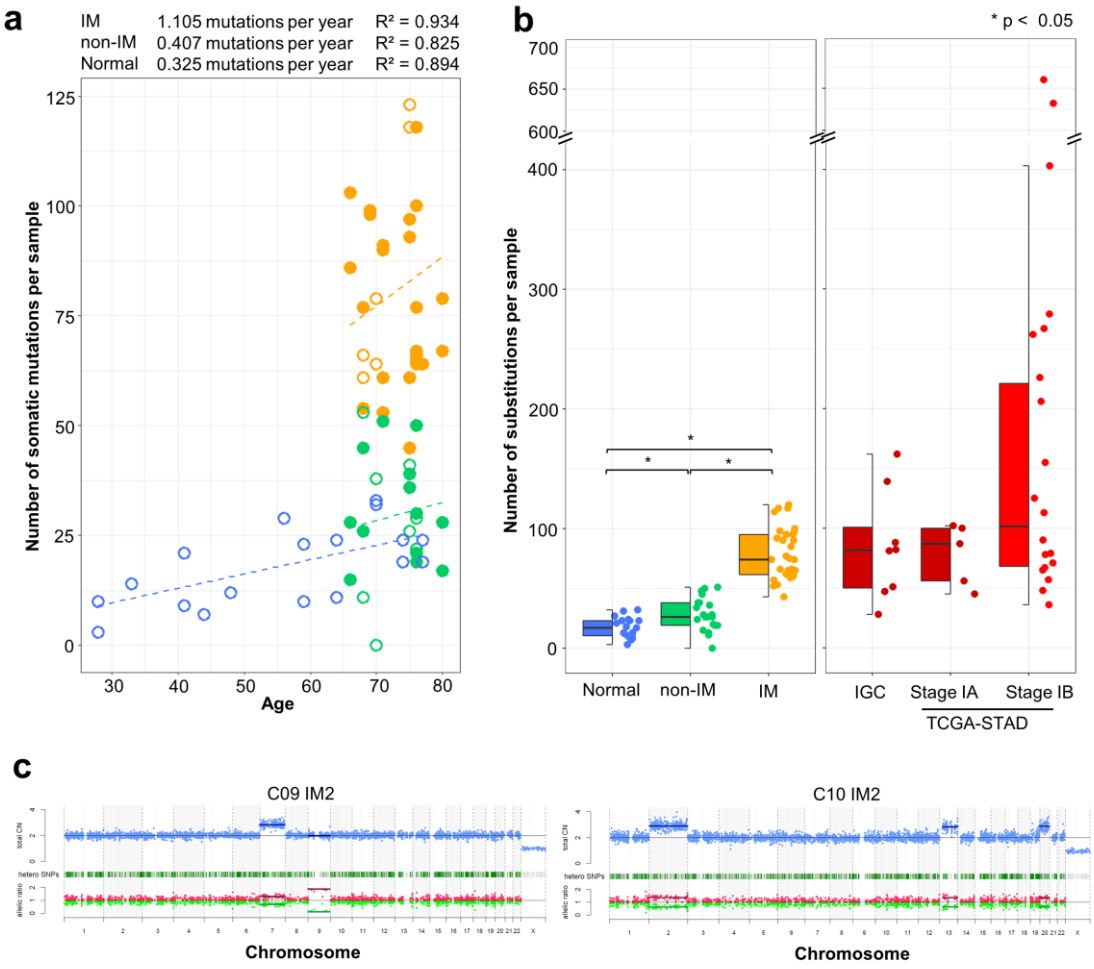

a
IM 1.105 mutations per year R² = 0.934
non-IM 0.407 mutations per year R² = 0.825
Normal 0.325 mutations per year R² = 0.894
Number of somatic mutations per sample
Age
b
* p < 0.05
Number of substitutions per sample
Normal
non-IM
IM
IGC
Stage IA
Stage IB
TCGA-STAD
c
C09 IM2
C10 IM2
Chromosome
Chromosome

Figure 3

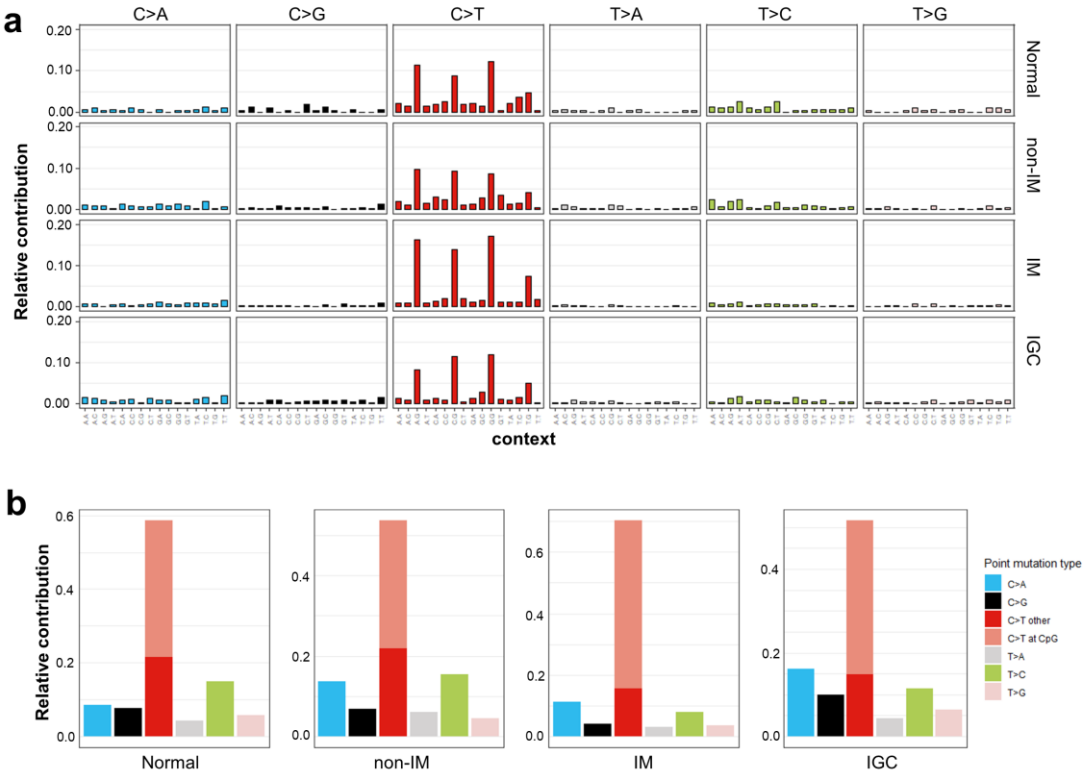

a
C>A
C>G
C>T
T>A
T>C
T>G
Normal
non-IM
IM
IGC
Relative contribution
0.20
0.10
0.00
context
b
Relative contribution
0.6
0.4
0.2
0.0
Normal
non-IM
IM
IGC
Point mutation type
C>A
C>G
C>T other
C>T at CpG
T>A
T>C
T>G

Figure 4

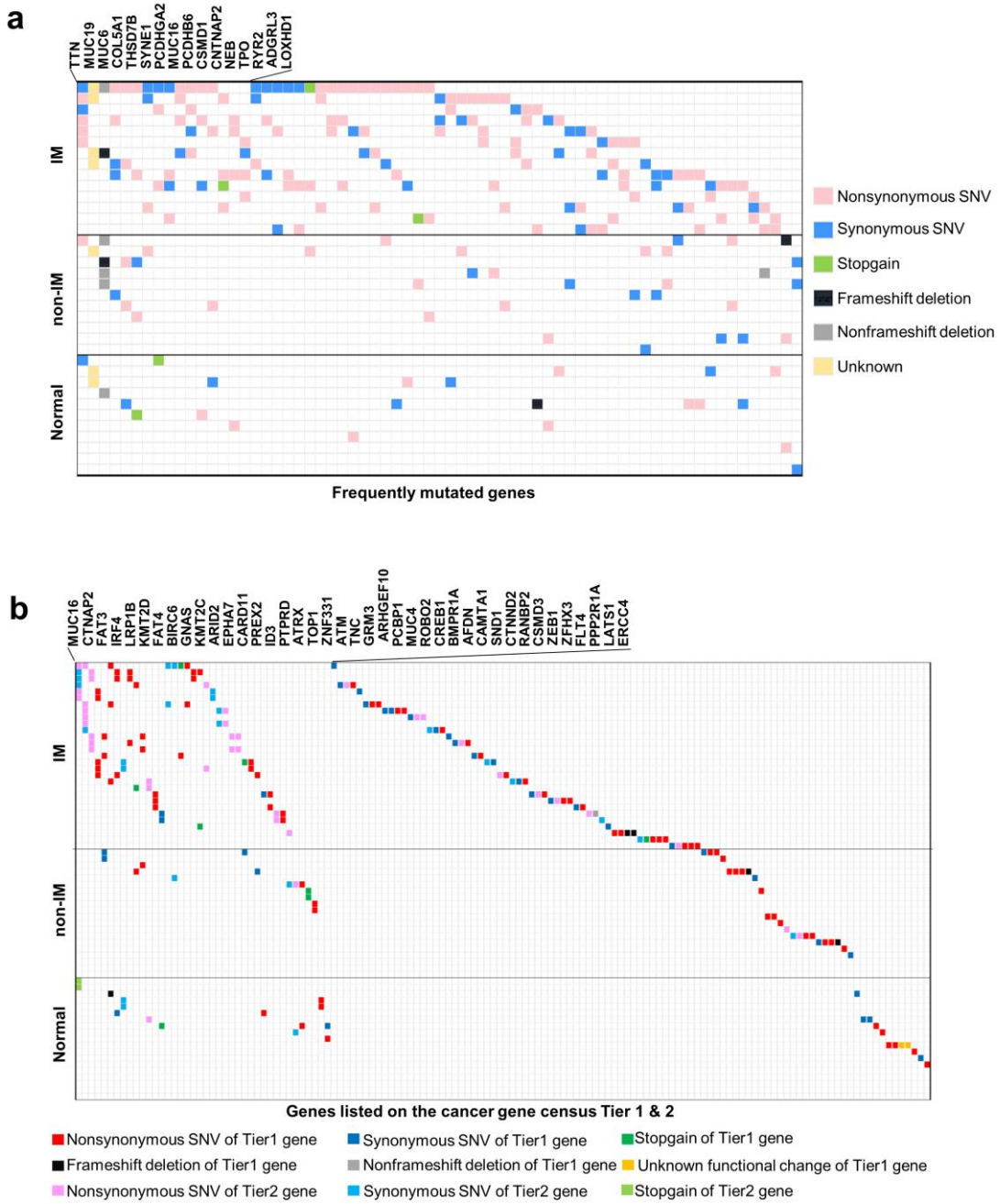
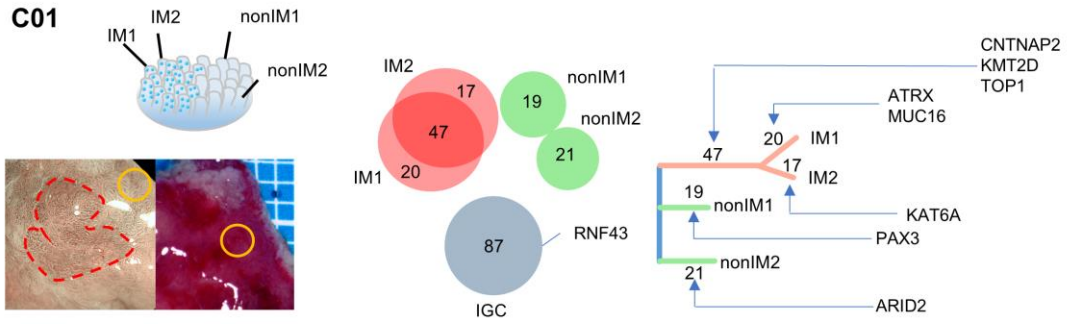
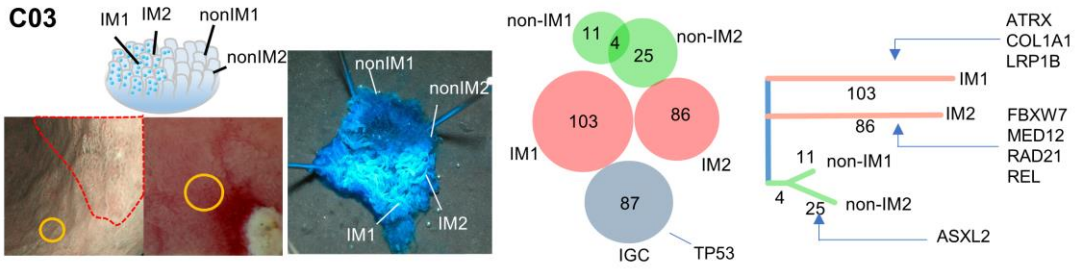
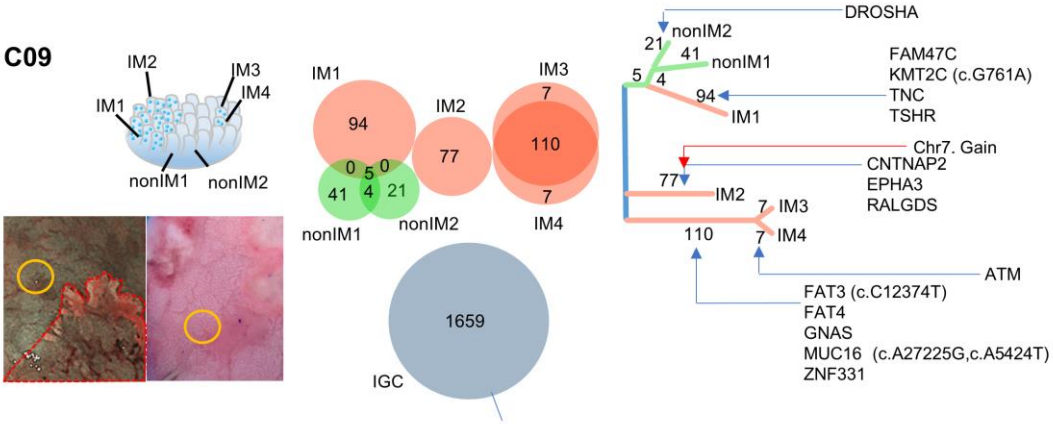

a
TTN
MUC19
MUC6
COL5A1
THSD7B
SYNE1
PCDHGA2
MUC16
PCDHB6
CSMD1
CNTNAP2
NEB
TPO
RYR2
ADGRL3
LOXHD1
IM
non-IM
Normal
Frequently mutated genes
Nonsynonymous SNV
Synonymous SNV
Stopgain
Frameshift deletion
Nonframeshift deletion
Unknown
b
MUC16
CTNAP2
FAT3
IRF4
LRP1B
KMT2D
FAT4
BIRC6
GNAS
KMT2C
ARID2
EPHA7
CARD11
PREX2
ID3
PTPRD
ATRX
TOP1
ZNF331
ATM
TNC
GRM3
ARHGEF10
PCBP1
MUC4
ROBO2
CREB1
BMPR1A
AFDN
CAMTA1
SND1
CTNND2
RANBP2
CSMD3
ZEB1
ZFHX3
FLT4
PPP2R1A
LATS1
ERCC4
IM
non-IM
Normal
Genes listed on the cancer gene census Tier 1 & 2
Nonsynonymous SNV of Tier1 gene
Synonymous SNV of Tier1 gene
Stopgain of Tier1 gene
Frameshift deletion of Tier1 gene
Nonframeshift deletion of Tier1 gene
Unknown functional change of Tier1 gene
Nonsynonymous SNV of Tier2 gene
Synonymous SNV of Tier2 gene
Stopgain of Tier2 gene

Figure 5







ACVR2A	ASXL1	BIRC6	CLP1	FAT3 (c.T12278C)	KMT2D		MYH11	PRPF40B	SALL4	UBR5
ANK1	ATM	CANT1	CLTC	FES	LARP4B		NBEANBN	PTPRK	SDC4	ZFHX3
APC	ATR	CCDC6	CTCF	FIP1L1	MAF		NSD3	RAF1	SETD1B	ZMYM2
AR	BAP1	CD79B	CTNNA2	FOXR1	MEN1		NTRK3	RARA	SETD2	ZNRF3
ARHGAP26	BAX	CDK12	ELL	GATA3	MET		NTRK3	RBM15	SND1	
ARHGAP5	BCL10	CDK6	EP300	HSP90AA1	MN1		OMD	RET	TGFBR2	
ARID1A	BCL11B	CHD4	ERC1	HSP90AB1	MUC16	(c.G37162A)	PMS1	RNF43	TNFAIP3	
ARID1B	BCORL1	CIITA	FANCD2	KMT2C (c.8390delA)		(c.32528delT)	PRDM2	RPL22	TSC2	

Figure 6

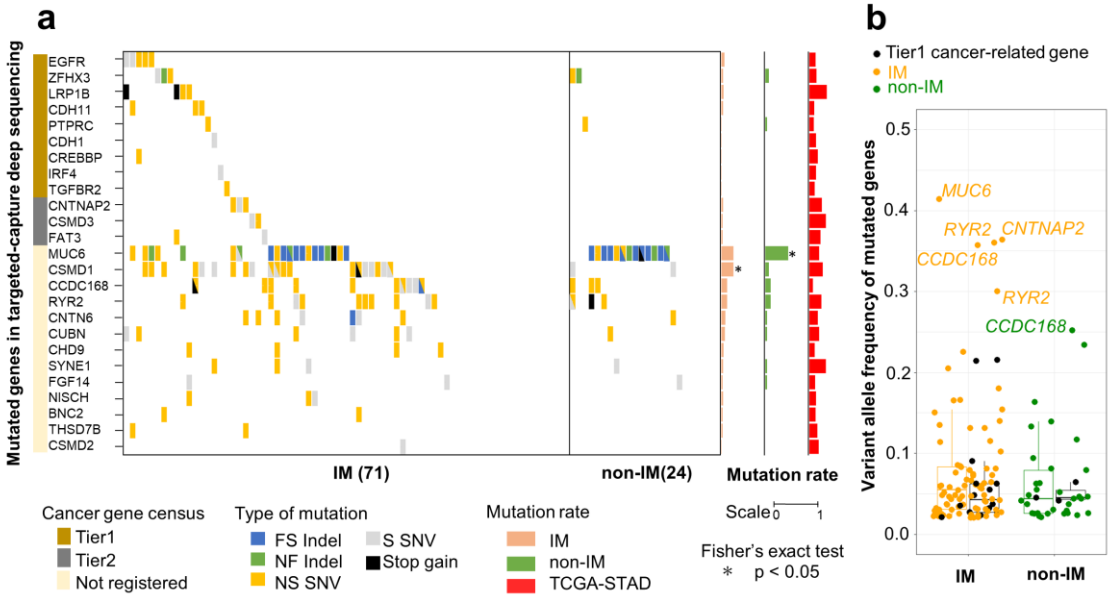

a
Mutated genes in targeted-capture deep sequencing
EGFR
ZFHX3
LRP1B
CDH11
PTPRC
CDH1
CREBBP
IRF4
TGFBR2
CNTNAP2
CSMD3
FAT3
MUC6
CSMD1
CCDC168
RYR2
CNTN6
CUBN
CHD9
SYNE1
FGF14
NISCH
BNC2
THSD7B
CSMD2
IM (71)
non-IM(24)
Mutation rate
Cancer gene census
Tier1
Tier2
Not registered
Type of mutation
FS Indel
NF Indel
NS SNV
S SNV
Stop gain
Mutation rate
IM
non-IM
TCGA-STAD
Scale 0 1
Fisher's exact test
* p < 0.05
b
Tier1 cancer-related gene
IM
non-IM
Variant allele frequency of mutated genes
0.5
0.4
0.3
0.2
0.1
0.0
MUC6
RYR2
CNTNAP2
CCDC168
RYR2
CCDC168
IM
non-IM

Figure 7

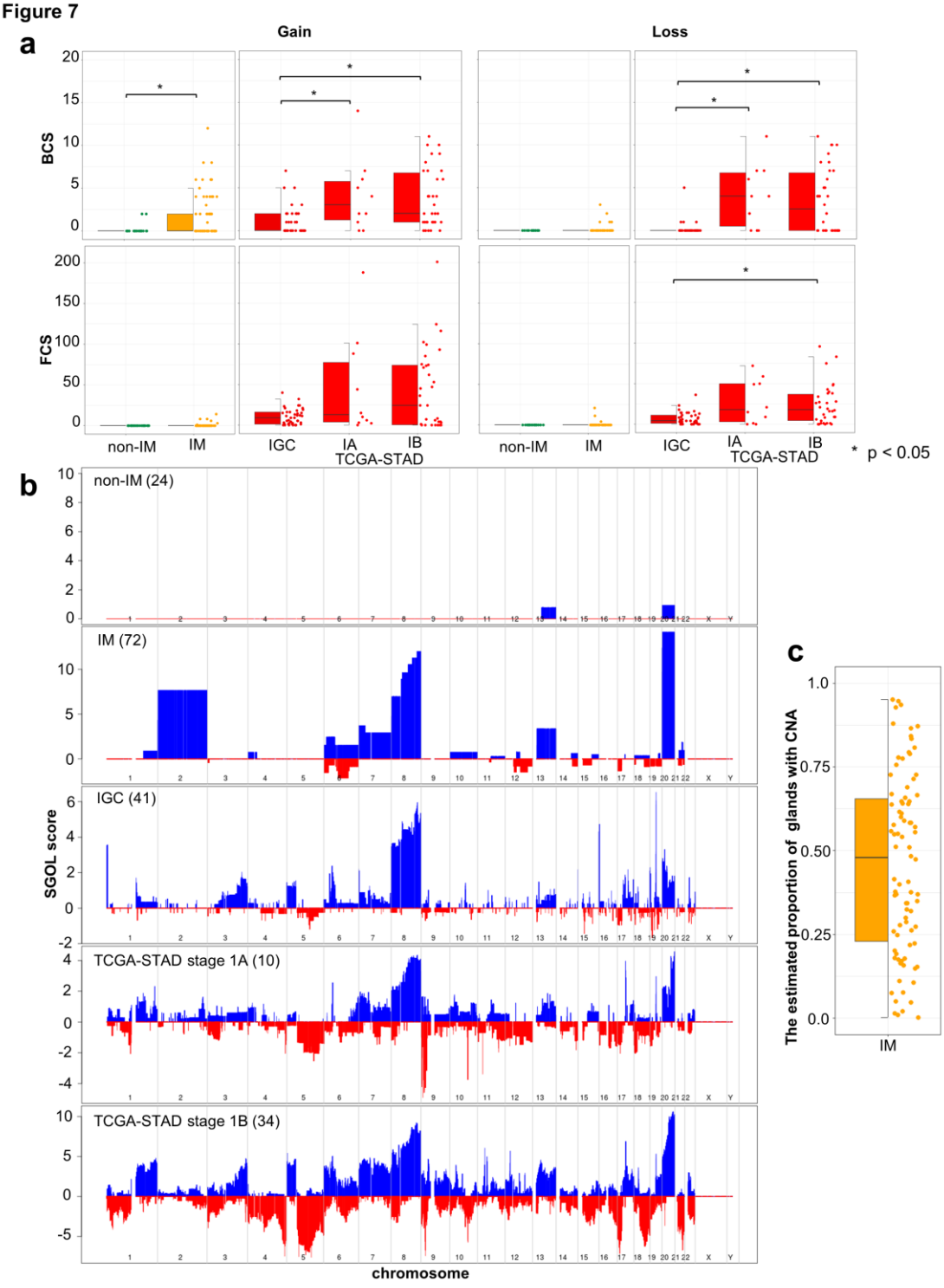
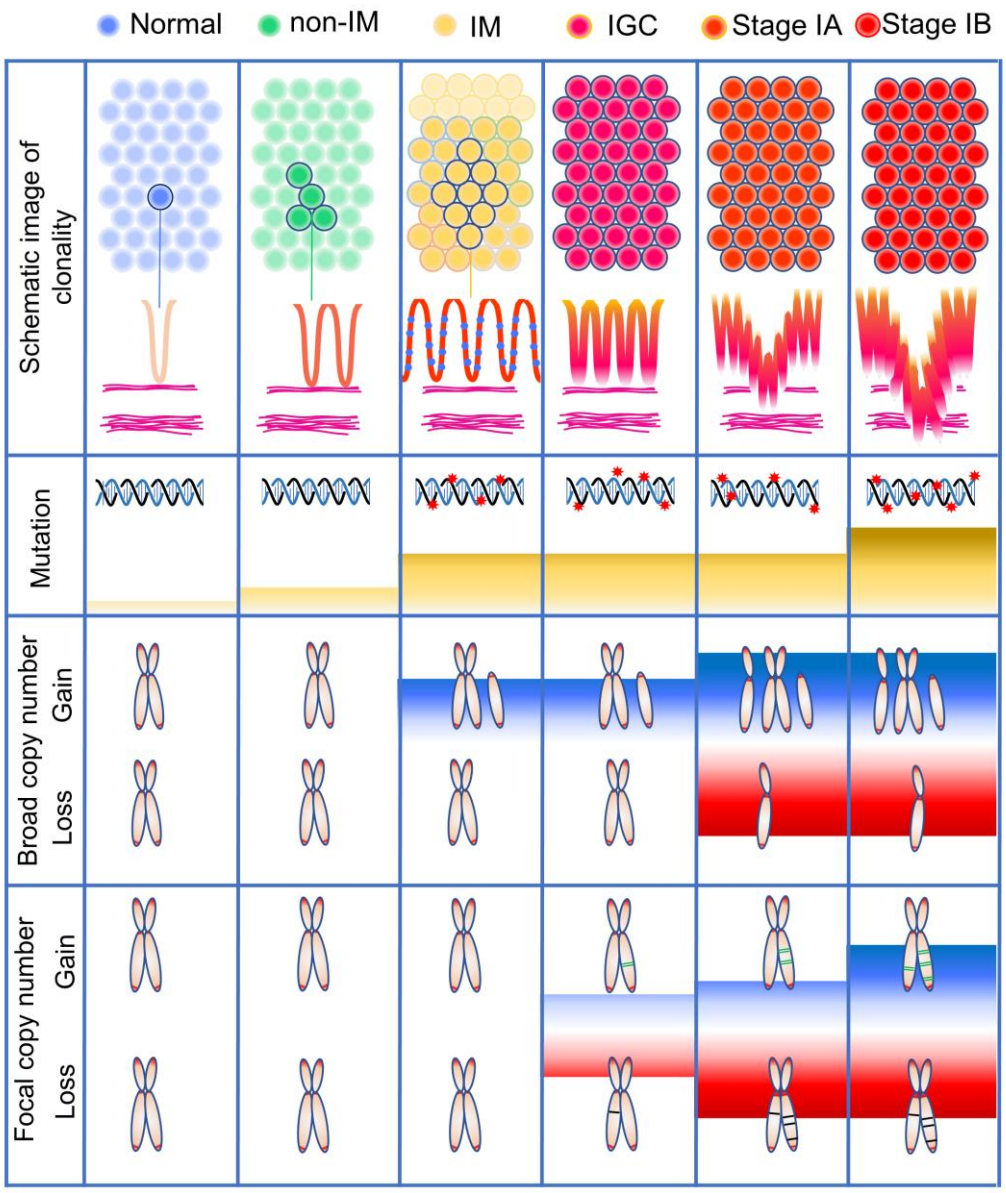

a
Gain
Loss
BCS
FCS
non-IM
IM
IGC
IA
IB
TCGA-STAD
* p < 0.05
b
non-IM (24)
IM (72)
IGC (41)
TCGA-STAD stage 1A (10)
TCGA-STAD stage 1B (34)
SGOL score
chromosome
c
The estimated proportion of glands with CNA
IM

Figure 8


Normal
non-IM
IM
IGC
Stage IA
Stage IB
Schematic image of clonality
Mutation
Broad copy number
Gain
Loss
Focal copy number
Gain
Loss

Supplementary material and methods

Whole-genome amplification (WGA) and whole-exome sequencing (WES) of gastric single glands

After WGA of extracted DNA, one amplified DNA sample was used for WES analysis, whereas the other was used for the validation study. gDNA from peripheral blood was extracted using the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) and used as a germline control. The quantity and purity of gDNA were assessed using a Qubit® 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA). In total, 2 μg of amplified DNA was fragmented on a E220 focused ultrasonicator Covaris (Covaris, Woburn, MA, USA), followed by end repair, A-tailing, and adapter ligation using Agilent SureSelect XT (Agilent Technologies, Santa Clara, CA, USA) according to the manufacturer's instructions. The fragmentation status was evaluated with the Agilent 2200 TapeStation system using the Genomic DNA ScreenTape assay (Agilent Technologies). WES libraries were prepared using Agilent SureSelect Human All Exon *v*.6 (Agilent Technologies), and sequencing of enriched exon fragments was performed on a NovaSeq 6000 system (Illumina, San Diego, CA, USA) in the 150-bp paired-end mode. After WGA of single-gland DNA, we conducted WES and detected genetic alterations by comparison with normal variants in matched peripheral leukocytes.

WES of formalin-fixed paraffin-embedded (FFPE) tumor samples

For genetic analysis of IGC, cancer components were macroscopically trimmed from 7-μm-thick FFPE tissues. Tumor purity in the samples was confirmed by two independent investigators to be greater than 0.4 through visual evaluation of hematoxylin

and eosin-stained slides. Genomic DNA (gDNA) was extracted from macro-dissected FFPE samples using the GeneRead DNA FFPE Kit (Qiagen, Hilden, Germany). In total, 200 ng of gDNA per FFPE tumor sample was fragmented on a E220 focused Covaris ultrasonicator (Covaris, Woburn, MA, USA) followed by end repair, A-tailing, and adapter ligation using the KAPA Library Preparation Kit (Kapa Biosystems Inc, Wilmington, MA, USA) according to the manufacturer's instructions. WES libraries were prepared using Agilent SureSelect Human All Exon *v*.6 (Agilent Technologies, Santa Clara, CA, USA) and sequencing of enriched exon fragments was performed on a NovaSeq 6000 system (Illumina, San Diego, CA, USA) in the 150-bp paired-end mode.

Targeted-capture deep sequencing of gastric clustered glands

In total, 200 ng gDNA was fragmented using an enzymatic fragmentation kit Lotus (IDT) and adapter-ligated according to the manufacturer's instructions. The libraries were constructed with xGen custom-designed Gene Capture Pools (IDT) to cover the 98 relevant targeted genes based on the whole-exome sequencing (WES) results from gastric single glands in the present study, TCGA (https://cancergenome.nih.gov/, RRID:SCR_003193), ICGC (https://dcc.icgc.org/, RRID:SCR_021722), COSMIC (http://cancer.sanger.ac.uk/cosmic, RRID:SCR_002260), and previous reports on gastric adenocarcinoma(1–3) and IM(4). After the sequencing libraries were quantified and their quality was checked using the 2200 TapeStation Instrument (High Sensitivity Assay), sequencing was performed on a NovaSeq 6000 system in the 150-bp paired-end mode.

Mutation calling

Mutation calling was performed using the Genomon2 pipeline (v.2.6), as described previously(5). Sequencing reads were aligned to the human reference genome (GRCh38)

for mutation calling. Somatic mutations were detected by eliminating polymorphisms and sequencing errors. To achieve this, Genomon2 first discards any low-quality, unreliable reads and variants according to the following criteria for non-cancerous samples: (i) mapping quality < 20, (ii) base call quality < 15. After further excluding variants that are not supported by (i) a sufficient number of reads (total reads ≥ 8 and variant reads ≥ 4); (ii) VAFs ≥ 0.25 for samples of single glands subjected to whole genome amplification (WGA), ≥ 0.02 for samples of clustered glands, ≥ 0.10 for FFPE tumor samples and TCGA-STAD dataset, and < 0.02 for the germline control, the remaining variants were interrogated to establish whether they were observed at significantly higher VAFs than expected for errors ($p < 0.01$ for single-gland samples subjected to WGA, $p < 0.05$ for clustered-gland samples, FFPE tumor samples and TCGA-STAD dataset). Putative germline variants were also excluded by comparing their VAFs with those of matched controls using Fisher's exact test ($p < 0.01$), which also eliminated any remaining sequencing errors. Clustered glands were analyzed using other samples from the same patients as controls to eliminate polymorphisms. Insertions were frequently introduced during WGA and were therefore excluded from single-gland analysis. The monoclonal origin of single-gland cells, which are derived from a single stem cell at a given time, was evident from the normal distribution of VAFs around 0.5. All detected mutations of single glands were checked manually using the Integrative Genomic Viewer (IGV)(6). The called mutations were annotated using ANNOVAR(7). Mutation signatures were analyzed using the Mutational Patterns R package(8) based on the Mutational Signatures version 2 on COSMIC. To evaluate the significantly mutated genes as driver mutations, we used the dNdScv R package(9) with the results of targeted-capture deep sequencing.

External data

For genetic analysis of early-stage gastric cancer, we re-analyzed the gastric adenocarcinoma dataset in TCGA-STAD. For mutation analysis, gastric cancer cases with microsatellite instability (MSI) or hypermutated-single-nucleotide variant predominant (HM-SNV) molecular features or those in which the DNA fractions with tumor cells were less than 0.4 were excluded. Thus, five cases of stage IA and 20 cases of stage IB were included (Supplementary Table S3). For CN analysis, we obtained the data of segmented log2 ratios for 10 cases of gastric cancer at stage IA, which were all submucosal gastric carcinoma, and 34 cases of gastric cancer at stage IB (Supplementary Table S4). For genetic analysis of non-inflamed gastric mucosa, we re-analyzed the dataset of the previous study (accession number JGAS000313) in which genomic data of gastric single glands were obtained in the same manner as the present study(10). Among the samples in the dataset, 13 samples of 8 individuals in which genetic alterations were matched at 90% or more in two independent analyses were selected in order to ensure the similar quality of samples with our cohort (Supplementary Table S1).

Transcriptome analysis

Total RNA was extracted from five gastric-clustered IM glands and three gastric-clustered non-IM glands from six patients with intramucosal gastric cancer using Sepasol-RNA I Super G (Nacalai Tesque, Kyoto, Japan). Libraries were constructed from a total of 100 ng total RNA using the TruSeq RNA Access Library Prep Kit (Illumina, San Diego, CA, USA). After the sequencing libraries were quantified and their quality was checked using the 2200 TapeStation Instrument (D1000 Assay), sequencing was performed on a NovaSeq 6000 system. The sequenced reads were aligned to the human reference genome

(GRCh38) as previously described(11). Read count data were normalized using the TCC R package (12). Hierarchical sample clustering based on the spearman correlation was performed using R. The MA plot, volcano plot, and heatmap were drawn using R. Gene Set Enrichment Analysis was performed based on the hallmark gene sets in Molecular Signatures Database v7.4 (13).

Validation of the mutations detected by targeted-capture deep sequencing using PCR-based deep sequencing

To verify the accuracy of targeted-capture deep sequencing, we randomly chose 48 of 176 mutations detected by targeted-capture deep sequencing and designed the primer sets for each using Primer3plus. The regions of interest were PCR-amplified, and 20 ng of each amplicon was pooled. The library was constructed with TruSeq Nano DNA Library Kit (Illumina) and sequencing was performed using a NovaSeq 6000 system (Illumina). All 48 mutations were confirmed by PCR-based deep sequencing with IGV (VAF > 0.02).

Sanger sequencing of formalin-fixed paraffin-embedded (FFPE) tumor samples

To investigate whether mutations detected in gastric clustered glands are present in IGC of the same patients, gDNA was extracted from macro-dissected FFPE tumor samples using the GeneRead DNA FFPE Kit (Qiagen). The regions of interest were PCR-amplified, followed by Sanger sequencing using 3730xl DNA analyzer (Thermo Fisher Scientific, Waltham, MA USA). Targeted mutations included 41 genes in Tier 1 or 2, MUC6 gene, and genes with high VAF detected in gastric clustered glands. Among them, we successfully examined 33 mutations in IGC by Sanger sequencing or WES due to the

qualification and quantification of DNA from FFPE samples (Supplementary Table S13).

Histological evaluation of atrophic gastritis

For histological evaluation, 5 μm thick sections of FFPE tissues were deparaffinized, rehydrated and stained with hematoxylin and eosin. For each patient, 5 fields were randomly selected and independently evaluated by 3 researchers (KK, TI, MT) on the basis of the Updated Sydney System visual analogue scale (0 for normal, 1 for mild, 2 for moderate, 3 for marked)(14). The average scores were defined as the atrophic score for each patient.

Mucin histochemistry

Expression of mucin was analyzed via Periodic acid Schiff-Alcian blue (PAS-AB) staining. FFPE tissue sections (5 μm) were deparaffinized, rehydrated and stained with Alcian blue for 3 min, followed by treatment with periodic acid for 3 min and with Schiff's reagent for 3 min.

We defined tissues where over 80% IM glands were complete type and incomplete type as "complete IM dominant" and "incomplete IM dominant", respectively. Other tissues with IM were defined as "mixed".

Immunohistochemistry

For immunohistochemistry studies, 5 μm thick sections of FFPE tissues were deparaffinized, rehydrated and submitted to antigen retrieval using Target Retrieval solution (Dako, S1699) in a pressure cooker. Blocking was performed using Protein Block Serum-Free (Dako, X0909) during 90 min at room temperature. The primary antibody

incubation was performed in Antibody Diluent with Background Reducing Components (Dako, S3022) overnight at 4 °C. Primary antibodies used were rabbit anti-c-Myc (1:200, ab32072; Abcam, RRID:AB_731658) and rabbit anti-SRC (1:1000, #2109; Cell Signaling Technology, RRID:AB_2106059). For detection, the sections were incubated with ImmPRESS HRP Horse Anti-Rabbit IgG Polymer Detection Kit, Peroxidase (Vector Laboratories, MP-7401), and visualized with the ImmPACT DAB Substrate Kit (Vector Laboratories). Slides were counterstained with hematoxylin.

Fluorescent in situ hybridization (FISH)

FISH analysis was performed with 3 μm-thick sections of FFPE tissues, using the ZytoLight FISH-Tissue Implementation Kit and ZytoLight SPEC MYC/CEN 8 Dual Color Probe (ZytoVision GmbH, Bremerhaven, Germany) according to the manufacturer's instruction. Fluorescence was observed with a fluorescent microscope (Olympus, Tokyo, Japan) at 1000× magnification.

Statistical analysis

We applied the Mann–Whitney *U*-test for statistical analysis unless otherwise stated. The correlation between age and the number of mutations was evaluated using Pearson's correlation coefficient model as a linear regression model, assuming zero as the intercept. To compare the number of somatic substitutions among normal single glands, single non-IM glands, and single IM glands, we used the Kruskal–Wallis nonparametric test and Steel–Dwass test. We used Fisher's exact test for categorical pairwise comparison.

Supplementary References

1. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **2014**;513:202–9.

2. Rokutan H, Abe H, Nakamura H, Ushiku T, Arakawa E, Hosoda F, et al. Initial and crucial genetic events in intestinal-type gastric intramucosal neoplasia. *J Pathol* **2019**;247:494–504.

3. Wang K, Yuen ST, Xu J, Lee SP, Yan HHN, Shi ST, et al. Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat Genet* **2014**;46:573–82.

4. Huang KK, Ramnarayanan K, Zhu F, Srivastava S, Xu C, Tan ALK, et al. Genomic and Epigenomic Profiling of High-Risk Intestinal Metaplasia Reveals Molecular Determinants of Progression to Gastric Cancer. *Cancer Cell* **2018**;33:137-150.e5.

5. Kakiuchi N, Yoshida K, Uchino M, Kihara T, Akaki K, Inoue Y, et al. Frequent mutations that converge on the NFKBIZ pathway in ulcerative colitis. *Nature* **2020**;577:260–5.

6. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **2013**;14:178–92.

7. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **2010**;38:e164.

8. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med* **2018**;10.

9. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, et al. Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **2018**;173:1823.

10. Nikaido M, Kakiuchi N, Miyamoto S, Hirano T, Takeuchi Y, Funakoshi T, et al. Indolent feature of Helicobacter pylori-uninfected intramucosal signet ring cell carcinomas with CDH1 mutations. *Gastric Cancer* **2021**;24:1102–14.

11. Shiraishi Y, Fujimoto A, Furuta M, Tanaka H, Chiba K-I, Boroevich KA, et al. Integrated analysis of whole genome and transcriptome sequencing reveals diverse transcriptomic aberrations driven by somatic genomic changes in liver cancers. *PLoS One* **2014**;9:e114263.

12. Sun J, Nishiyama T, Shimizu K, Kadota K. TCC: an R package for comparing tag count data with robust normalization strategies. *BMC Bioinformatics* **2013**;14:219.

13. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **2005**;102:15545–50.

14. Dixon MF, Genta RM, Yardley JH, Correa P. Classification and grading of gastritis. The updated Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994. *Am J Surg Pathol* **1996**;20:1161–81.

Supplementary Figure Legends

Supplementary Figure S1: Sample collection and isolation of single and clustered gastric glands

Flow chart of sample collection. MSS tumors for which the tumor purity was more than 0.4 were included for the mutation analysis of TCGA-STAD database. CN, copy number; CNV, copy number variation; IGC, intramucosal gastric cancer; MSI, microsatellite instability; MSS, microsatellite stable; RNAseq, RNA sequencing; TDS, Target capture deep sequencing; WES, whole-exome sequencing; WGA, whole genome amplification.

Supplementary Figure S2: Quality check of mutation calling with whole genome amplification

gDNA extracted from each gastric single gland was split into two aliquots, processed by WGA and WES independently, and compared. WES, whole-exome sequencing; WGA, whole genome amplification.

Supplementary Figure S3: Distribution of variant allele frequencies

(a) Variant allele frequency (VAF) distribution of mutation detected by two independent analyses per single gland. The results of C01IM1 and C01IM2 are shown. (b) VAF distribution of mutation matched by two independent analyses per single gland. The peak of the distribution was observed around the VAF of 0.50. (c) Correlation of VAF of mutations matched by two independent analyses. Red dotted line represents a VAF of 0.25. WGA, whole genome amplification.

Supplementary Figure S4: The number of mutations in gastric single gland

(a) Number of synonymous mutations in single IM glands (orange), non-IM glands (green), and normal glands (blue), respectively. Linear regression lines are shown with slope and R^2. (b) Number of mutations in single glands obtained from inflamed gastric mucosa with or without IGC. IGC, intramucosal gastric cancer. (c) Relationship between mutation number of single glands and grades of gastric atrophy.

Supplementary Figure S5: CNAs in gastric single glands

CNAs in normal single glands, single IM glands, and single non-IM glands. The upper

panel represents CN and the lower panel represents allelic ratio. CN, copy number; CNAs, copy number aberrations.

Supplementary Figure S6: Mutational signatures of gastric single gland and IGC

Relative contribution of the COSMIC mutational signature of the WES of IGC, single IM gland, single non-IM gland, and single normal gland. IGC, intramucosal gastric carcinoma; WES, whole-exome sequencing

Supplementary Figure S7: Transcriptome analysis of eight clustered glands

(a) Principal component analysis of clustered glands. (b) M-A plot of expressed genes. 548 genes with FDR<0.2 are colored. G1 and G2 are clustered non-IM and IM, respectively. (c) Heatmap of unsupervised hierarchical clustering based on 1000 expressed genes with the lowest q-value. (d) Volcano plot of transcriptome. Log2FC<−2 or >2, P < 10e-6. (e) Gene Set Enrichment Analysis based on the hallmark gene sets in the Molecular Signature Database. Enrichment plots of DNA repair, G2/M checkpoint and mitotic spindle were provided. NES, Normalized enrichment score.

Supplementary Figure S8: List of frequently mutated genes detected in more than 2 cases by WES of gastric single glands

High resolution image of Figure 4a.

Supplementary Figure S9: List of mutated cancer-related genes detected in gastric single glands by WES

High resolution image of Figure 4b.

Supplementary Figure S10: The relationship between number of mutations and IM subtypes

(a) Subtyping of IM by PAS-AB staining. Representative images of complete IM dominant (left), mixed (middle), and incomplete IM dominant (right) are shown. (b) Total number of mutations in “complete IM dominant” and “incomplete IM dominant”. (c) Number of mutations on cancer-related genes in “complete IM dominant” and “incomplete IM dominant”. (d) Number of nonsynonymous mutations on cancer related genes in “complete IM dominant” and “incomplete IM dominant”. PAS, periodic acid Schiff; AB, Alcian blue

Supplementary Figure S11: Clonal ordering analysis using somatic mutations detected in gastric single glands and IGC

Details of sample collection and the relationship of mutations between samples are shown. Each case represents the narrow-band image obtained by endoscopy, macroscopic and Alcian blue-stained stereomicroscopic images of resected specimens, the scheme of sampling locations, the Venn diagram, and the clonal ordering of somatic mutations and CNAs detected by WES of gastric single glands. Tissue samples including the border between IM and non-IM regions in the non-cancerous area were obtained from the black circle. The tumors are shown as the red dotted line in the endoscopic images. CNAs, copy number aberrations; WES, whole-exome sequencing.

Supplementary Figure S12: Relationship between shared mutations and distance between two glands using multiple single gland analysis

Relationship between the rate of shared mutations and distance between two glands in each case are shown. The rate of shared mutations was calculated by dividing the number of common mutations by the total number of mutations.

Supplementary Figure S13: Copy number aberrations in gastric clustered glands

CNAs in clustered non-IM glands and clustered IM glands. The upper panel represents CN and the lower panel represents allelic ratio. CN, copy number; CNAs, copy number aberrations.

Supplementary Figure S14: Copy number scores in clustered IM glands with subtypes

Copy number scores in clustered glands with “complete IM dominant” “mixed” and “incomplete IM dominant”. BCS, broad copy number score; FCS, focal copy number score.

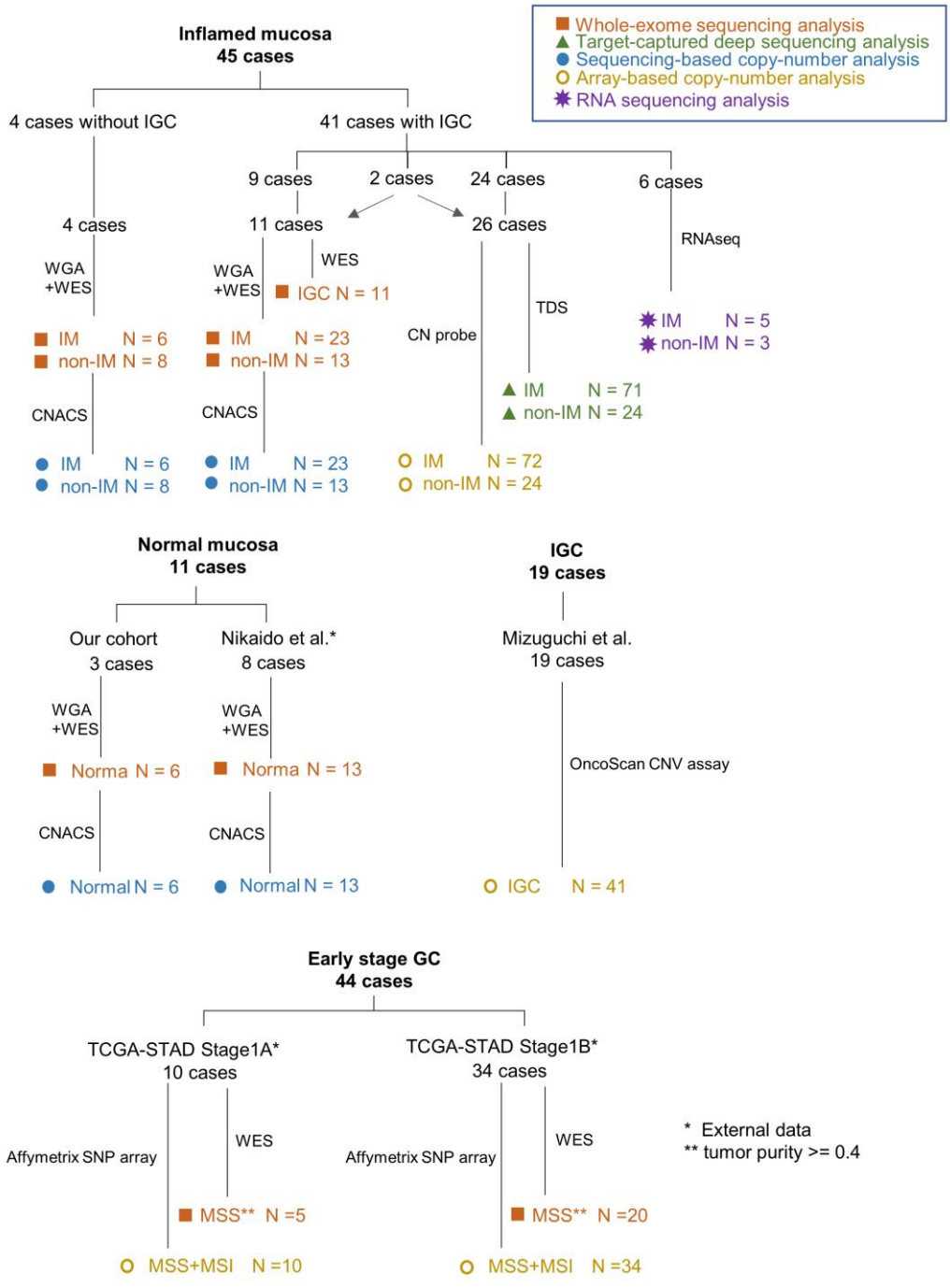
Supplementary Figure S15: Immunohistochemical analysis in gastric IM and non-IM glands

Immunohistochemistry for MYC and SRC proteins in representative cases are shown. MYC was expressed in the cytoplasm of IM glands, not in non-IM glands. SRC expression was found in both IM and non-IM glands, but was relatively high in IM glands, especially in the apical membrane. The black square represents the area of the lower panel. Scale bar represents 100 μm.

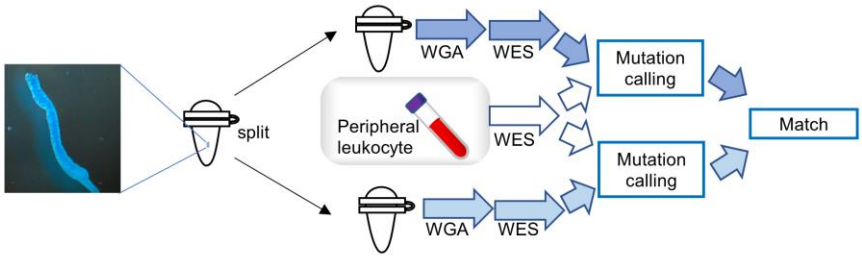
Supplementary Figure S16: Fluorescence in situ hybridization

(a) Fluorescence in situ hybridization analysis of IM and IGCs using MYC (red) and centromere of chromosome 8 (green) dual color probe at 1000× magnification and immunohistochemical analysis of MYC are shown. The red dotted line represents the border between IM (left) and IGC (right). Three dots of MYC or chromosome 8 represent their amplification. The representative cases (C10 and C11) are shown. Scale bar represents 100 μm. (b) Copy number aberration analysis based on the WES data of IGCs with the CNACS pipeline. IGC, intramucosal gastric cancer; WES, whole-exome sequencing.

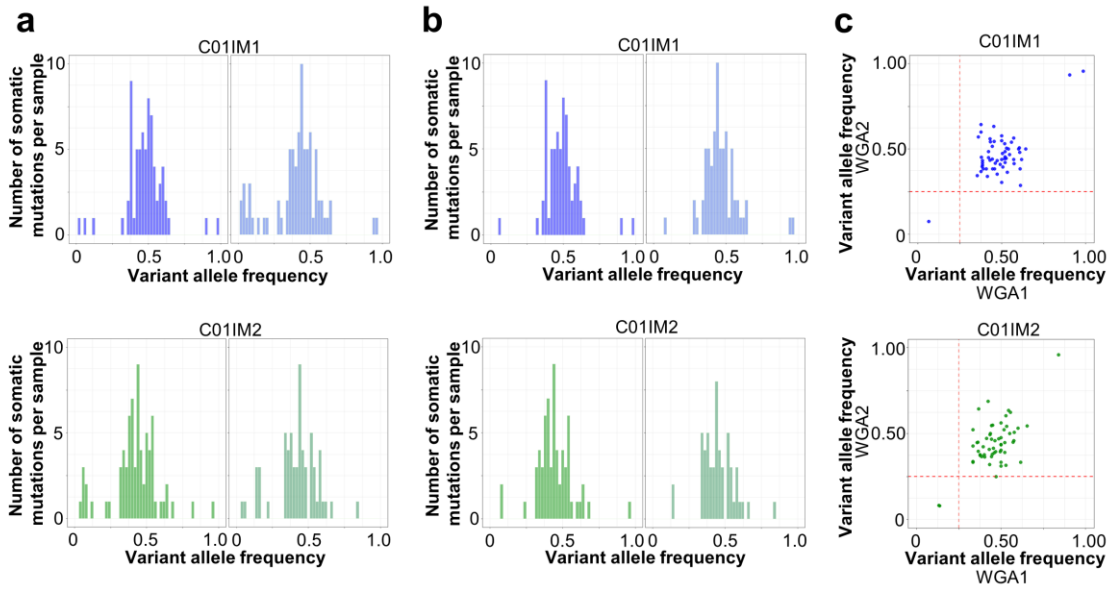
Supplementary Figure S1



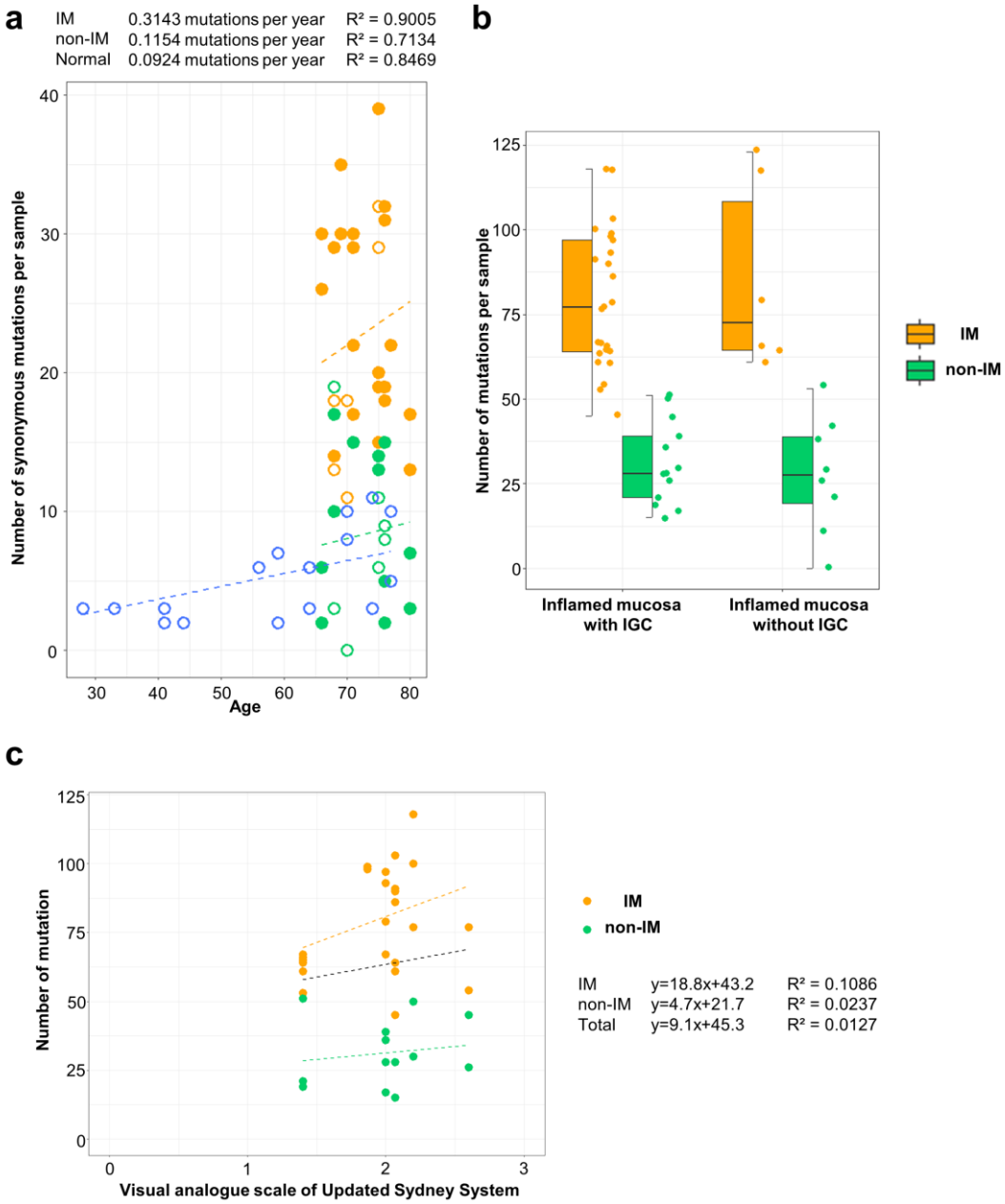
Supplementary Figure S2



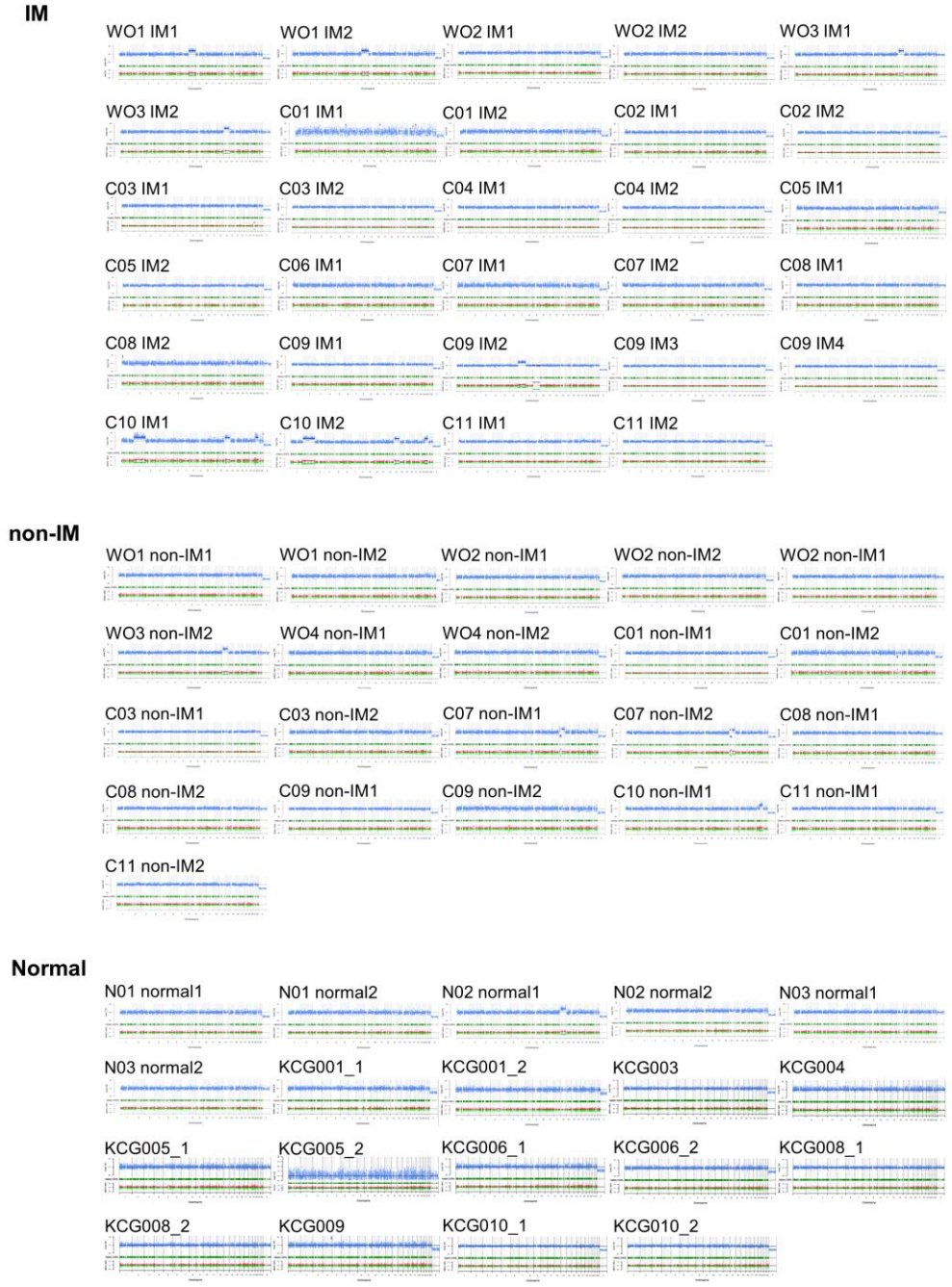
Supplementary Figure S3



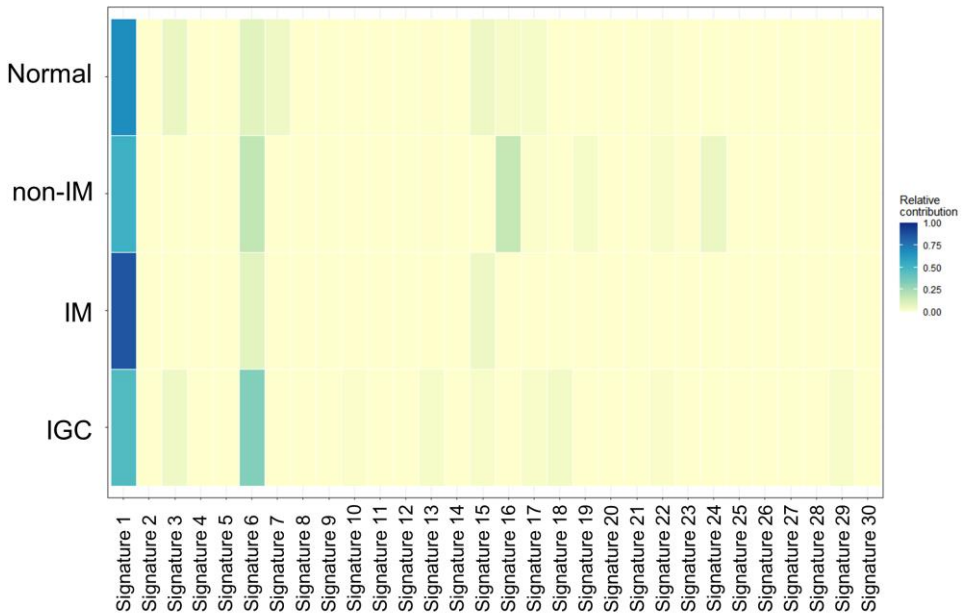
Supplementary Figure S4



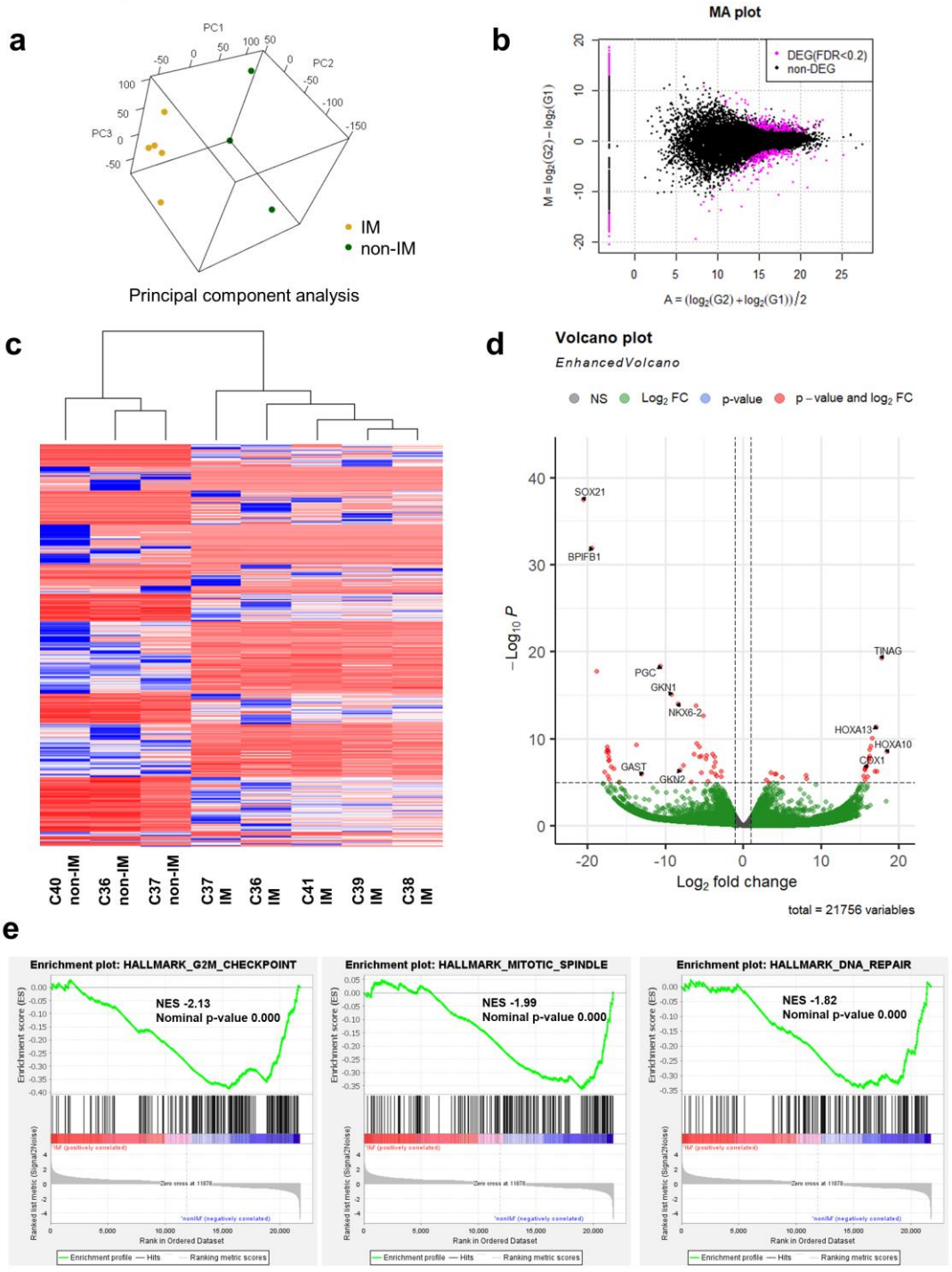
Supplementary Figure S5



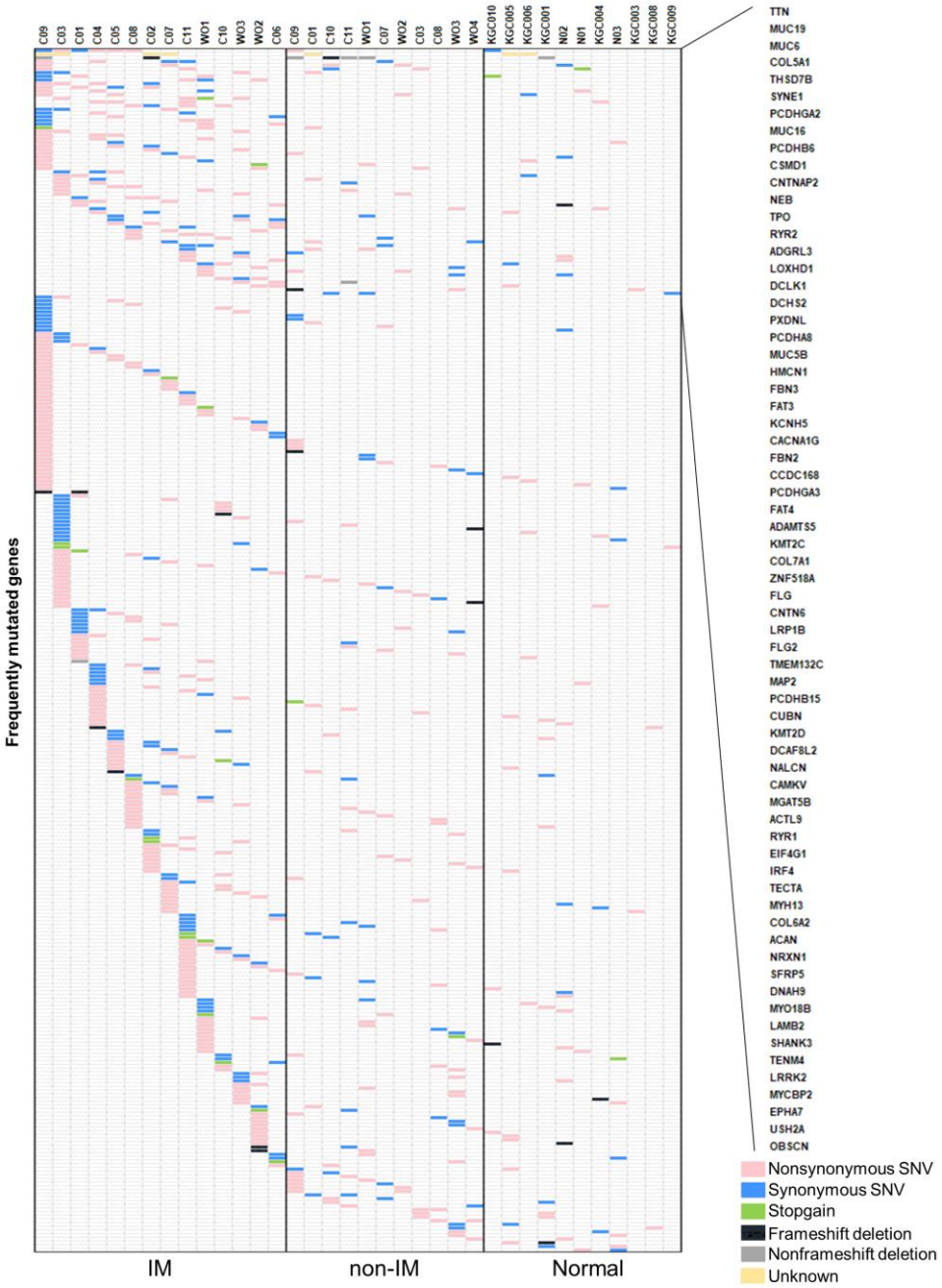
Supplementary Figure S6



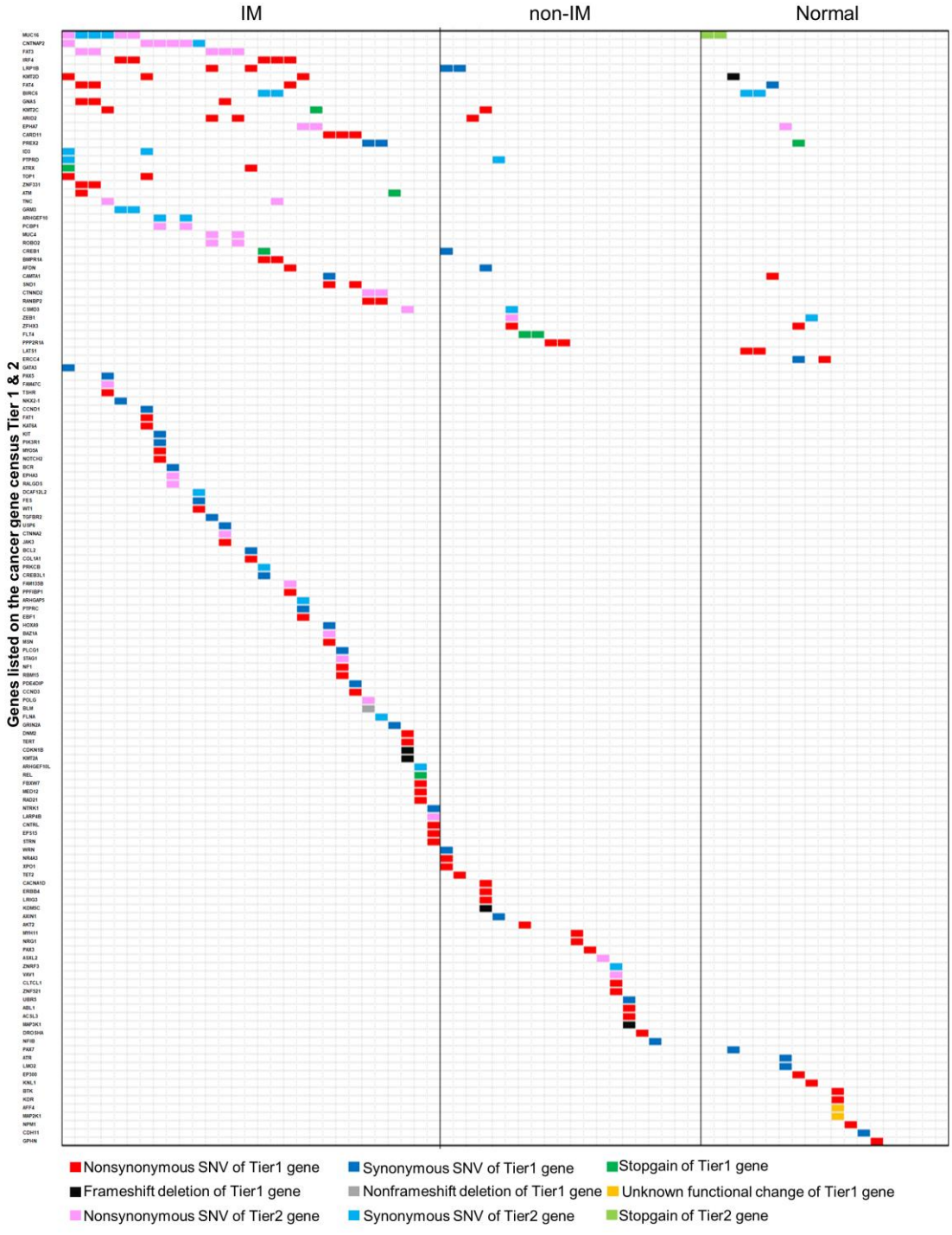
Supplementary Figure S7



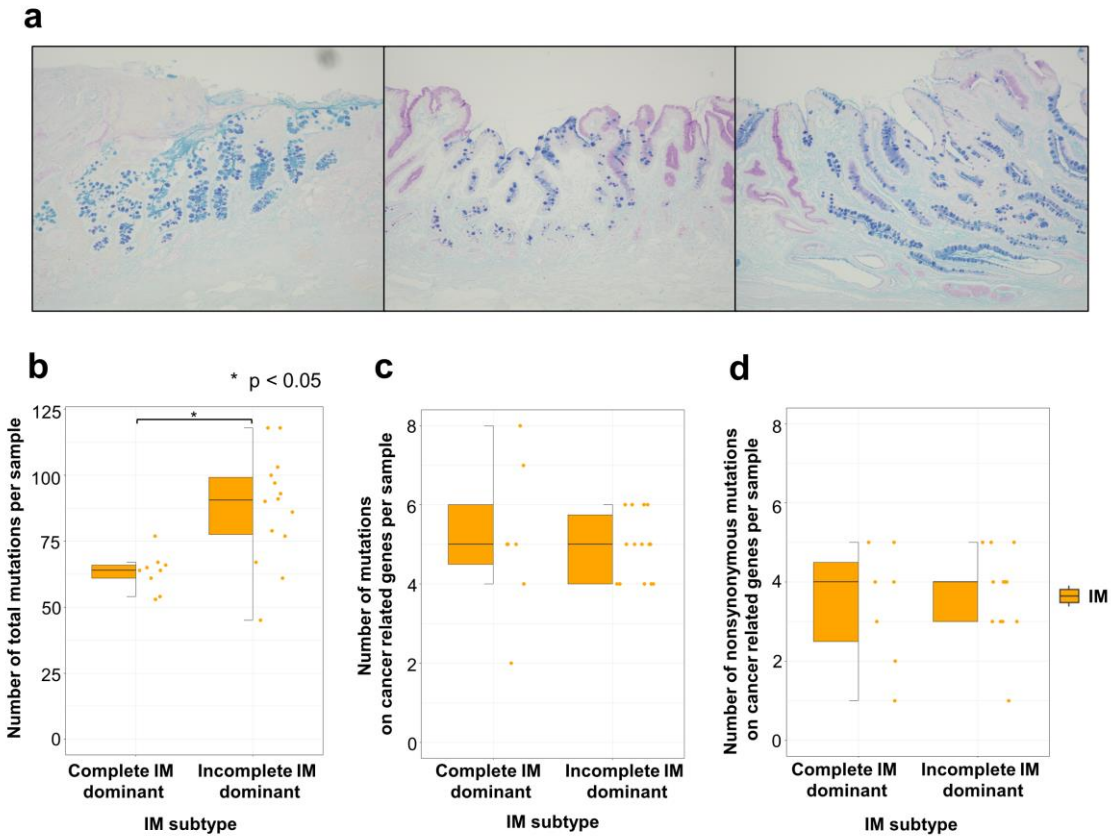
Supplementary Figure S8



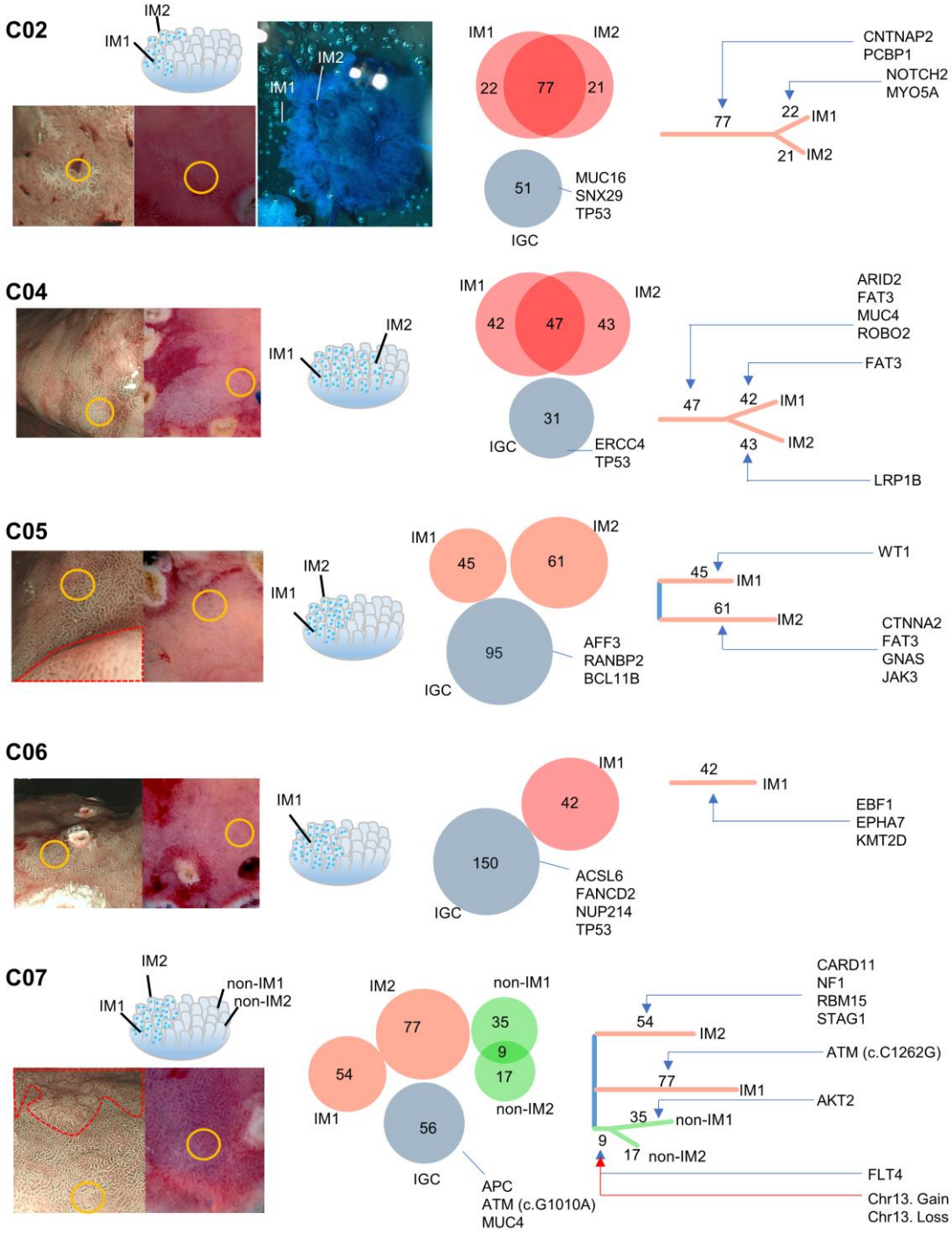
Supplementary Figure S9



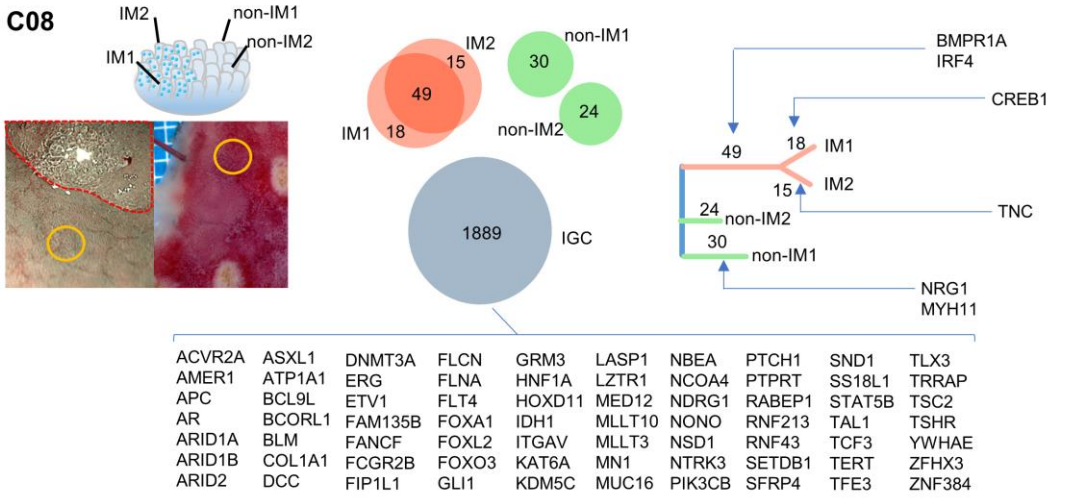
Supplementary Figure S10

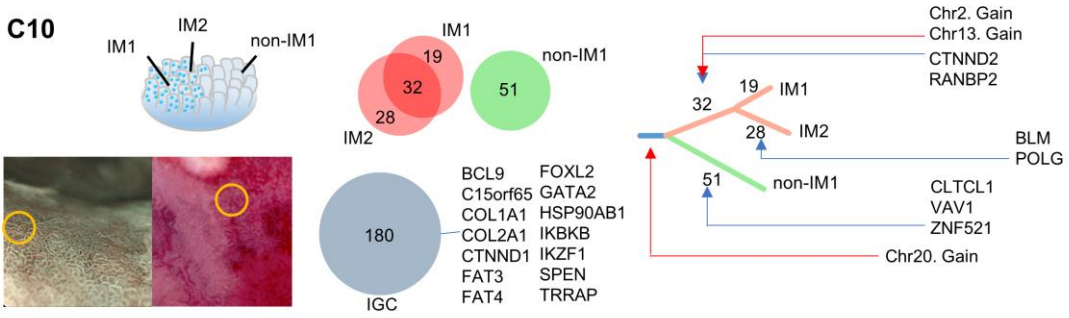


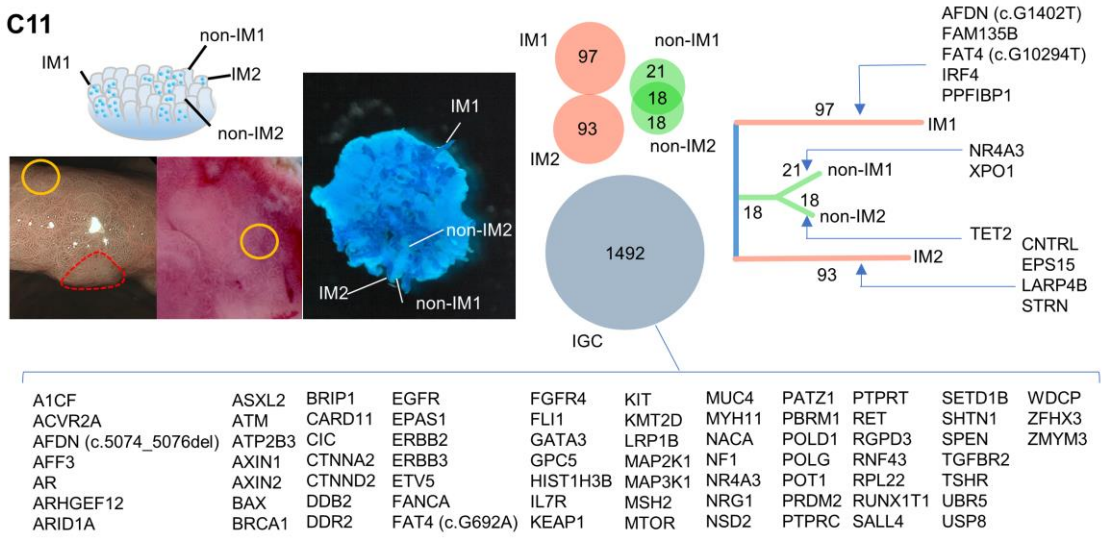
Supplementary Figure S11



Supplementary Figure S11 (Cont.)

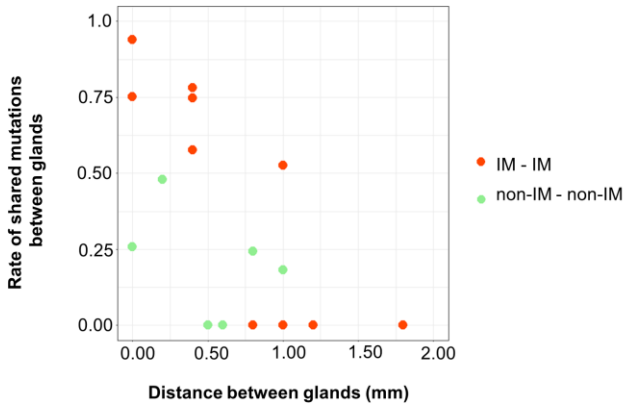






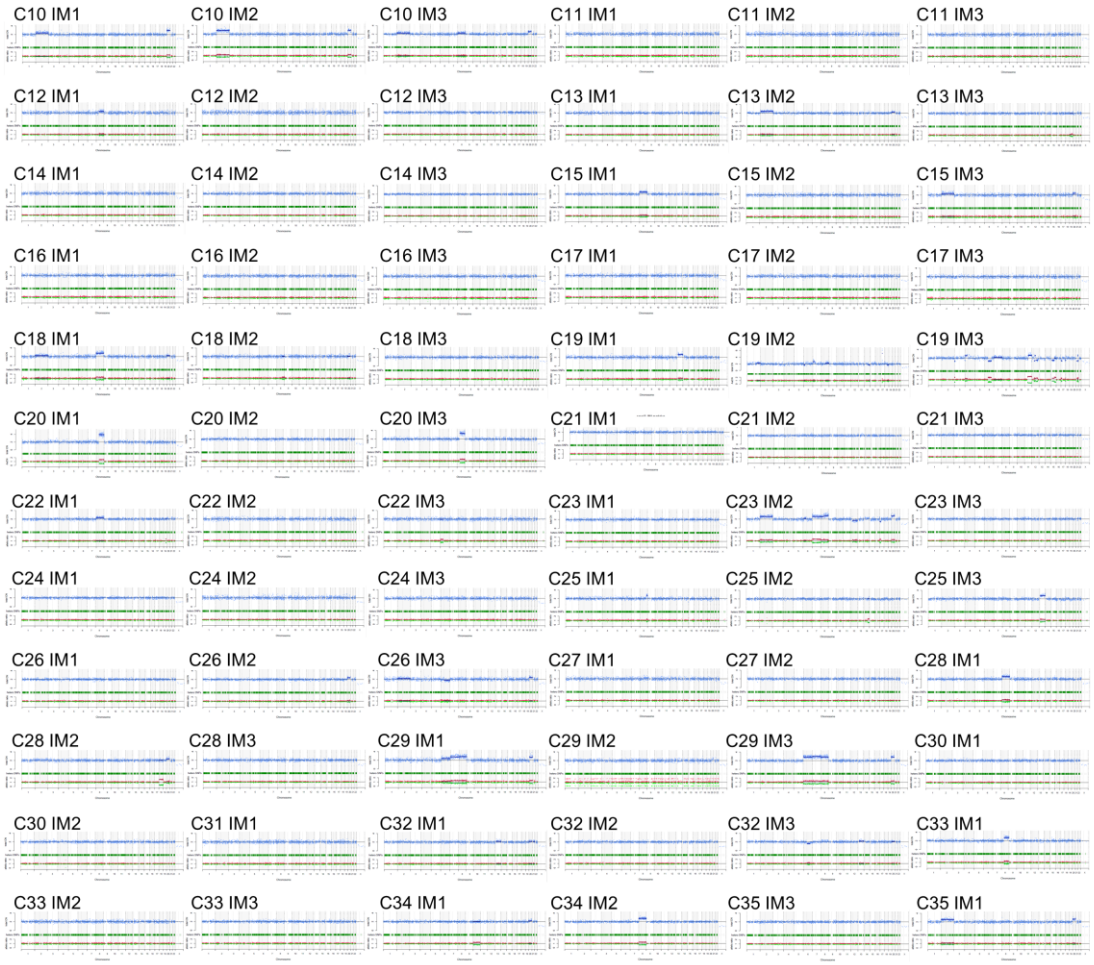
A1CF ASXL2 BRIP1 EGFR FGFR4 KIT MUC4 PATZ1 PTPRT SETD1B WDCP
ACVR2A ATM CARD11 EPAS1 FLI1 KMT2D MYH11 PBRM1 RET SHTN1 ZFHX3
AFDN (c.5074_5076del) ATP2B3 CIC ERBB2 GATA3 LRP1B NACA POLD1 RGPD3 SPEN ZMYM3
AFF3 AXIN1 CTNNA2 ERBB3 GPC5 MAP2K1 NF1 POLG RNF43 TGFBR2
AR AXIN2 CTNND2 ETV5 HIST1H3B MAP3K1 NR4A3 POT1 RPL22 TSHR
ARHGEF12 BAX DDB2 FANCA IL7R MSH2 NRG1 PRDM2 RUNX1T1 UBR5
ARID1A BRCA1 DDR2 FAT4 (c.G692A) KEAP1 MTOR NSD2 PTPRC SALL4 USP8

Supplementary Figure S12

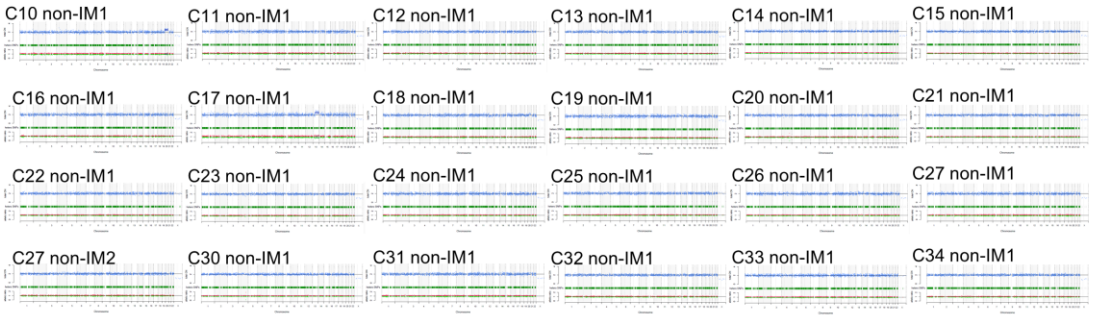


Supplementary Figure S13

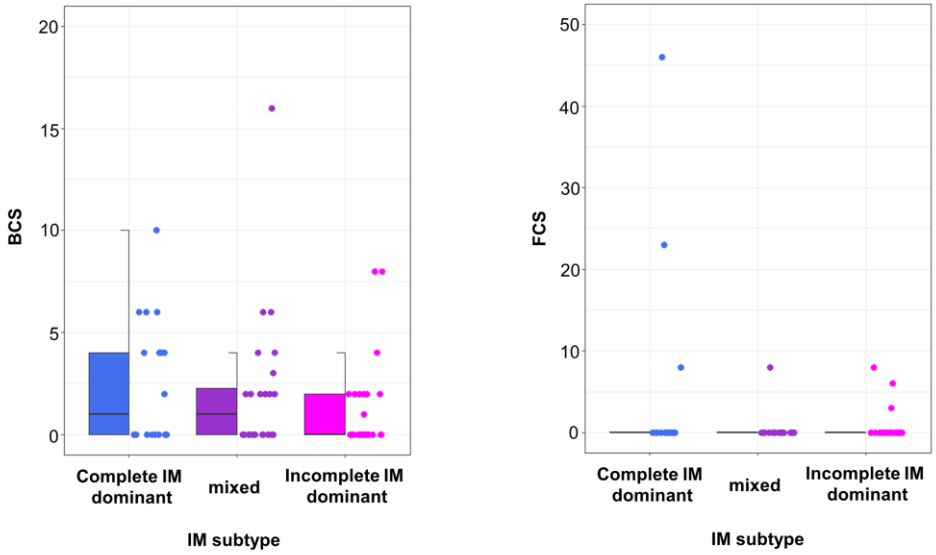
IM



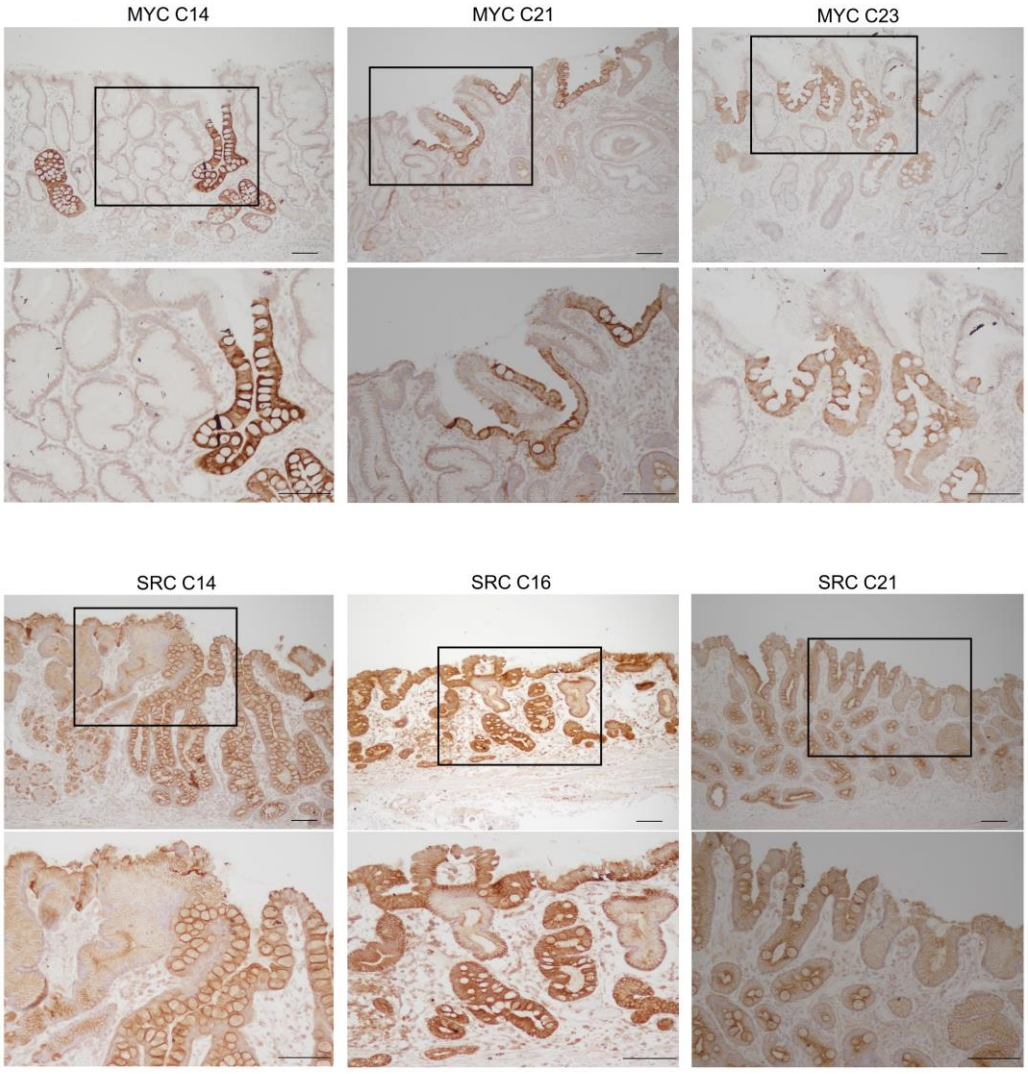
non-IM



Supplementary Figure S14



Supplementary Figure S15



Supplementary Figure S16