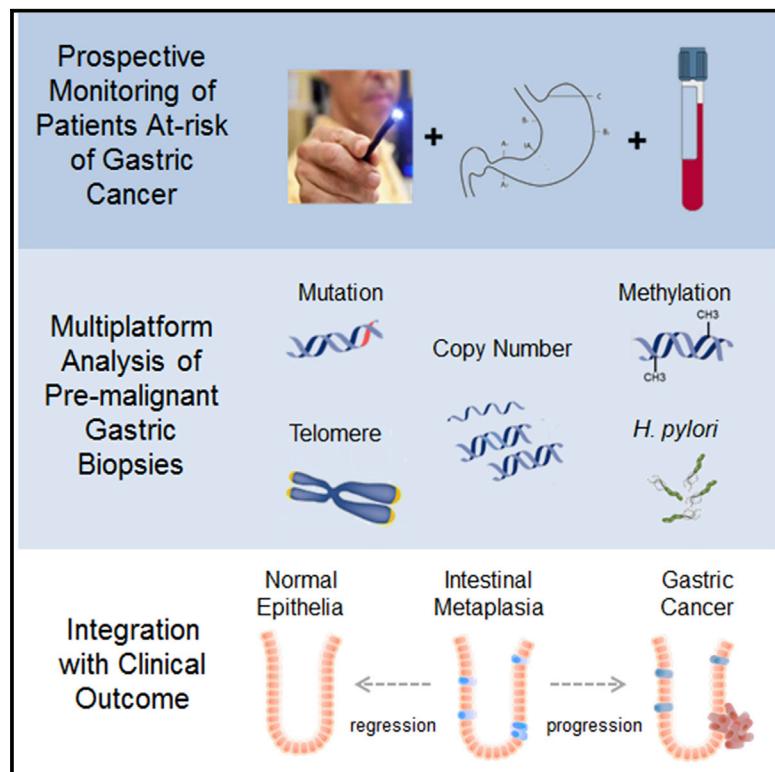


Cancer Cell

Genomic and Epigenomic Profiling of High-Risk Intestinal Metaplasia Reveals Molecular Determinants of Progression to Gastric Cancer

Graphical Abstract



Authors

Kie Kyon Huang,
Kalpana Ramnarayanan, Feng Zhu, ...,
Ming Teh, Khay Guan Yeoh, Patrick Tan

Correspondence

pattehm@nus.edu.sg (M.T.),
mdcykg@nus.edu.sg (K.G.Y.),
gmstanp@duke-nus.edu.sg (P.T.)

In Brief

Huang et al. perform molecular profiling of 138 intestinal metaplasias (IMs) from 148 gastric cancer (GC)-free patients and show that IMs with shortened telomeres and chromosomal alterations are associated with subsequent dysplasia or GC, whereas IMs with normal-like epigenomic patterns are associated with regression.

Highlights

- Genomic profiling reveals that IMs exhibit low mutational burdens compared with GCs
- Some IMs have *FBXW7* mutations, chromosome 8q amplifications, or shortened telomeres
- Sequencing detects more IM patients with active *H. pylori* infection than histology
- (Epi)genomic alterations in IM predict subsequent GC progression or regression

Genomic and Epigenomic Profiling of High-Risk Intestinal Metaplasia Reveals Molecular Determinants of Progression to Gastric Cancer

Kie Kyon Huang,^{1,23} Kalpana Ramnarayanan,^{1,23} Feng Zhu,^{2,23} Supriya Srivastava,^{2,3,23} Chang Xu,^{1,4} Angie Lay Keng Tan,¹ Minghui Lee,¹ Suting Tay,¹ Kakoli Das,¹ Manjie Xing,^{1,5,6} Aliya Fatehullah,⁷ Syed Muhammad Fahmy Alkaff,⁸ Tony Kiat Hon Lim,⁸ Jonathan Lee,⁹ Khek Yu Ho,^{2,9} Steven George Rozen,¹

(Author list continued on next page)

¹Program in Cancer and Stem Cell Biology, Duke-NUS Medical School, Singapore 169857, Singapore

²Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119228, Singapore

³Department of Pathology, National University of Singapore, Singapore 119228, Singapore

⁴Cancer Science Institute of Singapore, National University of Singapore, Singapore 117599, Singapore

⁵NUS Graduate School for Integrative Sciences and Engineering, Singapore 117456, Singapore

⁶Cancer Therapeutics and Stratified Oncology, Genome Institute of Singapore, Singapore 138672, Singapore

⁷Institute of Medical Biology, A*STAR, Singapore 138648, Singapore

⁸Department of Anatomical Pathology, Singapore General Hospital, Singapore 169608, Singapore

⁹Department of Gastroenterology and Hepatology, National University Health System, Singapore 119074, Singapore

¹⁰Centre for Regenerative Medicine, Edinburgh EH16 4UU, UK

¹¹Department of Gastroenterology and Hepatology, Tan Tock Seng Hospital, Singapore 308433, Singapore

¹²Department of Gastroenterology and Hepatology, Singapore General Hospital, Singapore 169854, Singapore

¹³Department of Gastroenterology and Hepatology, Changi General Hospital, Singapore 529889, Singapore

(Affiliations continued on next page)

SUMMARY

Intestinal metaplasia (IM) is a pre-malignant condition of the gastric mucosa associated with increased gastric cancer (GC) risk. We performed (epi)genomic profiling of 138 IMs from 148 cancer-free patients, recruited through a 10-year prospective study. Compared with GCs, IMs exhibit low mutational burdens, recurrent mutations in certain tumor suppressors (*FBXW7*) but not others (*TP53*, *ARID1A*), chromosome 8q amplification, and shortened telomeres. Sequencing identified more IM patients with active *Helicobacter pylori* infection compared with histopathology (11%–27%). Several IMs exhibited hypermethylation at DNA methylation valleys; however, IMs generally lack intragenic hypomethylation signatures of advanced malignancy. IM patients with shortened telomeres and chromosomal alterations were associated with subsequent dysplasia or GC; conversely patients exhibiting normal-like epigenomic patterns were associated with regression.

INTRODUCTION

Gastric cancer (GC) is a leading cause of global cancer mortality (Ferlay et al., 2015). Intestinal-type gastric adenocarcinoma, the most common form of GC, has been proposed to develop via

defined histological states, starting from gastritis and progressing to intestinal metaplasia (IM), dysplasia, and cancer (Tan and Yeoh, 2015). Among these, IM is a recognized precancerous condition, defined as the replacement of gastric mucosa by epithelium-resembling intestinal morphology. Patients with IM

Significance

IM is commonly diagnosed in patients; however, only a small proportion of IM patients will eventually develop GC. Here, by linking genomic and epigenomic features of IMs from cancer-free patients to subsequent clinical outcomes revealed by longitudinal surveillance, we identified molecular patterns predictive of IM clinical progression, regression, and disease stability. When benchmarked against gold-standard histopathology, genomic analysis also identified additional IM patients with active *H. pylori* infection. These findings suggest that genomic data may prove useful in risk-stratifying IM patients, by identifying individuals at high GC risk who may benefit from targeted screening, and also additional patients who could receive anti-*H. pylori* therapy. These results thus enable a molecular framework for the precision prevention of GC.

Bin Tean Teh,¹ Nick Barker,^{7,10} Chung King Chia,¹¹ Christopher Khor,¹² Choon Jin Ooi,¹² Kwong Ming Fock,¹³ Jimmy So,¹⁴ Wee Chian Lim,¹¹ Khoon Lin Ling,¹² Tiing Leong Ang,¹³ Andrew Wong,¹⁵ Jaideepraj Rao,¹⁶ Andrea Rajnakova,¹⁷ Lee Guan Lim,¹⁸ Wai Ming Yap,¹⁹ Ming Teh,^{3,*} Khay Guan Yeoh,^{2,9,22,*} and Patrick Tan^{1,4,6,20,21,22,24,*}

¹⁴Department of Surgery, National University of Singapore, Singapore 119228, Singapore

¹⁵Department of Surgery, Changi General Hospital, Singapore 529889, Singapore

¹⁶Department of Surgery, Tan Tock Seng Hospital, Singapore 308433, Singapore

¹⁷Mt. Elizabeth Medical Centre, Singapore 228510, Singapore

¹⁸Raffles Hospital, Singapore 188770, Singapore

¹⁹Department of Pathology, Tan Tock Seng Hospital, Singapore 308433, Singapore

²⁰SingHealth/Duke-NUS Institute of Precision Medicine, National Heart Centre Singapore, Singapore 169856, Singapore

²¹Cellular and Molecular Research, National Cancer Centre, Singapore 169610, Singapore

²²Singapore Gastric Cancer Consortium, Singapore 119074, Singapore

²³These authors contributed equally

²⁴Lead Contact

*Correspondence: pattehm@nus.edu.sg (M.T.), mdcykg@nus.edu.sg (K.G.Y.), gmstanp@duke-nus.edu.sg (P.T.)

<https://doi.org/10.1016/j.ccr.2017.11.018>

are at increased GC risk, with the estimated annual risk of GC in IM patients being 0.13%–0.25% per year (de Vries et al., 2008).

Helicobacter pylori (Hp) infection is pivotal to both IM and GC development (Correa et al., 2010). Hp preferentially colonizes the distal gastric antrum, causing inflammation, disruption of cellular junctions, gene mutations, and aberrant DNA methylation (Niwa et al., 2010; Shimizu et al., 2014). Current clinical guidelines recommend that all IM patients should undergo screening for active Hp infection (Dinis-Ribeiro et al., 2012), as Hp antibiotic eradication reverses early histologic changes in patients with gastritis, and may slow the progression of IM to GC (Lee et al., 2016b). Assessment of Hp is conventionally achieved by histologic evaluation of gastric biopsies; however, disadvantages of histology include inter-observer variability, sampling errors, and decreased sensitivity when IM is concurrently present (Lee and Kim, 2015).

Previous studies have proposed IM classifications based on histology, dividing IMs into “low-risk” complete (type I) and “high-risk” incomplete (type II and type III) subtypes (Gonzalez et al., 2016). However, associations between histologic IM subtypes with GC risk are not universally accepted (Dinis-Ribeiro et al., 2012), and earlier molecular studies investigating IM have focused on single molecular levels (Busuttil et al., 2014; Kang, 2012; Tahara et al., 2016), largely analyzing IMs from patients with concurrent GC (i.e., patients with confirmed cancer). As the median time of progression from IM to GC is estimated at 6.1 years (Li et al., 2016a), it is possible that these earlier studies may not reflect the true clinical scenario where most diagnosed IM patients are cancer free. In this study, we attempted to accurately assess the IM molecular landscape, and to identify molecular features associated with GC progression and IM regression.

RESULTS

Study Design, Samples, and Platforms

The Gastric Cancer Epidemiology Program (GCEP) (Figure S1A) is a prospective multi-center longitudinal cohort study, monitoring 2,980 Chinese subjects from gastroenterology clinics, aged 50 or above, for a minimum of 5 years. The STAR Methods introduces the overall GCEP study, including aims, study population, and patient characteristics. Clinical data of GCEP participants were

collected at baseline and annual follow-up, and gastric biopsies were collected at enrollment, year 3 and year 5 following standardized protocols through surveillance endoscopies (STAR Methods). Biopsy samples were assessed for chronic gastritis (Sydney classification), Hp positivity, gastric atrophy (GA), and IM (Figure S1B). Histological assessments were independently performed by two trained pathologists (S.S. and M.T.) using the updated Sydney classification for IM grading and the revised Vienna classification for gastric dysplasia (Dixon, 2002; Dixon et al., 1996). At study completion, a total of 11,157 person-years of surveillance were performed with a compliance rate of 85%. About half of the subjects (52%) had IM at baseline, classified as mild, moderate, or marked. We observed longitudinal development of 13 high grade dysplasias, 1 intramucosal carcinoma, and 7 early GCs- based on pre-specified endpoints, these were collectively classified as early gastric neoplasias (EGNs).

Both IM and GAs are associated with increased GC risk (Lee et al., 2016a). We focused on IMs as IM patients have a higher risk of developing GC than GA patients (2.5 per 1,000 person-year for IM versus 1–1.73 per 1,000 person-year for GA) (de Vries et al., 2008; Lee et al., 2016a), and because IM is readily recognized by the presence of goblet and colonocyte-resembling cells that do not appear in normal gastric mucosa, with good inter-observer agreement (den Hoed et al., 2013).

To determine genetic and epigenetic changes in high-risk IM, we selected 203 gastric samples from 148 GCEP subjects reflecting different disease stages (Figure S1C). Clinical parameters of the GCEP cohort used in this study are reported in Table S1 and representative gastric biopsy images in Figure S2. The majority of samples ($n = 116$; 83 patients) were high-risk IMs, being moderate or marked IM, of which 77 were type II/III ($n = 69$ and 8). We also analyzed 61 IMs from different anatomical regions (27 antrum, 24 body, 10 cardia) in the same patients to explore intra-patient variability. Finally, to provide references, we profiled histologically normal gastric samples (“normal”; 65 samples/43 subjects) and cases of mild IM (“mild IM”; 22 samples/22 patients). None of the mild IM subjects had GA, while 81% of the moderate-marked IM subjects had concurrent GA (67/83 subjects). Samples were processed for DNA sequencing (164 samples/135 patients), methylation profiling (191 samples/139 patients), and SNP arrays (51 samples/49 patients), depending on sample availability (Figure S1C).

Genomic Alterations in IM

We sequenced the coding exons of 766 genes previously shown to be involved in various gastrointestinal cancers (Table S2) across 164 GCEP gastric biopsies, of which 161 were matched against germline DNA sequences from patient bloods. These included 105 high-risk IMs, 22 mild IMs, and 37 gastric normal epithelia from 135 patients, and also IMs from multiple sites (antrum, body, or cardia) in 24 subjects. We generated over 44 million mappable sequence reads per sample, corresponding to an average sequencing depth of 365 \times . We processed the sequencing data using rigorous sequencing filters used in other studies (Figure S3A) (Taylor-Weiner et al., 2016), to reduce sequencing artifacts such as OxoG lesions (Costello et al., 2013). We identified 1,431 somatic single-nucleotide variants and 75 somatic insertion-deletion sequences (indels) in the gastric biopsies. Orthogonal validation of >300 predicted somatic mutations using Ion Torrent sequencing indicated a specificity of 98.4% (309/314), and despite being distinct sequencing platforms, we observed strong correlations in mutation allele frequencies (MAFs) between Illumina and Ion Torrent data (Figure S3B). Resequencing of selected samples using molecular barcoded sequencing also confirmed strong correlations between MAFs estimated with or without molecular barcodes, indicating minimal PCR amplification bias (Figure S3C; n = 184 variants across 8 IMs). Using MAFs as a marker for IM cellularity, we estimated the average IM purity in our dataset to be 22.2%. MAF levels in moderate/marked IM (estimated cellularity >30%) were significantly higher than mild IM (estimated cellularity <30%) (Figure S3D). Across individual samples, the MAFs of three key driver genes (*FBXW7*, *ARID1A*, and *TP53*; see later) were also significantly correlated with the IM cellularities of samples where these mutations were detected (Figure S3E).

Within the genomic regions analyzed, the IM mutation rate was 2.6 mutations per Mb, significantly higher than mild IM or normal gastric mucosae (0.9 and 0.4 mutations/Mb; p values 2.0×10^{-4} and 3.6×10^{-12}), but significantly lower than non-hypermutated (MSI) GC (6.9 mutations/Mb; p = 2.3×10^{-9}), the latter inferred from a GC cohort sequenced using an analogous panel at similar depth and validated using common samples profiled on both platforms (Figures 1A, S3F, and S3G). Most IM mutations were C > T transitions in the context of CpG dinucleotides, a mutation signature associated with age of diagnosis (Figure S3H). However, while somatic mutation rates were significantly correlated with age in normal gastric samples, they were not in the IM samples indicating the presence of additional mutational processes in IM (Figure S3I).

TP53 and *ARID1A* are two of the most frequently mutated tumor-suppressor genes in GC, exhibiting alterations in 22%–47% of GCs (Figure S3J) (Cancer Genome Atlas Research Network, 2014). In contrast, only two cases of *TP53* mutations (2%) and three cases of *ARID1A* mutations (3%) were detected in the IM samples, suggesting that clonal *TP53* and *ARID1A* mutations are likely infrequent in IM. In contrast, when advanced GCs were sequenced using the same protocol, *TP53* and *ARID1A* mutations were readily detected (n = 6 GCs; *TP53* mutations in 3/6 GC samples) (Figure S3K). To confirm the occurrence of clonal mutations in IM cells, we performed laser-capture microdissection (LCM) and immunohistochemistry (IHC). LCM analysis of a *TP53*-mutated IM (MAF 5.5%) coupled with validation

Sanger sequencing confirmed the presence of *TP53* mutations in LCM-purified IM cells (Figure S4A), and molecular barcoded sequencing of another seven independent LCM-processed IMs identified *TP53* and *FBXW7* mutations in two cases (Figure S4B). At the protein level, IHC analysis specifically localized p53 overexpression (a hallmark of mutated *TP53*) in IM cells and not adjacent gastric mucosae, and similarly IHC analysis of *ARID1A*-mutated IMs confirmed loss of *ARID1A* protein expression in metaplastic cells (Figure S4C). These results support the presence of genetic alterations in IM; however, we note that they do not preclude the possibility of other mutations occurring at very low MAF levels (<1%) in other Hp-infected gastric cells, detectable only by ultra-deep sequencing (Shimizu et al., 2014; see Discussion).

MutSigCV analysis identified *FBXW7* as a significantly mutated gene in 4.7% of gastric biopsies (six moderate-marked IMs, one mild IM, one normal) (Table S2). IMs harboring *FBXW7* mutations were distinct from IMs exhibiting *TP53* or *ARID1A* mutations (Figure S4D). *FBXW7* encodes an E3 ubiquitin ligase responsible for c-MYC and cyclin E1 degradation, and *FBXW7* is an established tumor suppressor in various cancer types (Welcker and Clurman, 2008). The most frequent *FBXW7* mutations are missense point mutations, causing dominant negative activity (Akhoondi et al., 2007; Davis et al., 2014). Supporting their functional significance, most of the IM *FBXW7* mutations were monoallelic and occurred at these hotspot regions (4/6; 66.7%) (Figure 1B). In GC, *FBXW7* mutations have been observed in 9.2%–18.5% of tumors (Cancer Genome Atlas Research Network, 2014) and shown to regulate apoptosis, proliferation, and therapy resistance (Li et al., 2016b). *Fbxw7*^{+/−} mice are also more prone to developing IM, dysplasia and GC after carcinogen exposure (Jiang et al., 2017). Supporting a role for *FBXW7* in genomic safeguarding, in both *FBXW7* WT (wild-type) GC cell lines and normal gastric epithelial cell lines, we observed increased γH2Ax levels (a DNA damage marker) after small interfering RNA-mediated *FBXW7* depletion compared with non-targeting control treated cells (Figure 1C). This effect was further enhanced after DNA damage by cisplatin (Figure 1D). We also established GC lines stably expressing ectopic *FBXW7* WT or *FBXW7*-mutant (R505C or R465C, both mutation hotspots). Significant increases in proliferation were observed in *FBXW7* R505C or R465C overexpressing cells compared with *FBXW7* WT overexpressing cells (Figure 1E), supporting the dominant negative nature of the IM *FBXW7* mutations. We also observed decreased γH2Ax levels in *FBXW7* WT-expressing cells compared with *FBXW7* R505C or R465C-expressing cells, particularly after cisplatin treatment (Figure 1F). These results suggest that *FBXW7* mutations in IM are likely to functionally contribute to IM and GC development. We note, however, that while *FBXW7* is a driver gene as defined by statistical (and functional) criteria, in absolute frequencies the *FBXW7* mutation rate in IM remains relatively low (4.7%) compared with in GC (9.2%–18.5%), indicating that additional genomic alterations are likely required to fully progress to GC.

We used the sequencing data to characterize somatic copy-number alterations (sCNAs) in the IM samples, validating this approach by correlating sCNAs predicted by sequencing against SNP array data of the same samples (n = 51; Figure S4E). sCNAs were observed in 12.5% of IMs, but none of the mild IM or normal

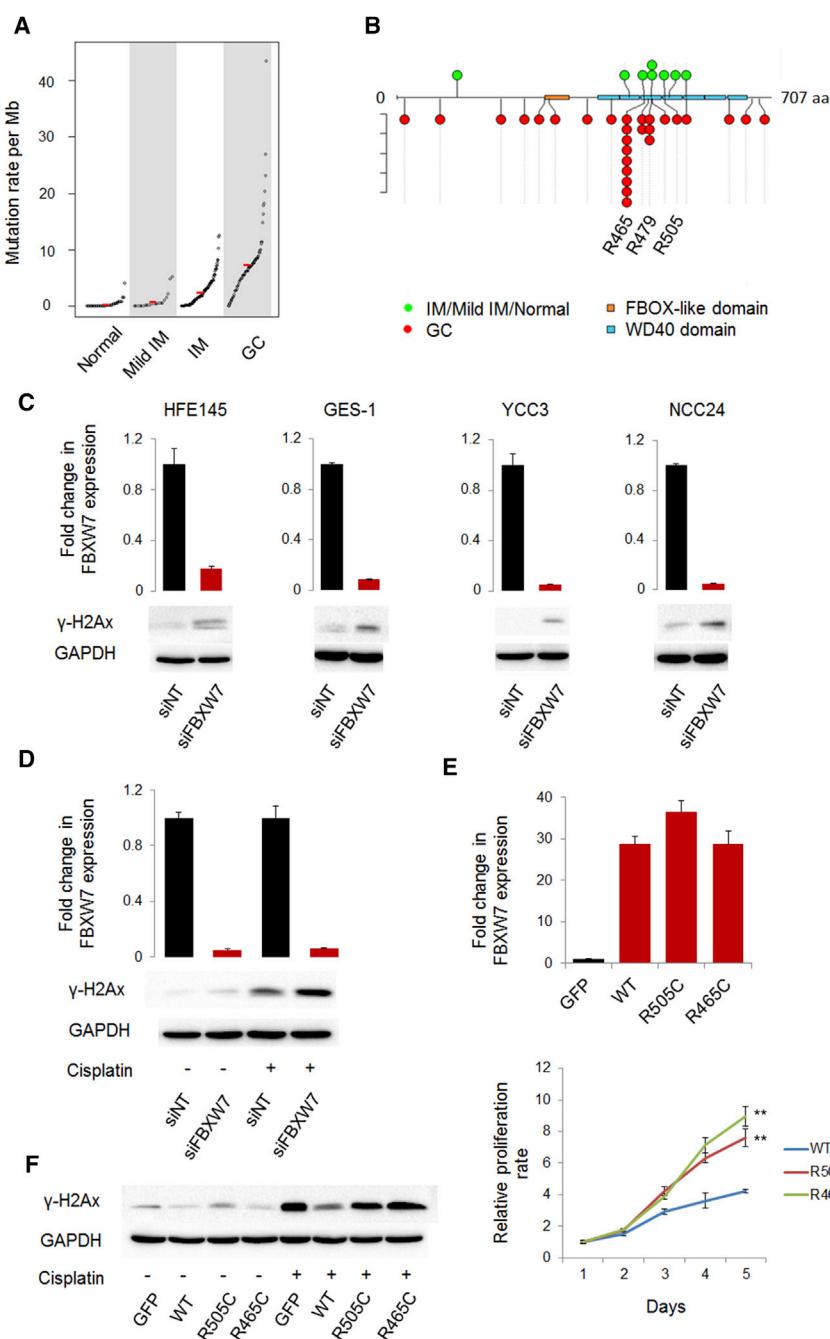


Figure 1. Genomic Alterations in IM

(A) Median mutation frequencies in each sample category are indicated by red bars. The p values were calculated using t tests.

(B) FBXW7 FBOX-like and WD40 domains (defined by UniProt) are in brown and blue. Somatic mutations specific to GC (TCGA) are depicted in red, while somatic mutations specific to pre-malignant gastric tissues (including normal, mild IM, and IM) are in green.

(C) FBXW7 mRNA expression (top) and phosphorylated γ-H2Ax (bottom) after small interfering RNA (siRNA)-mediated FBXW7 depletion. GAPDH serves as a loading control. All cell lines used are *FBXW7* WT. HFE145 and GES-1 are non-transformed gastric epithelial cells; YCC3 and NCC24 are transformed GC cells. Data are represented as mean ± SD.

(D) FBXW7 mRNA expression and phosphorylated γ-H2Ax in NCC24 cells after siRNA-mediated FBXW7 knockdown, with and without cisplatin treatment. GAPDH serves as loading control. Data are represented as mean ± SD.

(E) FBXW7 mRNA expression in *FBXW7* WT or R505C or R465C mutant stably overexpressing TMK1 GC cells (top). R505C and R465C are *FBXW7* mutation hotspots observed in IM and other cancers. Proliferation rates of *FBXW7* R505C or R465C overexpressing lines compared with *FBXW7* WT overexpressing cells (bottom); *p < 0.01; Welch t test. Data are represented as mean ± SD.

(F) Level of phosphorylated γ-H2Ax in *FBXW7* WT stably overexpressing lines compared with *FBXW7* R505C or R465C overexpressing lines and GFP control cells, with and without cisplatin treatment. GAPDH serves as a loading control.

See also Figures S1–S4; Tables S1 and S2.

We used TelSeq to estimate average telomere lengths in IMs, leveraging off-target reads (Ding et al., 2014). Telomere lengths were significantly shorter in IM samples compared with normal gastric biopsies, with IMs from the antrum exhibiting shorter telomeres than body or cardia IMs (Figure 2B). We observed a significant association between chromosome 8q amplification with telomere shortening (t test, p = 0.043). In contrast, similar telomere measurements in an independent

panel of GCs, and also in TCGA GC samples, revealed no significant changes in telomere length between paired GCs and adjacent normal samples (t test, p = 0.34 for in-house, p = 0.88 for TCGA) (Figure S4G). This suggests that telomere lengths are initially reduced in IM, but are likely subsequently restored during progression to malignancy. Supporting this interpretation, a recent study has shown that telomere lengths are eroded in early GC and lengthened in advanced GC (Mu et al., 2012).

To infer the degree of shared ancestry of IMs, we compared somatic mutations and sCNAs between IM samples at different sites from the same patients. In all 24 patients, IM samples

gastric samples (Figure 2A). The most common sCNA event was chromosome 8q amplification, observed in 9 IM samples (Figure 2A). Notably, the minimal amplified 8q locus (chr8q22.3 to chr8q24.3) included the *MYC* oncogene (chr8q24.21), and strong c-MYC expression was observed in IMs harboring 8q amplifications (n = 9) compared with IMs with normal 8q levels (n = 5; by % area; Welch's t test, p = 0.026). Similar to *TP53* and *ARID1A*, IHC analysis confirmed that c-MYC overexpression (associated with genomic amplification) was observed in IM cells and not adjacent gastric mucosa (Figure S4C). Other regions of focal amplification included *CTNNB1* (Figure S4F).

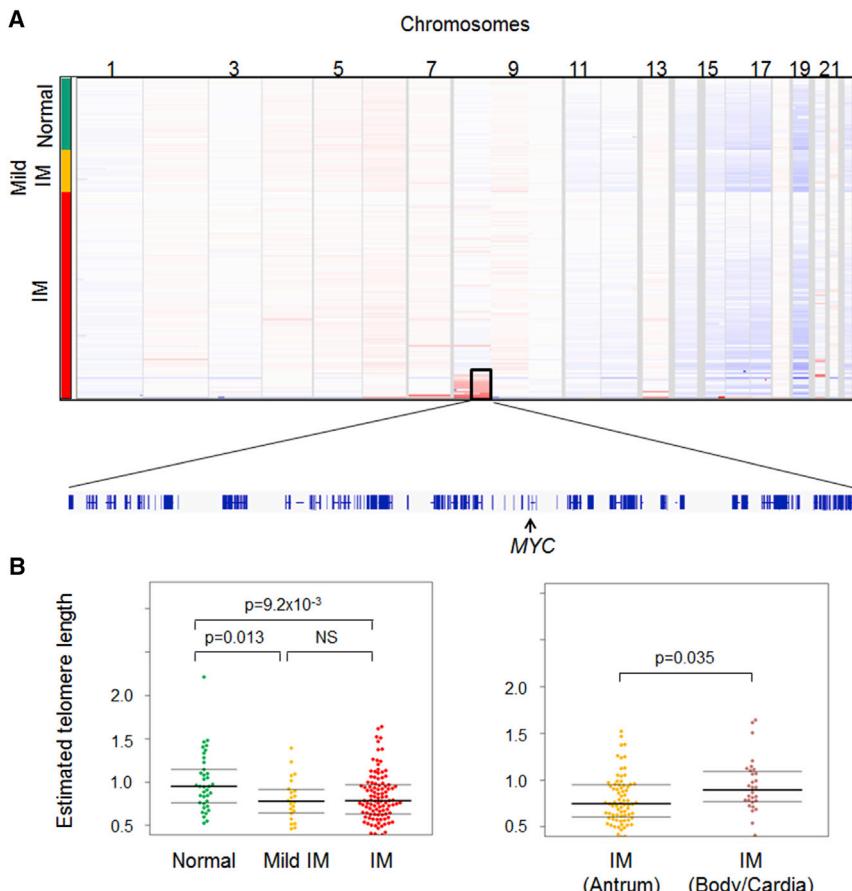


Figure 2. Copy-Number Alterations and Telomere Length in IM

(A) Heatmap representing sCNA profiles across 159 pre-malignant gastric tissues, arranged by increasing chromosomal instability. Sample categories of gastric tissues are on the left. A magnified view of the minimal amplified region on chromosome 8q and the genomic location of *MYC* is shown at the bottom (arrow).

(B) Estimated telomere lengths (normalized to median telomere length in normal gastric tissues) in gastric tissues of different sample categories (left) and IM tissues at different gastric sites (right). Thick bars indicate medians and thin bars indicate first and third quartiles. The p values were calculated using t tests.

See also Figure S4.

from different sites were clonally unrelated as no common mutations or sCNAs were detected, suggesting that IMs from the same patient are genetically dissimilar and likely associated with high intra-patient multiclonality. However, despite not showing common mutations, IMs at different stomach sites exhibited very similar mutational spectra (Figure S3H), dominated by CpG > TpG mutations with a smaller contribution from T > G (Sanger Institute Signature 17), which has been attributed to stomach acid. These findings suggest that IMs from the antrum and body/cardia may be caused by similar or related environmental exposures or biological processes. We also tested if whole-genome doubling (WGD) events can be observed in IMs from cancer-free patients. Using ABSOLUTE, we identified 1 IM (out of 104) showing evidence of WGD, in a sample also harboring a *TP53* mutation predicted to occur before the doubling event. In contrast, WGD events were observed in 39.5% of GCs (Figure S3J), suggesting that WGD events are likely rare in IMs from cancer-free patients.

Detection of Hp by Sequencing

We explored the use of sequencing as an alternative method for detecting Hp besides histology. In GCEP, Hp status was determined using two methods. First, all GCEP subjects were screened using a Hp serology test kit that has high sensitivity (92%–100%) and specificity (80%–96%) (Veijola et al., 2008), but cannot differentiate between active or previous Hp infection. Second, to diag-

nose subjects with active Hp infection, histological examination of stomach biopsies was performed following a standardized biopsy protocol (five biopsies). Histology was chosen as it is the gold standard for diagnosing active Hp infection (Calvet et al., 2009). Of the 135 subjects analyzed, all 135 were positive for Hp serology, as this was one of the selection criteria. Of these 135, 15 subjects (11%) were diagnosed with active Hp infection by histology, while the remaining 120 were diagnosed as negative for active Hp infection. Subsequent re-eradication proved successful in all 15 cases (Figure 3A).

Hp sequence reads were detected in 42% of IM samples (44 cases), a higher proportion than mild IM (14%), normal gastric mucosae (18%), or blood (0%), with IM samples also exhibiting the highest numbers of Hp reads ($p = 0.003$, Wilcoxon test) (Figure S5A). The Hp sequence reads were distributed throughout the Hp genome, covering Cag pathogenicity islands and antibiotic resistance genes (e.g., *rpsL*, *rpsE*, *rplF*, *rplV*, *rplC*, and *gyrA*) (Figures 3B and S5B). IM samples with Hp infection were more likely to harbor sCNAs (28.6%) compared with tissues with low or no Hp (8.4%) (Figure 3C), consistent with Hp playing an active role in cancer progression. Hp reads were more often detected in body or cardia IMs (63%) at higher levels compared with antral IMs (33%; Fisher's exact test $p = 8.1 \times 10^{-3}$, Wilcoxon test $p = 4.9 \times 10^{-3}$) (Figure 3D). We also performed a similar analysis for Epstein-Barr virus (EBV), another pathogen in GC (Cancer Genome Atlas Research Network, 2014) (Table S3).

Notably, all 15 subjects diagnosed with active Hp infection by histology were also positive for Hp sequence reads, indicating 100% concordance (Histo⁺/NGS⁺). In three subjects with histologically confirmed infection where multiple biopsies were available, Hp reads were observed in all biopsies, indicating that Hp detection in these individuals is not confined to a single biopsy site. In addition, sequencing of two histologically confirmed cases pre- and post-successful Hp eradication confirmed the complete absence of Hp sequence reads after eradication,

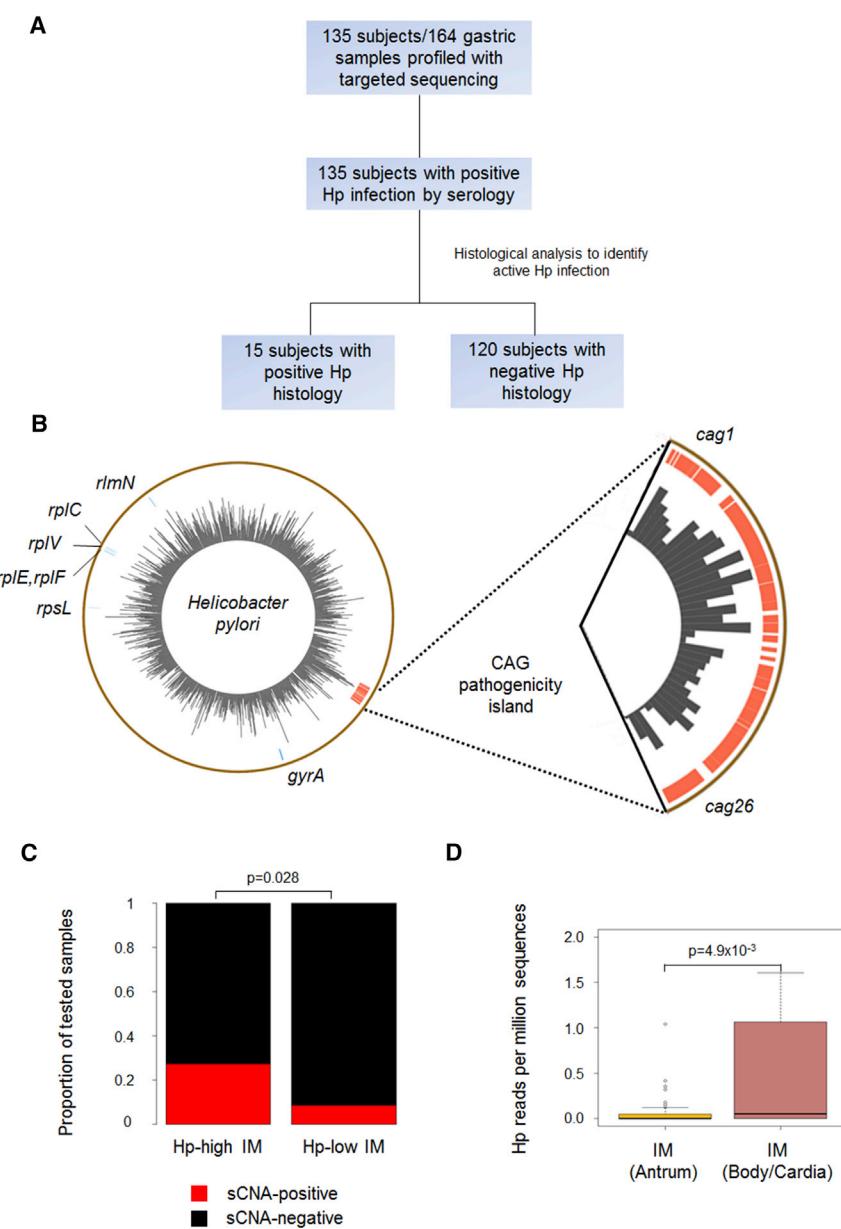


Figure 3. Genomic Detection of Hp in IM

(A) GCEP Hp prevalence.
(B) Circos plot of the Hp genome shows collective coverage depth using 1 kb windows (0× to 28×) from 22 gastric samples. Locations of Cag pathogenicity islands and antibiotic resistance genes are indicated in brown and blue. Magnified views show coverage of the Cag pathogenicity island (*cag1* to *cag26*).
(C) sCNA variation in Hp-high samples (left; n = 21) and Hp-low samples (right; n = 83). A cutoff density of 0.18 Hp sequences per million reads was used to delineate Hp-high and Hp-low samples. The p values were calculated using Fisher's exact test.
(D) Boxplots showing Hp densities (number of Hp sequences per million sequencing reads). Outlier samples with high Hp density (>2 Hp reads per million sequences) were omitted to improve visualization. Boxes indicate interquartile ranges (IQR) and center lines indicate the median. Whiskers are extended to within 1.5 IQR of the upper and lower quartiles. Data points falling outside this range are displayed independently. p Values were calculated using Wilcoxon tests.

See also Figure S5 and Table S3.

rate may increase from 11% to 14% (conservative lower limit if positivity is confined to those four patients with high Hp reads) or 11%–27% (upper limit if positivity is based on having both Hp sequence reads and subsequent Giemsa confirmation).

DNA Methylation Subtypes of IM

DNA methylation alterations, induced by environmental agents such as Hp, have been observed in GC and may precede the onset of genomic alterations (Suzuki et al., 2006). To characterize epigenetic changes in IM, we profiled 191 samples from 139 patients using Illumina methylation 450K arrays, including IMs (108), mild IMs (22), and normal gastric mucosae (61). We also included samples from multiple sites (antrum, body, and cardia) from

indicating that residual Hp DNA in the human gastric intestinal tract is likely transient (Figure S5C). Notably, of the 120 patients diagnosed by histology as negative for active Hp infection, 33 patients still exhibited Hp sequence reads (Histo[−]/NGS⁺). Even using a more stringent cutoff of 0.18 Hp reads per million sequences corresponding to the lowest Hp level found in a Hp Histo⁺/NGS⁺ sample, we still detected Hp in 4 of the 33 Histo[−]/NGS⁺ subjects (7 samples) (Figure S5B). To validate our results, we performed Hp reassessment using modified Giemsa staining, a specialized method used to detect adherent bacteria. Giemsa staining confirmed the presence of focal Hp infection in 22 out of the 33 subjects (25/37 IM samples) (Figure S5D). These results suggest that in the context of IM, sequencing may identify additional patients with low-level but nevertheless active infection. Based on this study, the improvement in Hp diagnostic

38 patients. Similar to the sequencing data, IM cellularities inferred from DNA methylation data were significantly higher in moderate/marked IM compared with mild IM (Figure S6A; 0.37 versus 0.2) (Houseman et al., 2014). IM cellularity values inferred from either sequencing or methylation data were also significantly correlated (Figure S6B).

Principal-component analysis of the normal gastric biopsies revealed clear clustering according to different anatomic regions, with antral samples segregating apart from body or cardia samples (Figure 4A). Within each anatomic side (antrum, body/cardia), IMs consistently exhibited global hypermethylation relative to normal gastric samples (Figure 4B), and IMs consistently segregated away from normal samples at both stomach sites (antrum or body/cardia) (Figure S6C). Notably, despite there being substantially more differentially methylated probes in antral

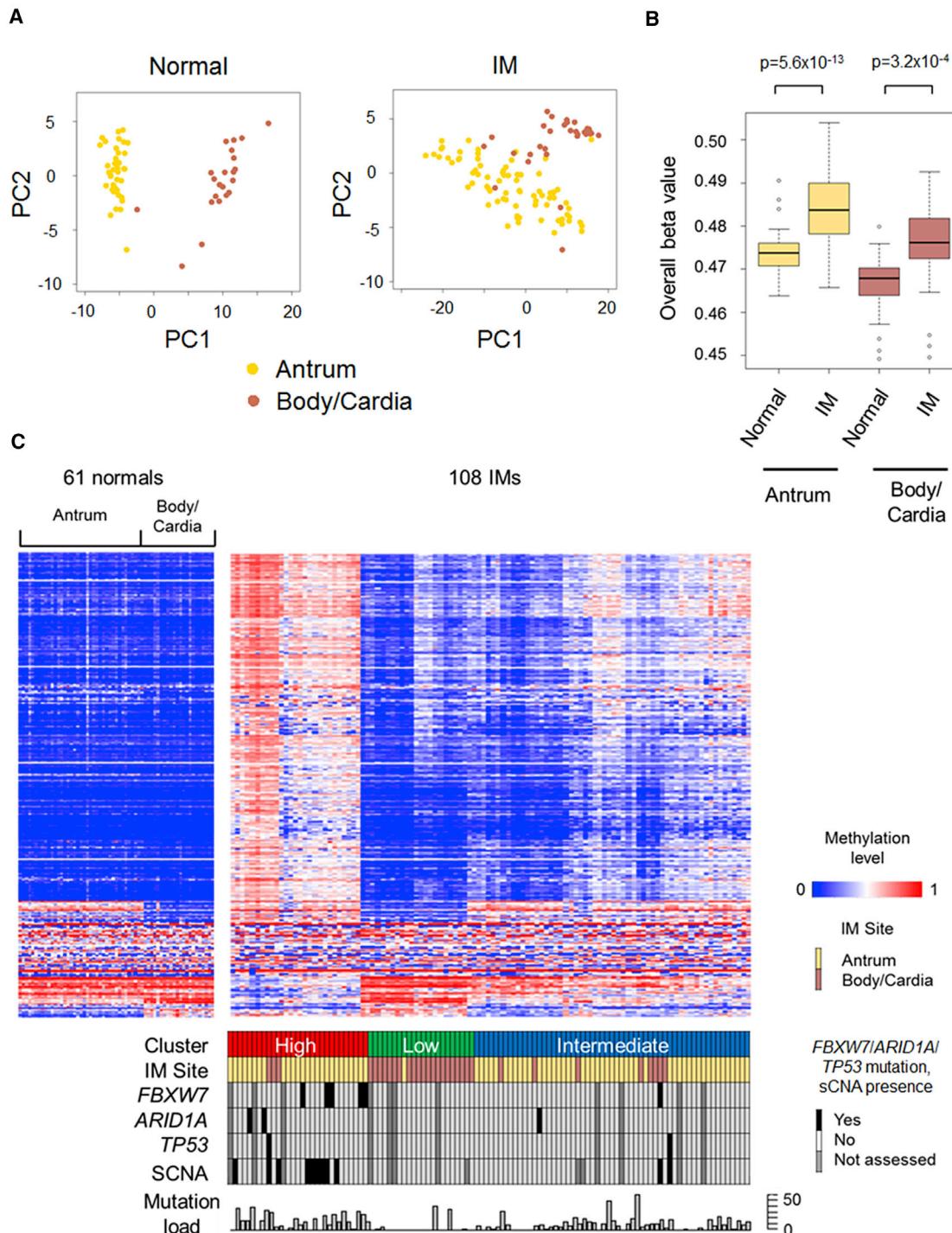


Figure 4. DNA Methylation Subtypes in IM

(A) Principal-component analysis (PCA) plot of DNA methylation profiles in normal (left) and IM (right) gastric biopsies. The 10,000 most variable CpG sites were used as input. Samples from antral regions are yellow and samples from body/cardia regions are brown.

(B) Boxplots show overall average β values (an index of DNA methylation levels) for normal and IM samples in antrum (yellow) and body/cardia (brown). Each box is the IQR, with lines representing the median. Whiskers are extended to within 1.5 IQR of the upper and lower quartiles. Data points falling outside this range are displayed independently. The p values were calculated using t tests.

(C) Heatmaps represent DNA methylation profiles of 108 IM samples based on unsupervised hierarchical clustering. DNA methylation profiles of the same sites in 61 normal gastric mucosae from the antrum and body/cardia are on the left heatmap. IMs are classified into high (red), intermediate (blue), and low-methylation (green) subgroups. Associations between clusters and genomic alterations are shown in the lower box.

See also Figure S6.

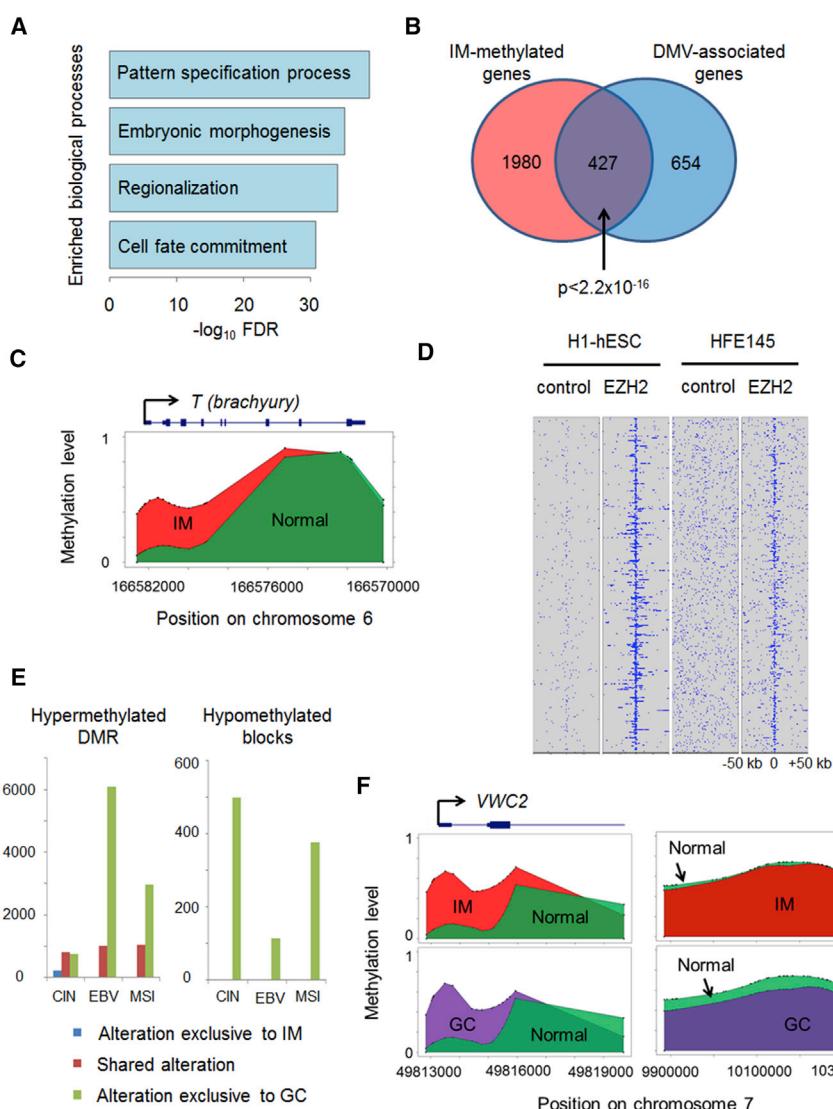


Figure 5. Biological Pathways Associated with IM DNA Methylation Changes

(A) Bar plot shows gene ontology (biological process) enrichments for hypermethylated regions. False discovery rate (FDR) values were calculated using a hypergeometric test. Pathways shown are $FDR < 1 \times 10^{-30}$.

(B) Overlap of genes hypermethylated in IM and genes associated with DNA methylation valleys (Xie et al., 2013). The p values were calculated using Fisher's exact test.

(C) Example of an IM DMR showing regional hypermethylation in IM. Methylation levels from 0 to 1 are shown across the developmental gene *T (brachyury)*, using 30 Illumina 450K probes.

(D) Heatmap of EZH2 ChIP-seq enrichment at IM DMRs using data from stem cells and HFE145 normal gastric cells. Each row represents a 100 kb window centered on an IM DMR. Input DNA from corresponding cells were used as controls.

(E) Bar plots show the number of hypermethylated regions (left) and hypomethylated blocks (right) in IM and GC (CIN, EBV, and MSI).

(F) Example of a shared hypermethylated DMR in IM and GC (left) and a GC-specific hypomethylated block. Methylation levels from 0 to 1 are shown across the gene *VWC2* (21 Illumina 450K probes) and also across a long-range hypomethylated block on chromosome 7 (95 Illumina 450K probes). See also Figure S7 and Table S4.

IM, there was a highly significant overlap in sites hypermethylated in both antral and gastric/body IMs, suggesting the existence of a common IM methylation signature (Figure S6D). Approximately one-third (34%) of IMs from the body and cardia clustered with antral IMs (Figure 4A) in contrast to nearly all normal samples segregating by tissue site (98%, 60/61), possibly reflecting the phenomenon of “antralization” (Xia et al., 2000). Unsupervised hierarchical clustering of all 108 IM samples identified three distinct clusters characterized by different DNA methylation levels (methylation-high, -intermediate, and -low; Figure 4C). The methylation-high and methylation-intermediate clusters were dominated by IMs from the antrum, while the low-methylation cluster was predominantly body and cardia.

To avoid confounding due to epigenetic differences between distinct sites-of-origin, we focused on antral and antral-like IMs in the methylation-high and methylation-intermediate clusters. Methylation-high antral IMs harbored significantly more somatic mutations ($p = 0.04$) and sCNAs ($p = 5.2 \times 10^{-4}$),

including *FBXW7* mutations ($p = 3.5 \times 10^{-3}$) and chromosome 8q amplifications ($p = 1.0 \times 10^{-3}$) (Figure 3C). We identified 2,537 hypermethylated regions in methylation-high IMs relative to normal antral tissues; pathway analysis of these regions revealed significant enrichment of biological processes including pattern specification process, cell fate commitment, and transcription factor activity (Figure 5A; hypergeometric false discovery rate $< 1 \times 10^{-5}$).

We noticed a strong overlap between genes hypermethylated in methylation-high IMs and genes associated with DNA methylation valleys (Figure 5B; odds ratio = 4.9), which are CpG-rich genomic regions that remain unmethylated in normal cell lineages but which acquire methylation in cancer (Xie et al., 2013). Examples of genes in these regions include stem cell lineage markers (e.g., *EOMES*, *T (brachyury)*, and *GSC*) (Figure 5C) and transcription factor families (e.g., *HOX*, *FOX*, and *GATA*). We further observed that genes hypermethylated in methylation-high IMs exhibited a strong enrichment of genes exhibiting bivalent promoter chromatin states in stem cells (Figure S7A), with *SUZ12* and *EZH2* as two potential transcription factors binding to the differentially methylated regions (Figure S7B). We confirmed associations of *EZH2* with IM DMRs using chromatin immunoprecipitation sequencing (ChIP-seq) data from H1 embryonic cells and also HFE145 normal gastric epithelial cells (Figure 5D). *EZH2* and *SUZ12* are part of polycomb repressive complex 2, which establishes H3K27me3 repressive chromatin marks (Margueron and Reinberg, 2011); similar

H3K27me3 enrichments were observed using ChIP-seq data for H1 cells and SNU484 GC cells (Figure S7C). These results suggest that hypermethylation events in IM are likely to target early developmental processes normally regulated by polycomb complexes.

GCS can be divided into four molecular subgroups referred to as CIN, EBV, MSI, and GS (Cancer Genome Atlas Research Network, 2014). To compare IM methylation alterations with those in malignant GC, we leveraged public TCGA data (450K methylation data from 248 patients). For this analysis, GS tumors (representing diffuse-type GCS) were excluded as these may develop through an IM-independent pathway (Grabsch and Tan, 2013). Compared with normal gastric antrum, we identified 1,553, 7,057, and 3,945 hypermethylated regions in CIN, EBV, and MSI type GC. Of these 78.3%, 98.6%, and 99.0% of regions hypermethylated in IM were also hypermethylated in CIN, EBV, and MSI tumors, respectively, suggesting that IM hypermethylation signatures are recapitulated in developed GC (Figures 5E and 5F). In contrast, we did not detect any hypomethylated regions in IM, compared with EBV, MSI, and CIN tumors which exhibited hundreds of globally hypomethylated regions particularly at intragenic regions (Figures 5E and 5F). This suggests that, while aberrant hypermethylation may occur early in IMs, global intragenic hypomethylation may occur during the subsequent transition of IMs to cancer.

We correlated results from molecular profiling against histological classifications of IM, specifically complete (type I) and incomplete (type II/III) IM. There were no significant associations between the IM histologic classes with mutation burden, copy-number changes, Hp density, and DNA methylation levels, although we noted a trend for incomplete IMs to have shorter telomeres ($p = 0.16\text{--}0.17$; Table S4). These results suggest that IM molecular data is likely to capture patterns of heterogeneity independent from histology.

IM Genomic/Epigenomic Profiles and Risk Prediction

Finally, to identify molecular alterations correlated with IM clinical outcome, we focused on IMs from the stomach antrum (82 IMs from distinct subjects). We chose to concentrate initially on antral IMs only, as they (1) represent the largest group and (2) samples from the antrum and body/cardia show very different DNA methylation profiles irrespective of being normal and IM, which may lead to confounding of the epigenomic data if antral and body/cardia IMs are analyzed in a combined fashion. We classified subjects into regressive, persistent, and progressive groups, based on outcomes at completion of surveillance (Figure 6A). Subjects showing at least a two-tier reduction of IM histological grade or absence of IM after two subsequent assessments were classified as regressive ($n = 15$), in accordance with the general principles defined by Rugge et al. (2003), who defined regression as a situation where in at least two subsequent assessments, dysplasia is either no longer detected or found to a lesser degree. The progressive group was defined as subjects with antral IM that eventually developed EGN ($n = 6$). All other subjects were considered as persistent ($n = 61$).

At the univariate level, somatic mutation levels and DNA methylation alterations were both significantly correlated with IM regression. IMs that eventually regressed exhibited significantly lower overall mutation burdens (average mutation = 8.9;

Figure 6B, top) and DNA methylation changes ($\beta = 0.480$; Figure 6B, bottom), compared with persistent IM (average mutation = 14.6, $\beta = 0.485$). Strikingly, >50% (10/18) of IMs exhibiting normal-like patterns of DNA methylation eventually regressed (10/18), compared with only 10% (5/52) of antral IMs exhibiting aberrant methylation profiles ($p = 0.0003$). We also performed multivariate logistical regression analyses to examine the impact of multiple factors on IM regression. After considering mutation burden, DNA methylation level, and Hp density for regression, no one single factor was significantly associated with IM regression, although a trend persisted for low methylation levels ($p = 0.16$; Table S5).

For subjects who developed EGN (progressive IMs), we found that telomere lengths were significantly shorter in progressive IMs than persistent IM (Figure 6C, top). The presence of sCNAs also correlated with higher risk of progression (Figure 6C, bottom). After logistic regression considering telomere length, presence of sCNAs, and histology as risk factors for progression, only the presence of sCNAs remained significant for IM progression (Table S5). Besides the antral IMs, we also investigated if similar trends might also apply for IMs in the body/cardia. Despite the much smaller sample size which prohibited rigorous statistical analysis (34 body/cardia IMs, of which 2 were progressive), we observed that the two progressive body-cardia IMs also had shortened telomeres, similar to antral IMs. In an analysis combining IMs from the antrum/body/cardia, we again found that both shortened telomeres and increased copy-number alterations were significantly associated with progression (Table S6; $p = 1.1 \times 10^{-5}$ for shortened telomeres; Welch's t test and $p = 1.6 \times 10^{-2}$ for increased sCNA; Fisher's test). These observations suggest that telomere shortening and chromosomal instability may identify those IM patients with the highest risk of progression to GC. We emphasize that these results are likely limited to intestinal-type GC, as diffuse-type GC, despite also being associated with Hp infection is not known to be associated with IM (Grabsch and Tan, 2013).

DISCUSSION

IM is an established pre-malignant condition, and known IM risk factors include Hp infection, high salt intake, alcohol consumption, and chronic bile reflux (Correa et al., 2010). However, despite epidemiological data, it remains controversial whether IMs are direct cellular precursors of dysplasia and GC. While animal models have demonstrated that Hp-induced chronic inflammation can induce metaplasia (Yoshizawa et al., 2007), these animal models typically did not progress to GC (Goldenring et al., 2010). Instead, it has been proposed that GC may arise from stem cells (Hayakawa et al., 2015) or differentiated “reserve” stem cells (Leushacke et al., 2017). Under these models, Hp infection may induce a regenerative response by stem cells, and the expansion of stem cells subsequently predisposes cells to acquire additional genetic alterations, giving rise to metaplasia and dysplasia (Hayakawa et al., 2017).

Here, to understand molecular alterations in human IM, we integrated genomic (somatic mutations, copy-number alterations, and telomere length), epigenomic (DNA methylation), and pathogen (Hp) alterations to prospective outcome data, in a cohort of cancer-free patients. We found that metaplastic cells

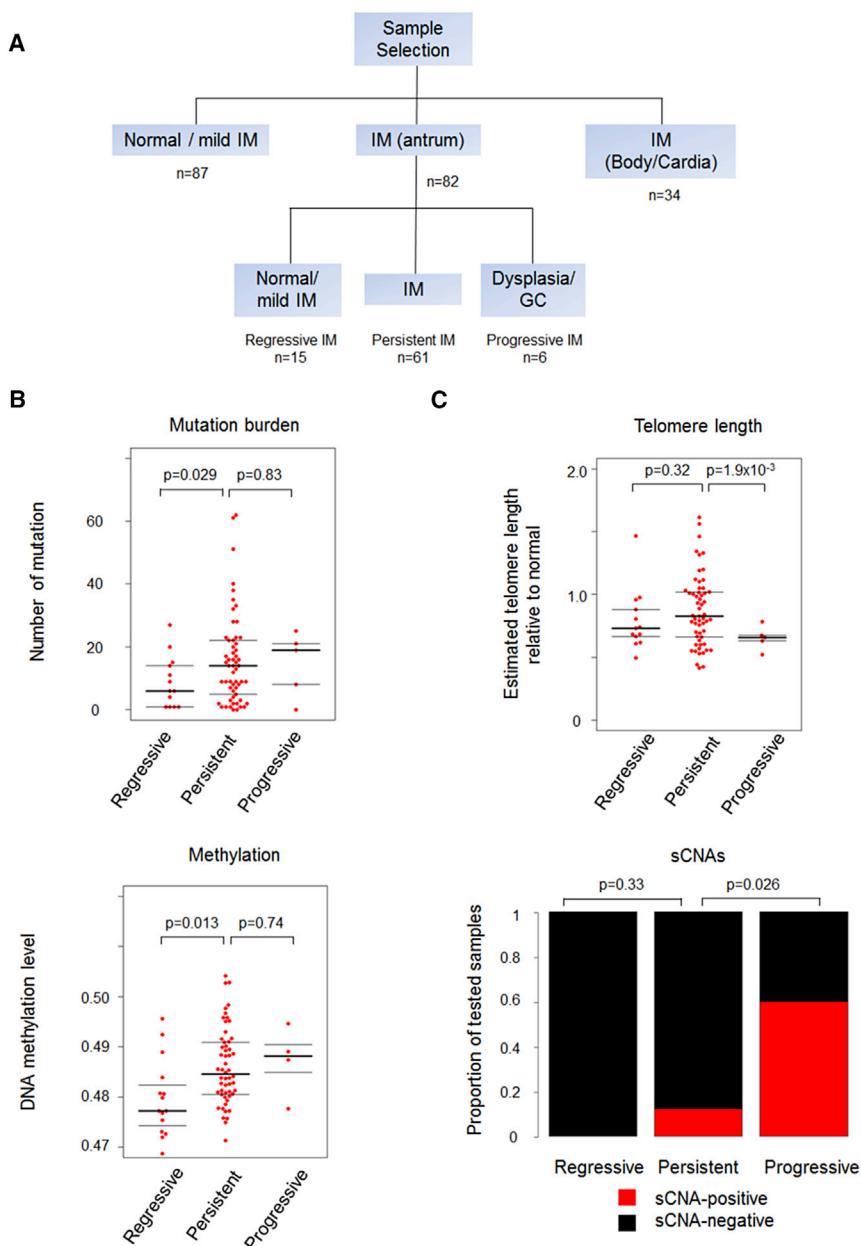


Figure 6. Molecular IM Features Correlated to Clinical Outcomes

(A) A total of 82 patients with baseline IM in the antrum were included for assessment of correlates between molecular features and clinical outcome. At the end of surveillance, 6 patients had developed EGN (progressive), 61 patients had no changes in IM status (persistent), and 15 patients exhibited a regression in IM severity (regressive). (B) Analyses of regression with somatic mutations (top) and methylation changes. The p values were calculated using t tests. Thick bars indicate medians and thin bars indicate first and third quartiles. (C) Analyses of progression with telomere length (top) and sCNAs (bottom). The p values were calculated using either t tests (telomeres) or Fisher's exact test (sCNAs). Thick bars indicate medians and thin bars indicate first and third quartiles.

See also Tables S5 and S6.

can exhibit both clonal mutations (i.e., MAF > 3%–4%) and epigenetic alterations, and that these aberrations are correlated with IM cellularity between histopathological categories (mild IM versus marked/moderate IM), across individual samples, or observed in LCM-purified IM cells. However, although clonal mutations can occur in IM (e.g., *TP53*, *ARID1A*, and *FBXW7*), our data also indicate that such mutations are infrequent (2%–4.7%) and largely present in distinct samples. Also, IMs from the same patient exhibit high intra-patient heterogeneity and are often genetically distinct.

The observed low rate of IM clonal mutations is consistent with previous reports (Busuttil et al., 2014; Shimizu et al., 2014), for example, Shimizu et al. (2014), reported that, by using whole-exome sequencing at a mean coverage of 41.2×, they observed

TP53 mutations in 3/5 GC samples, but in 0/5 non-cancerous gastritis mucosa. However, it is worth noting that, in our study, we used a mean sequencing coverage of 365×, which detected mutations at a median MAF of 4%. This sensitivity, while sufficient to detect clonal mutations, is likely insufficient to detect mutations at very low MAF levels (<1%), as these are only detectable by ultra-deep sequencing. As such, our results also do not preclude the possibility that additional low-abundance mutations may also occur in Hp-infected gastric tissues. Indeed, when Shimizu et al. (2014) analyzed Hp-infected non-cancerous gastritis mucosa using an ultra-deep sequencing platform (mean coverage 4024×, almost 100× higher than conventional exome sequencing), 15/44 cases exhibited *TP53* mutations at very low MAF values (0.1%–0.26%). Supporting this finding, in an exploratory analysis on a subset of our IMs sequenced at much

higher depth (1,855×) using molecularly barcoded sequencing, we too observed *TP53* G245D mutations occurring at allele frequencies between 0.42% and 0.67% in 6/8 samples (MAFs 0.42%–0.67%). G245D is the most common mutation reported in Hp-infected gastric mucosa (Shimizu et al., 2014). Hp infection may thus cause a plethora of latently accumulating low-abundance mutations, facilitating the subsequent propagation of clonal mutations in certain IMs.

Our results have three broad areas of translational and clinical significance. First, our findings may contribute toward a better understanding of the “point-of-no-return”; a hypothesized stage where Hp eradication is no longer able to reverse mucosal damage and subsequent GC development. Evidence supporting a point-of-no-return has been suggested through cohort studies

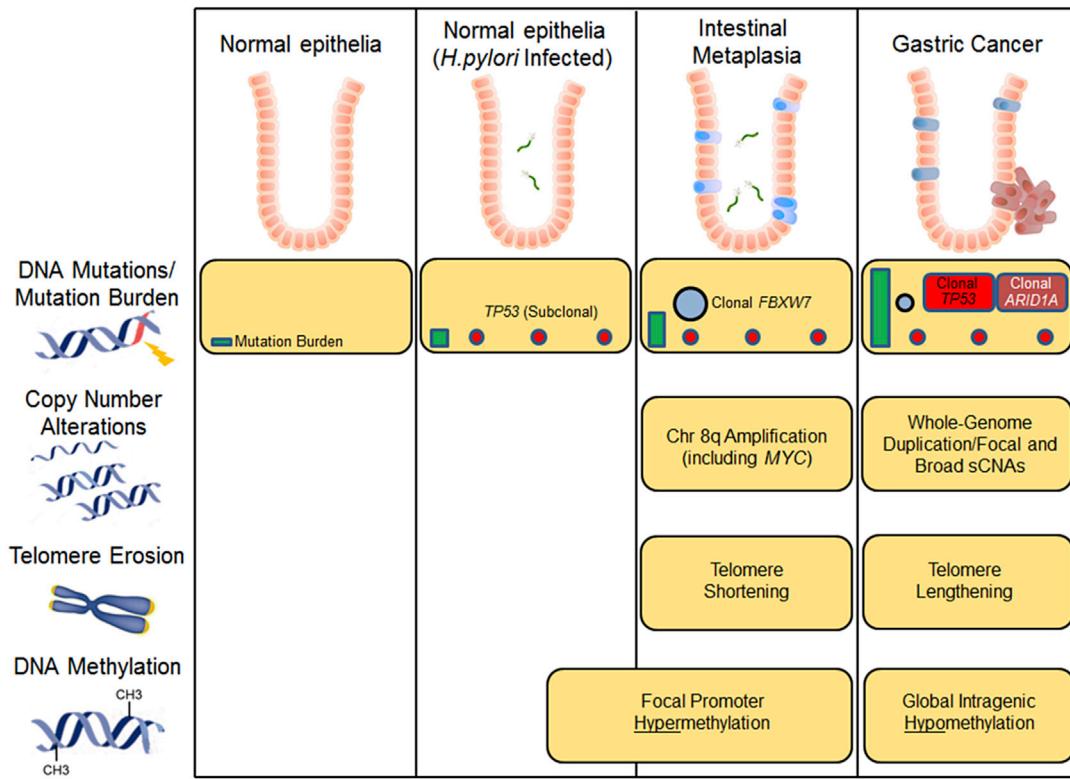


Figure 7. Molecular Alterations across Gastric Pathology Stages

Infection of gastric epithelium (orange cells) by Hp are proposed to cause subclonal mutations and initial DNA methylation changes contributing to IM development (blue). Certain IMs exhibit clonal mutations (e.g., *FBXW7*), but these are still infrequent and mutational burdens remain relatively low, with initial sCNAs (8q) and telomere erosion. Transition to GC (red) likely requires additional mutations in key GC tumor-suppressor genes (*TP53* and *ARID1A*), telomere restoration, and intragenic DNA hypomethylation that will enhance genomic instability (e.g., WGD). Due to high IM intra-patient heterogeneity and lack of definitive evidence that IMs directly transit to GC, this model does not assume a strict linear progression where IMs exhibiting clonal mutations are the direct cellular precursors of GC.

assessing the efficacy of Hp mass eradication. For example, Wong et al. (2004) showed that Hp eradication displayed a protective role only in subjects without precancerous lesions (GC, IM, or dysplasia), while Lee et al. (2013) reported that Hp eradication led to significant reductions in Hp infection and GA, but no difference in levels of IM severity and gastric dysplasia. Other studies, however, have challenged the point-of-no-return. For example, Kang et al. (2012) reported that Hp eradication might prove beneficial even in moderate to severe IM patients, and a recent meta-analysis concluded that Hp eradication is associated with GC incidence reduction in all levels of baseline risk, including IM patients (Lee et al., 2016b). Molecularly, the exact point-of-no return still remains to be defined, but may correspond to irreversible molecular changes (gene mutations, telomere reduction, or chromosomal rearrangements) occurring during pre-malignancy. In our study, a subset of patients with intermediate levels of DNA methylation changes tended to show histological regression, while patients with high levels of DNA methylation changes tended to be persistent/progressive. Notably, we found that Hp infection levels were correlated with DNA methylation levels in the intermediate group, but not in the methylation-high group, suggesting that Hp eradication may potentially reverse or delay methylation changes in the former but not the latter. Levels of aberrant DNA methylation may thus also influence the point-of-no-return.

Second, in the context of GC screening, our results suggest that integrating genomic and epigenomic data may represent a molecularly guided risk stratification strategy for identifying IM patients at maximal risk of GC. Early detection is a principal strategy for improving GC patient outcomes; however, the need for a risk stratification approach to GC screening is particularly important for countries such as Singapore, where the absolute GC incidence is lower than in Japan. This is because early GC diagnosis is best achievable by endoscopic screening and due to the inherent costs, invasiveness, and risks of current endoscopic technologies, implementing population-wide GC screening programs based on endoscopy is only cost-effective in countries with extremely high GC incidence rates (Gupta et al., 2011). In moderate-risk countries (e.g., Singapore), general population screening is not justifiable, and in these countries, early GC detection may be facilitated by focusing screening efforts only on high-risk populations to maximize the probability of detecting early GCs (i.e., risk stratification). Unfortunately, there is as yet no recognized consensus criteria for identifying the optimal subgroup at highest risk of GC that might benefit from endoscopic screening. Currently, IMs are phenotypically stratified by histologic criteria (diffuse, severe, or incomplete) or OLGIM scores (operative link on gastric intestinal metaplasia assessment) (Capelle et al., 2010). Our data suggest that patients whose

IMs exhibit telomere shortening and chromosomal instability are significantly more likely to progress to GC. Such findings illustrate the potential of stratifying IMs using molecular criteria, setting the stage for precision approaches to GC prevention.

Third, we found that sequencing revealed additional patients with Hp DNA in IM tissues beyond those identified by histology. While acknowledging current controversies regarding the reversibility of IM after Hp eradication (Dinis-Ribeiro et al., 2012), we nevertheless note that meta-analyses of randomized clinical trials (Lee et al., 2016b), as well as real world population screening studies (Lee et al., 2013), provide convincing evidence of the benefits of Hp eradication in both asymptomatic and in early GC patients. Accurate diagnosis of Hp is therefore crucial in the effective management of GC pre-malignancy. Our results highlight the potential utility of sequencing as an alternative method for detecting Hp, which may prove particularly useful in IM patients, as histological diagnosis of Hp is more challenging when IM is concurrently present (Lee and Kim, 2015). Besides providing increased sensitivity, we also note that sequencing also provides information on antibiotic resistance and virulence factors that may also influence clinical outcomes and treatment decisions. If independently validated, our results thus suggest that molecular methods (e.g., sequencing) may provide a diagnostic gain in patients with histologically negative IM biopsies, identifying additional patients for Hp eradication.

In conclusion, we propose that in infected gastric mucosa, Hp-induced inflammation causes the generation of subclonal mutations and DNA methylation alterations, contributing to IM development (Figure 7). This may lead to clonal mutations (e.g., *FBXW7*) and epigenetic alterations in IM; however, these mutations are still infrequent and mutation burdens remain relatively low. Further acquisition of other mutations, large-scale chromosomal alterations, and additional epigenetic changes are then required to progress to GC. Due to significant IM intra-patient genetic heterogeneity, it remains open if IMs exhibiting clonal mutations are the direct cellular precursors to eventual GC. However, supporting the model that chromosomal alterations occur after the development of IM, we observed a lack of intragenic DNA hypomethylation in IM compared with GCs, as hypomethylated DNA has also been associated with chromosomal instability (Eden et al., 2003). Finally, there is emerging interest in using molecular data to guide cancer prevention, through efforts such as the Pre-Cancer Atlas (Spira et al., 2017). The observation of increased DNA methylation in a significant proportion of IMs raises the potential for low-dose epigenetic therapy as a potential chemopreventive strategy, using clinically approved demethylating agents such as 5-azacitidine and decitabine.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - GCEP Cohort
 - Sample Selection
 - Cell Lines

● METHOD DETAILS

- Selection of Genomic Targets
- DNA Extraction, Target Capture and Library Construction
- Read Mapping and Somatic Mutation Detection
- Validation of Somatic Variants with Ion Torrent Resequencing
- Copy Number, Telomere Length and Hp Density Estimation
- Illumina Infinium Assay (Copy Number)
- Illumina Infinium HD Methylation Assay
- Immunohistochemistry and Stains
- Laser Capture Microdissection (LCM)
- Sanger Sequencing of LCM Material
- siRNA Transfection and Cisplatin Treatment
- Molecular Cloning and Stable Cell Line Establishment
- Quantitative RT-PCR
- Western Blotting
- Cell Proliferation Assay
- ChIP-seq Signal Aggregation

● QUANTIFICATION AND STATISTICAL ANALYSIS

● DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and six tables and can be found with this article online at <https://doi.org/10.1016/j.ccr.2017.11.018>.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation Singapore under its Translational and Clinical Research (TCR) Flagship Program administered by the Singapore Ministry of Health's National Medical Research Council (TCR/009-NUHS/2013) and grant NMRC/STaR/0026/2015. Other sources include Duke-NUS Medical School, and the Cancer Science Institute of Singapore, NUS, under the National Research Foundation Singapore and the Singapore Ministry of Education under its Research Centers of Excellence initiative. We thank Hassan Ashktorab, Alfred Cheng, and Sun Young Rha for biological reagents. We thank all members of the Singapore Gastric Cancer Consortium and Soh Ee Lee and Jia Wei Lee for technical assistance.

AUTHOR CONTRIBUTIONS

Conceptualization, M.T., K.G.Y., and P.T.; Methodology, K.K.H., K.R., F.Z., S.S., C.X., A.L.K.T., M.L., S.T., K.D., M.X., A.F., S.M.F.A., T.K.H.L., N.B., M.T., K.G.Y., and P.T.; Investigation, K.K.H., K.R., F.Z., S.S., C.X., A.L.K.T., M.L., S.T., M.T., K.G.Y., and P.T.; Resources, J.L., K.Y.H., S.G.R., B.T.T., C.K.C., C.K., C.J.O., K.M.F., J.S., W.C.L., K.L.L., T.L.A., A.W., J.R., A.R., L.G.L., and W.M.Y.; Writing – Original Draft, K.K.H., K.R., F.Z., S.S., M.T., K.G.Y., and P.T.; Writing – Review & Editing, K.K.H., K.R., F.Z., S.S., M.T., K.G.Y., and P.T.; Funding Acquisition, K.G.Y. and P.T.

Received: June 9, 2017

Revised: October 2, 2017

Accepted: November 28, 2017

Published: December 28, 2017

REFERENCES

- Akalin, A., Franke, V., Vlahovicek, K., Mason, C.E., and Schubeler, D. (2015). Genomation: a toolkit to summarize, annotate and visualize genomic intervals. *Bioinformatics* 31, 1127–1129.
- Akhoondi, S., Sun, D., von der Lehr, N., Apostolidou, S., Klotz, K., Maljukova, A., Cepeda, D., Fiegl, H., Dafou, D., Marth, C., et al. (2007). FBXW7/hCDC4 is a general tumor suppressor in human cancer. *Cancer Res.* 67, 9006–9012.

- Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D., and Irizarry, R.A. (2014). Minfi: a flexible and comprehensive bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 30, 1363–1369.
- Busuttil, R.A., Zapparoli, G.V., Haupt, S., Fennell, C., Wong, S.Q., Pang, J.M., Takeno, E.A., Mitchell, C., Di Costanzo, N., Fox, S., et al. (2014). Role of p53 in the progression of gastric cancer. *Oncotarget* 5, 12016–12026.
- Calvet, X., Sanchez-Delgado, J., Montserrat, A., Lario, S., Ramirez-Lazaro, M.J., Quesada, M., Casalots, A., Suarez, D., Campo, R., Brullet, E., et al. (2009). Accuracy of diagnostic tests for *Helicobacter pylori*: a reappraisal. *Clin. Infect. Dis.* 48, 1385–1391.
- Cancer Genome Atlas Research Network. (2014). Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 513, 202–209.
- Capelle, L.G., de Vries, A.C., Haringsma, J., Ter Borg, F., de Vries, R.A., Bruno, M.J., van Dekken, H., Meijer, J., van Grieken, N.C., and Kuipers, E.J. (2010). The staging of gastritis with the OLGA system by using intestinal metaplasia as an accurate alternative for atrophic gastritis. *Gastrointest. Endosc.* 71, 1150–1158.
- Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A., et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* 30, 413–421.
- Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* 31, 213–219.
- Correa, P., Piazuelo, M.B., and Wilson, K.T. (2010). Pathology of gastric intestinal metaplasia: clinical implications. *Am. J. Gastroenterol.* 105, 493–498.
- Costello, M., Pugh, T.J., Fennell, T.J., Stewart, C., Lichtenstein, L., Meldrim, J.C., Fostel, J.L., Friedrich, D.C., Perrin, D., Dionne, D., et al. (2013). Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res.* 41, e67.
- Davis, H., Lewis, A., Behrens, A., and Tomlinson, I. (2014). Investigation of the atypical FBXW7 mutation spectrum in human tumours by conditional expression of a heterozygous propellar tip missense allele in the mouse intestines. *Gut* 63, 792–799.
- de Vries, A.C., van Grieken, N.C., Looman, C.W., Casparie, M.K., de Vries, E., Meijer, G.A., and Kuipers, E.J. (2008). Gastric cancer risk in patients with premalignant gastric lesions: a nationwide cohort study in the Netherlands. *Gastroenterology* 134, 945–952.
- den Hoed, C.M., Holster, I.L., Capelle, L.G., de Vries, A.C., den Hartog, B., Ter Borg, F., Biermann, K., and Kuipers, E.J. (2013). Follow-up of premalignant lesions in patients at risk for progression to gastric cancer. *Endoscopy* 45, 249–256.
- DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498.
- Ding, Z., Mangino, M., Aviv, A., Spector, T., and Durbin, R.; Consortium, U.K. (2014). Estimating telomere length from whole genome sequence data. *Nucleic Acids Res.* 42, e75.
- Divis-Ribeiro, M., Areia, M., de Vries, A.C., Marcos-Pinto, R., Monteiro-Soares, M., O'Connor, A., Pereira, C., Pimentel-Nunes, P., Correia, R., Ensari, A., et al. (2012). Management of precancerous conditions and lesions in the stomach (MAPS): guideline from the European Society of Gastrointestinal Endoscopy (ESGE), European Helicobacter Study Group (EHSG), European Society of Pathology (ESP), and the Sociedade Portuguesa de Endoscopia Digestiva (SPED). *Endoscopy* 44, 74–94.
- Dixon, M.F. (2002). Gastrointestinal epithelial neoplasia: Vienna revisited. *Gut* 51, 130–131.
- Dixon, M.F., Genta, R.M., Yardley, J.H., and Correa, P. (1996). Classification and grading of gastritis. The updated Sydney system. International workshop on the histopathology of gastritis, Houston 1994. *Am. J. Surg. Pathol.* 20, 1161–1181.
- Eden, A., Gaudet, F., Waghmare, A., and Jaenisch, R. (2003). Chromosomal instability and tumors promoted by DNA hypomethylation. *Science* 300, 455.
- Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., and Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* 136, E359–E386.
- Gehring, J.S., Fischer, B., Lawrence, M., and Huber, W. (2015). SomaticSignatures: inferring mutational signatures from single-nucleotide variants. *Bioinformatics* 31, 3673–3675.
- Goldenring, J.R., Nam, K.T., Wang, T.C., Mills, J.C., and Wright, N.A. (2010). Spasmolytic polypeptide-expressing metaplasia and intestinal metaplasia: time for reevaluation of metaplasias and the origins of gastric cancer. *Gastroenterology* 138, 2207–2210, 2210.e1.
- Gonzalez, C.A., Sanz-Anquela, J.M., Companioni, O., Bonet, C., Berdasco, M., Lopez, C., Mendoza, J., Martin-Arranz, M.D., Rey, E., Poves, E., et al. (2016). Incomplete type of intestinal metaplasia has the highest risk to progress to gastric cancer: results of the Spanish follow-up multicenter study. *J. Gastroenterol. Hepatol.* 31, 953–958.
- Grabsch, H.I., and Tan, P. (2013). Gastric cancer pathology and underlying molecular mechanisms. *Dig. Surg.* 30, 150–158.
- Gupta, N., Bansal, A., Wani, S.B., Gaddam, S., Rastogi, A., and Sharma, P. (2011). Endoscopy for upper GI cancer screening in the general population: a cost-utility analysis. *Gastrointest. Endosc.* 74, 610–624.e2.
- Hayakawa, Y., Ariyama, H., Stancikova, J., Sakitani, K., Asfaha, S., Renz, B.W., Dubeykovskaya, Z.A., Shibata, W., Wang, H., Westphalen, C.B., et al. (2015). Mist1 expressing gastric stem cells maintain the normal and neoplastic gastric epithelium and are supported by a perivascular stem cell niche. *Cancer Cell* 28, 800–814.
- Hayakawa, Y., Fox, J.G., and Wang, T.C. (2017). The origins of gastric cancer from gastric stem cells: lessons from mouse models. *Cell Mol. Gastroenterol. Hepatol.* 3, 331–338.
- Houseman, E.A., Molitor, J., and Marsit, C.J. (2014). Reference-free cell mixture adjustments in analysis of DNA methylation data. *Bioinformatics* 30, 1431–1439.
- Jiang, Y., Qi, X., Liu, X., Zhang, J., Ji, J., Zhu, Z., Ren, J., and Yu, Y. (2017). Fbxw7 haploinsufficiency loses its protection against DNA damage and accelerates MNU-induced gastric carcinogenesis. *Oncotarget* 8, 33444–33456.
- Kang, G.H. (2012). CpG island hypermethylation in gastric carcinoma and its premalignant lesions. *Korean J. Pathol.* 46, 1–9.
- Kang, J.M., Kim, N., Shin, C.M., Lee, H.S., Lee, D.H., Jung, H.C., and Song, I.S. (2012). Predictive factors for improvement of atrophic gastritis and intestinal metaplasia after *Helicobacter pylori* eradication: a three-year follow-up study in Korea. *Helicobacter* 17, 86–95.
- Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499, 214–218.
- Lee, J.Y., and Kim, N. (2015). Diagnosis of *Helicobacter pylori* by invasive test: histology. *Ann. Transl. Med.* 3, 10.
- Lee, T.Y., Wang, R.C., Lee, Y.C., Lin, J.T., Ho, H.J., Hsieh, M.C., and Wu, C.Y. (2016a). The incidence of gastric adenocarcinoma among patients with gastric intestinal metaplasia: a long-term cohort study. *J. Clin. Gastroenterol.* 50, 532–537.
- Lee, Y.C., Chen, T.H., Chiu, H.M., Shun, C.T., Chiang, H., Liu, T.Y., Wu, M.S., and Lin, J.T. (2013). The benefit of mass eradication of *Helicobacter pylori* infection: a community-based study of gastric cancer prevention. *Gut* 62, 676–682.
- Lee, Y.C., Chiang, T.H., Chou, C.K., Tu, Y.K., Liao, W.C., Wu, M.S., and Graham, D.Y. (2016b). Association between *Helicobacter pylori* eradication and gastric cancer incidence: a systematic review and meta-analysis. *Gastroenterology* 150, 1113–1124.e5.

- Leushacke, M., Tan, S.H., Wong, A., Swathi, Y., Hajamohideen, A., Tan, L.T., Goh, J., Wong, E., Denil, S., Murakami, K., and Barker, N. (2017). Lgr5-expressing chief cells drive epithelial regeneration and cancer in the oxyntic stomach. *Nat. Cell Biol.* **19**, 774–786.
- Li, D., Bautista, M.C., Jiang, S.F., Daryani, P., Brackett, M., Armstrong, M.A., Hung, Y.Y., Postlethwaite, D., and Ladabaum, U. (2016a). Risks and predictors of gastric adenocarcinoma in patients with gastric intestinal metaplasia and dysplasia: a population-based study. *Am. J. Gastroenterol.* **111**, 1104–1113.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760.
- Li, H., Wang, Z., Zhang, W., Qian, K., Xu, W., and Zhang, S. (2016b). Fbxw regulates tumor apoptosis, growth arrest and the epithelial-to-mesenchymal transition in part through the RhoA signaling pathway in gastric cancer. *Cancer Lett.* **370**, 39–55.
- Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343–349.
- McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501.
- Mu, Y., Zhang, Q., Mei, L., Liu, X., Yang, W., and Yu, J. (2012). Telomere shortening occurs early during gastic carcinogenesis. *Med. Oncol.* **29**, 893–898.
- Niwa, T., Tsukamoto, T., Toyoda, T., Mori, A., Tanaka, H., Maekita, T., Ichinose, M., Tatematsu, M., and Ushijima, T. (2010). Inflammatory processes triggered by *Helicobacter pylori* infection cause aberrant DNA methylation in gastric epithelial cells. *Cancer Res.* **70**, 1430–1440.
- Rugge, M., Cassaro, M., Di Mario, F., Leo, G., Leandro, G., Russo, V.M., Pennelli, G., and Farinati, F.; Interdisciplinary Group on Gastric Epithelial Dysplasia (IGGED) (2003). The long term outcome of gastric non-invasive neoplasia. *Gut* **52**, 1111–1116.
- Shimizu, T., Marusawa, H., Matsumoto, Y., Inuzuka, T., Ikeda, A., Fujii, Y., Minamiguchi, S., Miyamoto, S., Kou, T., Sakai, Y., et al. (2014). Accumulation of somatic mutations in TP53 in gastric epithelium with *Helicobacter pylori* infection. *Gastroenterology* **147**, 407–417.e3.
- Spira, A., Yurgelun, M.B., Alexandrov, L., Rao, A., Bejar, R., Polyak, K., Giannakis, M., Shilatifard, A., Finn, O.J., Dhopakar, M., et al. (2017). Precancer Atlas to drive precision prevention trials. *Cancer Res.* **77**, 1510–1541.
- Suzuki, K., Suzuki, I., Leodolter, A., Alonso, S., Horiochi, S., Yamashita, K., and Perucho, M. (2006). Global DNA demethylation in gastrointestinal cancer is age dependent and precedes genomic damage. *Cancer Cell* **9**, 199–207.
- Tahara, T., Shibata, T., Kawamura, T., Horiguchi, N., Okubo, M., Nakano, N., Ishizuka, T., Nagasaka, M., Nakagawa, Y., and Ohmiya, N. (2016). Telomere length shortening in gastric mucosa is a field effect associated with increased risk of gastric cancer. *Virchows Arch.* **469**, 19–24.
- Tan, P., and Yeoh, K.G. (2015). Genetics and molecular pathogenesis of gastric adenocarcinoma. *Gastroenterology* **149**, 1153–1162.e3.
- Taylor-Weiner, A., Zack, T., O'Donnell, E., Guerriero, J.L., Bernard, B., Reddy, A., Han, G.C., AlDubayan, S., Amin-Mansour, A., Schumacher, S.E., et al. (2016). Genomic evolution and chemoresistance in germ-cell tumours. *Nature* **540**, 114–118.
- Van Loo, P., Nordgard, S.H., Lingjaerde, O.C., Russnes, H.G., Rye, I.H., Sun, W., Weigman, V.J., Marynen, P., Zetterberg, A., Naume, B., et al. (2010). Allele-specific copy number analysis of tumors. *Proc. Natl. Acad. Sci. USA* **107**, 16910–16915.
- Veijola, L., Oksanen, A., Sipponen, P., and Rautelin, H. (2008). Evaluation of a commercial immunoblot, Helicoblot 2.1, for diagnosis of *Helicobacter pylori* infection. *Clin. Vaccin. Immunol.* **15**, 1705–1710.
- Welcker, M., and Clurman, B.E. (2008). FBW7 ubiquitin ligase: a tumour suppressor at the crossroads of cell division, growth and differentiation. *Nat. Rev. Cancer* **8**, 83–93.
- Wong, B.C., Lam, S.K., Wong, W.M., Chen, J.S., Zheng, T.T., Feng, R.E., Lai, K.C., Hu, W.H., Yuen, S.T., Leung, S.Y., et al. (2004). *Helicobacter pylori* eradication to prevent gastric cancer in a high-risk region of China: a randomized controlled trial. *JAMA* **291**, 187–194.
- Xia, H.H., Kalantar, J.S., Talley, N.J., Wyatt, J.M., Adams, S., Chueng, K., and Mitchell, H.M. (2000). Antral-type mucosa in the gastric incisura, body, and fundus (antralization): a link between *Helicobacter pylori* infection and intestinal metaplasia? *Am. J. Gastroenterol.* **95**, 114–121.
- Xie, W., Schultz, M.D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., Whitaker, J.W., Tian, S., Hawkins, R.D., Leung, D., et al. (2013). Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134–1148.
- Yoshizawa, N., Takenaka, Y., Yamaguchi, H., Tetsuya, T., Tanaka, H., Tatematsu, M., Nomura, S., Goldenring, J.R., and Kaminishi, M. (2007). Emergence of spasmolytic polypeptide-expressing metaplasia in Mongolian gerbils infected with *Helicobacter pylori*. *Lab. Invest.* **87**, 1265–1276.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
p53	DAKO	Cat# NCL-p53-DO7-FITC; RRID:AB_876938
c-Myc	Abcam	Cat# NB110-57172; RRID:AB_843803
ARID1A	Abcam	Cat# ab182561
Biological Samples		
Gastric mucosa tissues (normal, mild intestinal metaplasia and intestinal metaplasia)	This study	GCEP
Gastric cancer and normal gastric mucosa tissues	In-house data	-
Gastric cancer and normal gastric mucosa tissues	TCGA (https://portal.gdc.cancer.gov)	TCGA-STAD
Deposited Data		
GCEP targeted sequencing data	https://www.ncbi.nlm.nih.gov/sra	PRJNA383939
GCEP DNA methylation data	https://www.ncbi.nlm.nih.gov/geo/	GSE103186
Experimental Models: Cell Lines		
HFE145	gift from Dr. Hassan Ashktorab, Howard University	-
GES-1	gift from Dr. Alfred Cheng, Chinese University of Hong Kong	RRID:CVCL_EQ22
TMK1	Korean Cell Line Bank	RRID:CVCL_4384
NCC24	Japan Health Science Research Resource Bank	RRID:CVCL_8899
Software and Algorithms		
BWA-mem	(Li and Durbin, 2009)	http://bio-bwa.sourceforge.net/
GATK	(DePristo et al., 2011).	https://software.broadinstitute.org/gatk/
Picard	-	http://broadinstitute.github.io/picard/
Mutect2	(Cibulskis et al., 2013)	https://software.broadinstitute.org/gatk/documentation/tooldocs/current/org_broadinstitute_gatk_tools_walkers_cancer_m2_MuTect2.php
MutsigCV	(Lawrence et al., 2013)	http://archive.broadinstitute.org/cancer/cga/mutsig
SomaticSignatures	(Gehring et al., 2015)	http://bioconductor.org/packages/release/bioc/html/SomaticSignatures.html
ABSOLUTE	(Carter et al., 2012)	http://software.broadinstitute.org/cancer/software/genepattern/modules/docs/ABSOLUTE
Telsq	(Ding et al., 2014)	https://github.com/zd1/telsq
ASCAT	(Van Loo et al., 2010)	https://github.com/Crick-CancerGenomics/ascat
Minfi	(Aryee et al., 2014)	https://bioconductor.org/packages/release/bioc/html/minfi.html
GREAT	(McLean et al., 2010)	http://bejerano.stanford.edu/great/public/html/
RefFreeEWAS	(Houseman et al., 2014)	https://cran.r-project.org/web/packages/RefFreeEWAS/index.html
Genomation	(Akalin et al., 2015)	https://bioconductor.org/packages/release/bioc/html/genomation.html

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Patrick Tan (gmstang@duke-nus.edu.sg).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

GCEP Cohort

The Gastric Cancer Epidemiology Programme (GCEP) is a prospective endoscopic surveillance programme targeted at the Chinese population aged 50 years or above in Singapore, a country with an intermediate risk of GC. A pre-defined "high-risk" cohort comprising 3000 Chinese subjects of age > 50 years was recruited to explore the feasibility and yield of endoscopic screening in the Singapore population. GCEP aims to identify an optimum approach and cost-effective algorithm for targeted screening of GC in intermediate-risk populations. Besides generation of clinical data, biological materials were also banked from GCEP patients both at initial recruitment and subsequent follow-up visits, including blood, gastric biopsies, and gastric juices (Figure S1A).

Tissues and clinical information (Table S1) used in this study were collected under the GCEP study, with patient informed consent and approvals from the Domain Specific Review Board (DSRB) of the National Healthcare Group and Centralised Institutional Review Board (CIRB) of Singapore Health Services. All tissues were obtained by conventional upper gastrointestinal endoscopy performed by specialist endoscopists at endoscopy centres using standard forward viewing videoendoscopes. Six biopsies including 5 gastric mucosal biopsies and 1 cardia biopsy were taken for histology examination according to the updated Sydney System for classification and grading of gastritis. An additional 3 biopsies were taken from the antrum, body and cardia, and kept in liquid nitrogen for correlative studies.

Biopsy samples were assessed by two independent experienced pathologists for the degree of chronic gastritis, the presence of *H. pylori* (Hp) infection, the presence of gastric atrophy (GA), and intestinal metaplasia (IM). Histologic findings were scored using the updated Sydney System classification. IMs was classified as mild, moderate or marked, based on hematoxylin and eosin (H&E) staining. The first 331 moderate or marked IM cases were also classified into Type I (complete IM), Type II (incomplete) and Type III (incomplete) IM by H&E/HID/AB staining (I-36%; II-61%, III-2%). Because IMs are heterogeneous, classifications were based on the predominant type of IM. Degrees of dysplasia were diagnosed by the revised Vienna classification and classification of carcinomas were performed according to the WHO classification of tumors. A total of 2987 subjects (51% male, mean age, 59.9 ± 7.0) were recruited under the GCEP study, of which about half (52%) had IM at baseline. Out of 1553 subjects with IM, 725 (47%) had GA while the remaining 828 (53%) either did not have GA or lacked GA assessment. Conversely, among 870 GCEP subjects with GA, 725 (83%) had IM while the remaining 145 (17%) did not. In the GCEP cohort, 67% of subjects were Hp serology positive (1814/2699), and at baseline examination 22% (657/2980) exhibited active Hp infection based on histology analysis. After 5-7 years' surveillance, there were 13 high grade dysplasias (HGD) and 8 early gastric cancers (GC). HGD and early GC were classified as early gastric neoplasia (EGN) which was the endpoint of the study.

Sample Selection

For the current study, IM cases of type I and II were selected based on the following criteria: 1) subjects must be Hp serology positive; 2) have IM at antral sites; 3) the average severity of IM at the antrum must be at least moderate ($\geq 30\%$ area); 4) IM presence at both the antral lesser curve and antral greater curve. Using this criteria, 25 type I and 54 type II IMs were selected. All 8 type III IMs were also included. IM biopsies from the body and cardia sites for 28 of these subjects were also included for comparative analyses. Mild IM biopsies were selected from GCEP subjects who were Hp serology positive, free of EGN and any other cancers, had mild IM, no GA, no dysplasia, and no polyps based on at least two surveillance endoscopies. The chosen samples were all from the antral stomach. Normal gastric mucosae biopsies were selected from GCEP subjects who were Hp serology positive, free of EGN and any other cancers, no IM, no GA, no dysplasia, and no polyps based on at least two surveillance endoscopies. Both mild IM and normal gastric biopsies were frequency matched to IM cases based on age and gender.

Cell Lines

HFE145 cells were a gift from Dr. Hassan Ashktorab, Howard University. GES-1 cells were a gift from Dr. Alfred Cheng, Chinese University of Hong Kong. YCC3 cells were a gift from Yonsei Cancer Centre, South Korea. NCC24 cells were obtained from Korean Cell Line Bank. IM95, MKN1, and TMK1 cell lines were obtained from the Japan Health Science Research Resource Bank. Cell line identities were confirmed by STR DNA profiling performed at the Centre for Translational Research and Diagnostics, NUS. All cell lines were negative for mycoplasma contamination as assessed by the MycoAlertTM Mycoplasma Detection Kit (Lonza) and the MycoSensor qPCR Assay Kit (Agilent Technologies).

METHOD DETAILS

Selection of Genomic Targets

We performed comprehensive literature and database reviews to identify genes biologically and clinically relevant to GC, including genes involved in key oncogenic signaling pathways, oncogenes, tumor suppressor genes and genes from kinase and chromatin

remodeler families. Data repositories including The Cancer Genome Atlas (<https://cancergenome.nih.gov/>; TCGA), cBio Portal (<http://www.cbioportal.org/>), Catalogue of Somatic Mutations in Cancer (COSMIC; <http://cancer.sanger.ac.uk/cosmic>), and mutations of kinases in cancer (<http://strubiol.icr.ac.uk/extra/mokca/>; MokCa) were mined to identify genes exhibiting recurrent somatic mutations in GC. Agilent SureSelect E-array software was used to design unique RNA baits for the 766 gene panel. Biotinylated RNA baits were synthesized by Agilent for use with the SureSelect Target Enrichment system (Agilent, USA).

DNA Extraction, Target Capture and Library Construction

Genomic DNA from tissues and blood samples were extracted using the DNeasy kit (Qiagen), and concentrations of extracted DNA determined by Qubit. For library preparation, 100 ng of DNA was sheared using a Covaris S2 instrument (Covaris, USA) to a size distribution (150 to 200 bp) which is optimal for target enrichment. Size-selected, adapter-ligated libraries were then incubated with the custom-designed Sureselect baits for 24 hr. The captured libraries were then PCR amplified and pooled at equimolar concentrations with unique index tags. The libraries were then sequenced using an Illumina HiSeq 2500 sequencer (Illumina, San Diego, California, USA). For molecular barcoding, the Agilent Sureselect XT-HS target enrichment assay was used. 10 ng of sample was sheared with Covaris and the samples were end repaired and A-Tailed. Molecular barcoded adaptors were then ligated and amplified for targeted capture against custom designed sureselect baits. The captured libraries were then PCR amplified and sequenced using an Illumina Hiseq 3000.

Read Mapping and Somatic Mutation Detection

Raw sequencing data were aligned to the reference human genome (hs37d5) using Burrow-Wheeler Aligner Maximal Exact Match (bwa-mem) (Li and Durbin, 2009) and processed with Picard and GATK (DePristo et al., 2011). Alignments were sorted and converted to BAM format. The Mutect2 (Cibulskis et al., 2013) algorithm was implemented to identify somatic SNVs and indels for matched tumor-blood samples. All candidate SNVs and indels with at least 10 somatic variant-supporting reads were further filtered for OxoG artifacts to retain only highly confident predictions. Similar filters have been used by others (Taylor-Weiner et al., 2016). Functional annotation of mutations was performed with Oncotator (<http://www.broadinstitute.org/cancer/cga/oncotator>). Mutational significance analysis was conducted using MutsigCV (Lawrence et al., 2013), which considers gene expression, replication time and chromatin state when calculating the background mutation rate. Mutational spectra were visualized using the SomaticSignatures R package (Gehring et al., 2015). For samples with molecular barcodes, Picard's MarkDuplicates tool using molecular barcode-aware settings was used to flag PCR duplicates (reads mapping to the same start position and having the same molecular barcodes).

Validation of Somatic Variants with Ion Torrent Resequencing

Custom primers were designed using the Ion Ampliseq designer software. Briefly, 10 ng of DNA was used for library preparation. Targets of interest were PCR amplified with the appropriate primer pool, partially digested and adaptor ligated. The adaptor ligated libraries were then purified by AMPure beads and quantified by a Bioanalyser instrument (Life Technologies). Samples were then pooled and template prepared for sequencing using the Ion PGM 200 Sequencing Kit (Ion Torrent). Pooled samples were loaded on the 318 chip and sequenced on the Ion Torrent PGM 200 system. Data processing was performed using the Torrent Suite v3.6.5.

Copy Number, Telomere Length and Hp Density Estimation

Copy number levels were estimated using the GATK4 workflow for allelic copy number variation (ACNV; <https://software.broadinstitute.org/gatk>). Heterozygous germline SNPs in blood samples and the aggregate read counts at these sites in gastric tissues were identified. Somatic copy-number changes were estimated by segmenting the genome into regions of constant copy number and providing an estimate for the copy ratio and minor allele fraction in those regions. A subset of the predicted copy number alterations was validated using Illumina Infinium Assays (see below). ABSOLUTE (Carter et al., 2012) was used to infer whole genome doubling.

Telomere lengths were estimated using Telseq (Ding et al., 2014) on sequencing BAM files. Briefly, we calculated the proportion of sequencing reads containing greater than seven telomeric repeats (TTAGGG), relative to the number of GC-adjusted sequencing reads. The relative count was then multiplied with a GC-normalized genome length constant, resulting in an estimated telomere length.

Whole genome sequence data of Hp strain 26695 (NC_000915.1) was obtained from NCBI. We combined the human and Hp genomes to build a hybrid reference genome. The raw paired-end sequencing reads were aligned to the hybrid reference genome using bwa mem. Hp density was estimated as the proportion of reads mapping uniquely to the Hp genome with no mismatches relative to the number of genomic sequencing reads.

Illumina Infinium Assay (Copy Number)

DNA samples from blood and tissue (200 ng) were subjected to denaturation, neutralization and then isothermal amplification to prevent any amplification bias. The amplified DNA was then fragmented, precipitated and hybridised to a Infinium OmniExpress-24 v1.2 BeadChip (Illumina, Inc). The Beadchips were later washed to remove any unhybridised and non-specifically hybridised DNA and prepared for staining and extension. Single-base extension of the oligos on the BeadChip, using the captured DNA as a template

was performed. The bead chips were then imaged using an Illumina HiScan. DNA array data was processed using Illumina Genome Studio to compute the log2R ratio (LRR) and the B-allele frequency (BAF). ASCAT v2.0 ([Van Loo et al., 2010](#)) was used to identify regions of copy number gains or losses.

Illumina Infinium HD Methylation Assay

500 ng of genomic DNA was subjected to bisulfite conversion using the EZ DNA Methylation-Gold kit (Zymo Research). The bisulfite-converted genomic DNA was then profiled on Infinium HumanMethylation450 Beadchip (Illumina, Inc). All protocols were performed according to the manufacturer's instructions. Raw intensity data from the HumanMethylation 450K array were analyzed in R using the minfi ([Aryee et al., 2014](#)) package to obtain beta values (ratio of methylation). Samples were normalized using the functional normalization function. Small methylated regions and large methylation blocks were identified using the bumphunter and blockfinder techniques ([Aryee et al., 2014](#)), respectively. Regions were filtered by size using a minimum cut-off at least 4 CpGs and blocks were filtered to include only those larger than 100 kb. Significance was assigned based on permutation testing with a cutoff of adjusted p value of less than 0.05 (family-wise error rate). Analysis of functional enrichment of differentially methylated regions was performed using the Genomic Regions Enrichment of Annotations Tool (GREAT) ([McLean et al., 2010](#)). Cellularity of IMs was estimated using the RefFreeEWAS tool ([Houseman et al., 2014](#)) using normal samples as control.

Immunohistochemistry and Stains

PAS/AB

Sections were stained with Periodic Acid-Schiff/Alcian Blue (PAS/AB). Briefly, slides were immersed in AB solution for 15 min and subsequently treated with periodic acid followed by Schiff's reagent for 10 min each. Evaluation of neutral mucins and sialomucins was performed after counterstaining with hematoxylin for 1 min.

HID/AB

Sections were stained with High Iron Diamine/Alcian blue (HID/AB). Briefly, slides were immersed in HID solution for 18 to 20 hr, at 23–25°C. Slides were then rinsed with deionised water and stained with 1% Alcian blue (pH 2.5) for two min. The evaluation of sialomucins and sulphomucins was done in sections without haematoxylin counterstain.

c-MYC, p53, and ARID1A :Deparaffinized and rehydrated sections were treated with primary antibodies to p53 (DAKO; clone DO-7), c-MYC (Abcam, clone Y69), or ARID1A (Abcam, ab182561). Localization of antibodies was performed with peroxidase labelled Streptavidin-biotin system (DAKO, Glostrup, Germany) with 3-3' Diaminobenzidine (DAKO, Glostrup, Germany) according to the manufacturer's instructions and the slides were then counterstained with haematoxylin. Appropriate positive and negative control was run concurrently for each tested antibodies. Manual DAKO Envision reagents were used for visualization.

Giemsa

Unstained sections were deparaffinized and placed in Giemsa solution for 10 min (Benchmark Special Stainer, Ventana, USA). This was followed by dehydration and differentiation in ethanol and histoclear. Hp bacteria were detected as staining dark blue in colour.

Laser Capture Microdissection (LCM)

Gastric FFPE tissue blocks were sectioned at 5 µm and mounted on slides. IM cells in the hematoxylin and eosin- stained section were marked by a trained pathologist. Prior to laser capture, an unstained section was stained with the Arcturus™ Paradise™ PLUS FFPE LCM Staining Kit following the protocol indicated in the user manual. IM regions were dissected and collected on CapSure Macro LCM Caps. Genomic DNA was then extracted from the caps using the PicoPure™ DNA Extraction Kit in 20 µl of extraction buffer with periodic vortexing. Extracted DNA was then quantified by Qubit and used for Sanger sequencing. We also performed LCM analysis on an additional 7 IMs followed by molecular barcoded targeted sequencing (766 genes) and identified clonal (MAF>10%) TP53 or FBXW7 mutations in two samples. These mutations (TP53 C176Y and FBXW7 R505C) are likely to be somatic mutations because they are known oncogenic mutations observed in multiple cancer types but not in any blood samples profiled on the same platform in this study (n=148). FBXW7 R505C mutations were also observed in the non-LCM dissected IM samples ([Figure 1B](#)).

Sanger Sequencing of LCM Material

50 ng of FFPE gDNA was amplified by using hot start Taq colorless master mix (Promega) in 20 µl PCR reaction mixtures. The following primer sequences (Forward –CAAGTGGCTCCTGACCTGG and Reverse – CCTCATTTGGGCCTGTGTT) were used. PCR conditions were set as: 95°C for 5 min, followed by 40 cycles of 95°C for 50 s; 58°C for 50 s; 72°C for 60 s and a final extension at 72°C for 10 min. Purified PCR products were sequenced bi-directionally on an ABI 3730 DNA analyzer (3730s, life technologies).

siRNA Transfection and Cisplatin Treatment

ON-TARGETplus FBXW7 siRNA SMARTpool and ON-TARGETplus Non-targeting Pool(Dharmacon/Thermo Fisher Scientific) were transfected onto cells (2X10⁵) at 50 nM using Dharmafect 1 transfect reagent in 6-well plates, according to the manufacturer instructions. Cisplatin (2 µM) (Sigma) was added to the cells 48 hr after transfection, and cells were collected 72 hr after transfection for quantitative RT-PCR and Western Blot analysis.

Molecular Cloning and Stable Cell Line Establishment

The constructs FBXW7 WT/MUT1(R505C)/MUT2(R465C) were amplified from cDNAs of HFE145/IM95/MKN1 cells using the following primers: FBXW7-202-F- 5'-TTTAGATCTATGAATCAGGAAC TGCTCTGT-3', FBXW7-202-R:5'-TTTGTCGACTCACTT CATGTCCACATCAAAGT-3', and cloned into the pHR'CMVGFPIRESWSIn18 based vector (gift from Dr. Li Shang, Duke-NUS). Plasmids were transfected into HEK293T cells. Cells were infected and selected using hygromycin B (Nalgene) for two weeks.

Quantitative RT-PCR

Total RNA was extracted using RNeasy Mini Kit (Qiagen), and reverse transcribed using iScript Reverse Transcription Supermix for RT-qPCR (Biorad). cDNA was amplified using SYBRGreen PCR Master Mix (Applied Biosystems). Gene expression fold changes were normalized to GAPDH. Primers sequences are as follows: FBXW7: F-5'-CACGTTGCAGGGG CATACTA-3', R: 5'-CTCCACATCCAAACACGGA-3'; GAPDH: F-5' CCAGGGCTGCTT TAACTC 3', R-5' GCTCCCCCTGCAAATGA 3'.

Western Blotting

Cells were harvested and lysed in RIPA buffer (Sigma) for 10 min on ice. Protein concentrations were measured using the Pierce BCA protein assay (Thermo Scientific). Cell lysates were heated at 99°C for 10 min in 4xSDS sample buffer, and 40 µg per sample was loaded for assay. Anti-phospho-Histone H2A.X (Ser139) (1:100; clone JBW301, Millipore) and GAPDH (1:3000; 60004-1-Ig, Protein-tech Group) antibodies were used to probe the lysate.

Cell Proliferation Assay

2x10³ FBXW7 WT/MUT1(R505C)/MUT2(R465C) stably integrated TMK1 cells were seeded into 96-well plates. Cell proliferation was measured the next five days (Day 0-4) by the WST-8 assay (Cell Counting Kit-8, Dojindo). 10 µl of WST-8 solution was added per well and the absorbance reading was measured at 450 nm after 2 hr of incubation in a humidified incubator.

ChIP-seq Signal Aggregation

EZH2 and H3K27me3 ChIP-seq data (bam files) for H1-hESC cells were downloaded from the ENCODE data portal. EZH2 ChIP-seq data for HFE145 and H3K27me3 data for SNU484 were obtained from in-house data. Average ChIP-seq signals surrounding DMR sites in IM were calculated using the genomatix ([Akalin et al., 2015](#)) bioconductor package.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis was performed and graphics produced using R version 3.2.6. All hypothesis tests were 2-sided and statistical tests used are specified in the Results and Figure legends. For clinical associations, only factors significant in univariate analysis were used for multivariate analysis.

DATA AND SOFTWARE AVAILABILITY

Sequencing data has been deposited into sequence read archive (SRA) database under accession number PRJNA 383939. Methylation data has been deposited into Gene Expression Omnibus (GEO) database under accession number GSE103186. Software used in this study are noted in the [Method Details](#) section above and the [Key Resources Table](#).