

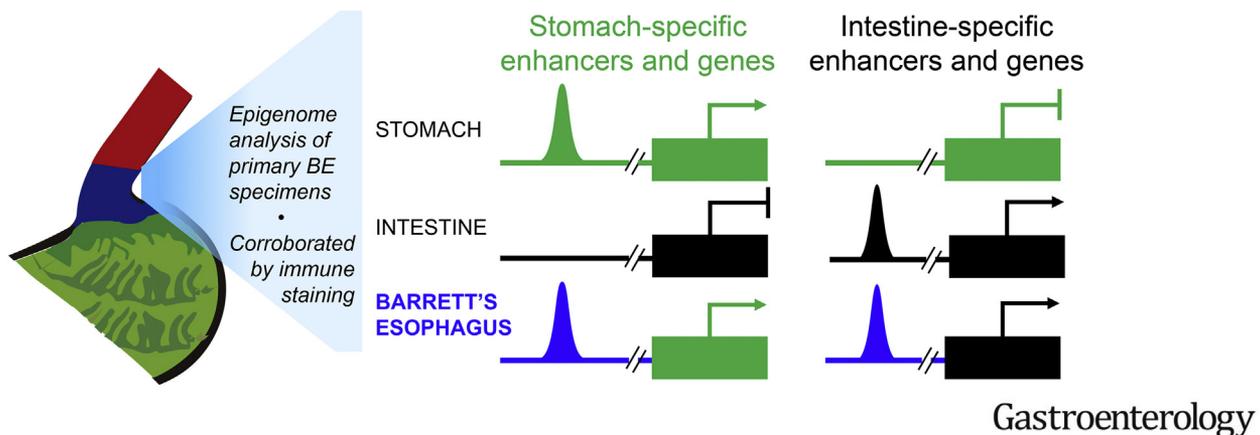


Hybrid Stomach-Intestinal Chromatin States Underlie Human Barrett's Metaplasia

Harshabad Singh,^{1,2} Kyungsik Ha,³ Jason L. Hornick,⁴ Shariq Madha,¹ Paloma Cejas,¹ Kunal Jajoo,² Pratik Singh,^{1,2} Paz Polak,⁵ Hwajin Lee,³ and Ramesh A. Shivdasani^{1,2,6}

¹Department of Medical Oncology and Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, Massachusetts; ²Department of Medicine Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts; ³Biomedical Knowledge Engineering Laboratory, Seoul National University, Seoul, Korea; ⁴Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts; ⁵Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, New York, New York; and ⁶Harvard Stem Cell Institute, Cambridge, Massachusetts

Co-existing gastric and intestinal chromatin states in Barrett's esophagus



BACKGROUND & AIMS: Tissue metaplasia is uncommon in adults because established *cis*-element programs resist rewiring. In Barrett's esophagus, the distal esophageal mucosa acquires a predominantly intestinal character, with notable gastric features, and is predisposed to developing invasive cancers. We sought to understand the chromatin underpinnings of Barrett's metaplasia and why it commonly displays simultaneous gastric and intestinal properties. **METHODS:** We profiled *cis*-regulatory elements with active histone modifications in primary human biopsy materials using chromatin immunoprecipitation followed by DNA sequencing. Mutations in Barrett's esophagus were examined in relation to tissue-specific enhancer landscapes using a random forest machine-learning algorithm. We also profiled open chromatin at single-cell resolution in primary Barrett's biopsy specimens using the assay for transposase-accessible chromatin. We used 1- and 2-color immunohistochemistry to

examine protein expression of tissue-restricted genes. **RESULTS:** Barrett's esophagus bears epigenome fingerprints of human stomach and intestinal columnar, but not esophageal squamous, epithelia. Mutational patterns were best explained as arising on the epigenome background of active gastric *cis*-elements, supporting the view that adjoining stomach epithelium is a likely tissue source. Individual cells in Barrett's metaplasia coexpress gastric and intestinal genes, reflecting concomitant chromatin access at enhancers ordinarily restricted to one or the other epithelium. Protein expression of stomach-specific mucins; CLDN18; and a novel gastric marker, ANXA10, showed extensive tissue and subclonal heterogeneity of dual stomach-intestinal cell states. **CONCLUSIONS:** These findings reveal mixed and dynamic tissue-restricted chromatin states and phenotypic heterogeneity in Barrett's esophagus. Pervasive intragland variation argues against stem-cell governance of this phenotype.

Keywords: Intestinal Metaplasia; Epigenetics; Tissue-Specific Epigenomes.

In Barrett's esophagus (BE), columnar mucosa of an intestinal type replaces the native stratified epithelium in the human distal esophagus.¹ Esophageal adenocarcinoma (EAC) arises almost exclusively within dysplastic BE foci, and its incidence in the West has increased nearly 8-fold in the last 4 decades.^{2,3} Because symptoms usually appear after EACs are inoperable, improved outcomes hinge on the prevention, detection, and monitoring of BE.^{1,3} Preventive strategies are, however, constrained by a limited understanding of BE origins and pathogenesis.

Historically, BE was thought to reflect trans-differentiation of squamous esophageal cells.^{4,5} Indeed, glands dissected from one BE specimen carried the same mutations in one nuclear (*CDKN2A*) and one mitochondrial (*COO*) gene as those found in adjoining stratified epithelium.⁶ Recent experiments in mouse esophagus reinforce this view,⁷ but the presence of significant gastric features in human BE specimens raises the alternative that BE reflects mucosal repair by stomach-derived columnar cells, which subsequently acquire ectopic intestinal character.^{8–11} This idea, difficult to prove in humans, is consistent with the similar metaplastic responses to chronic stomach or esophageal injury and with the strong similarities between esophageal and gastric adenocarcinomas.¹² A third possibility is that BE originates in esophageal submucosal glands (ESMGs),¹³ whose single-cell (sc) RNA profiles resemble those in BE.¹⁴ No animal model mimics BE perfectly, and rodents lack ESGMs; however, expansion of mouse fetal cellular rests,¹⁵ transitional squamocolumnar junction cells,¹⁶ or cells from the gastric cardia¹⁷ support non-squamous sources of mouse intestinal metaplasia (IM). Tissue-specific chromatin signatures can shed light on BE origins and the co-occurrence of gastric and intestinal properties.

Only a fraction of the millions of *cis*-regulatory elements in mammalian genomes is active in any tissue. Active sites have accessible chromatin, bind transcription factors (TFs), and bear characteristic covalent modifications such as methylated lysine 4 (H3K4me1/2) on histone 3.^{18,19} Each cell type uses thousands of distinct enhancers, and the specific combination constitutes a fingerprint.²⁰ Stretch or super enhancers carry dense active histone marks, control many lineage-restricted genes, and contribute disproportionately toward tissue signatures.^{21,22} Mutations that adult tissues accumulate from replication errors²³ tend to concentrate in regions that lack active enhancers, thus reflecting cell-of-origin (COO) chromatin landscapes for tumors and precancerous lesions.²⁴

Chromatin states might hence point to cellular sources of BE with more clarity than other methods possible in human tissues and, in individual cells, show the *cis*-regulatory basis for cellular diversity. We delineated the enhancer landscapes specific to human esophageal, stomach, and intestinal mucosae. BE specimens showed concomitant activation of enhancers specific to the intestine and stomach, but not

WHAT YOU NEED TO KNOW

BACKGROUND AND CONTEXT

Barrett's esophagus, the major risk factor for esophageal adenocarcinoma, has unclear chromatin underpinnings. We investigated the condition using epigenome analyses that reveal chromatin states and reflect tissue identity.

NEW FINDINGS

Barrett's esophagus is a unique tissue type in which individual cells manifest varying degrees of dual stomach and intestinal identity, both in chromatin states and marker gene expression.

LIMITATIONS

Human biopsy materials capture a limited window of time and only partially show the simultaneous and dynamic gastric-intestinal identity we identify in Barrett's esophagus.

IMPACT

Definition of novel mixed chromatin states and cell identities in Barrett's esophagus likely reflects the underlying plasticity of the Barrett's stem cell and provides new insights into adult stem cell biology, chromatin plasticity, and human metaplasias.

the esophagus, and the gastric epigenome best explained BE mutational backgrounds. In discrete BE cells resolved by sc assay for transposase-accessible chromatin (ATAC) sequencing (ATAC-seq), stomach- and intestine-restricted enhancers are coaccessible. Immunohistochemistry verified abundant coexpression of gastric and intestinal markers in the same cells. Together, these findings show a unique hybrid chromatin state as the basis for a human metaplasia with simultaneous gastric and intestinal cell identities and pervasive subclonal heterogeneity.

Methods

Biopsy Samples for Fixed-Tissue Chromatin Immunoprecipitation Sequencing and Single-Cell Assay for Transposase-Accessible Chromatin Sequencing

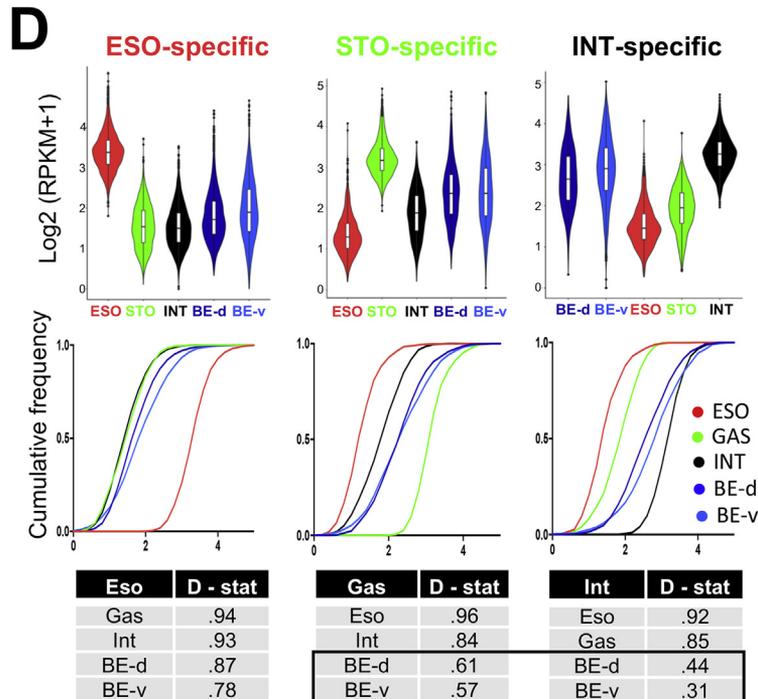
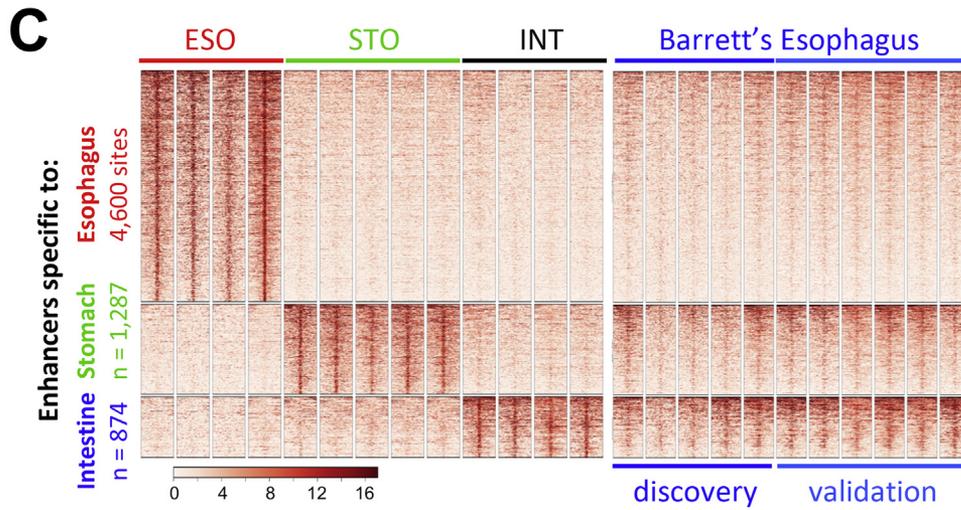
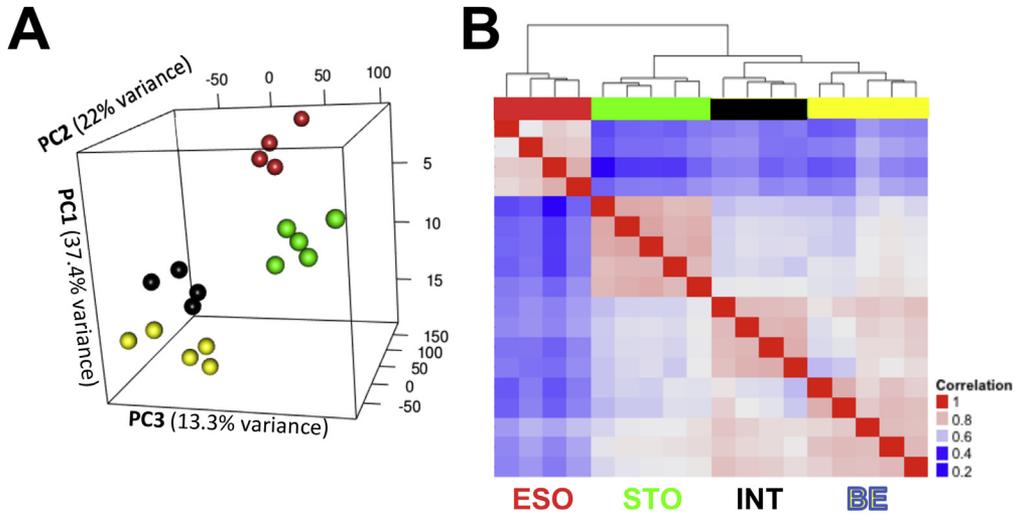
For fixed-tissue chromatin immunoprecipitation sequencing (FiT-seq), endoscopic mucosal biopsy specimens were fixed in formalin and embedded in paraffin. Institutional review boards (IRBs) approved research on anonymized samples from regions

Abbreviations used in this paper: ATAC, assay for transposase-accessible chromatin; BE, Barrett's esophagus; ChIP, chromatin immunoprecipitation; COO, cell of origin; EAC, esophageal adenocarcinoma; ESCC, esophageal squamous cell carcinoma; ESGM, esophageal submucosal gland; FiT, fixed-tissue chromatin immunoprecipitation; GIM, gastric intestinal metaplasia; IM, intestinal metaplasia; IRB, institutional review board; kb, kilo base pairs; mAb, monoclonal antibody; Mb, mega base pairs; mRNA, messenger RNA; sc, single cell; seq, sequencing; TF, transcription factor; tSNE, *t*-distributed stochastic neighbor embedding.

 Most current article

© 2021 by the AGA Institute
0016-5085/\$36.00

<https://doi.org/10.1053/j.gastro.2021.05.057>



of healthy tissue and BE (Dana-Farber Cancer Institute IRB 05-157). Tissue sections were chosen for >90% mucosal content or dissected to achieve this proportion (Supplementary Figure 1B). For BE, we required a ≥ 1 -cm esophageal segment with IM, including goblet cells,²⁵ and no dysplasia. For normal stomach, we used fundic (oxyntic) mucosa, avoiding the cardia. Ileum was used for normal small intestine. Histologically normal squamous epithelium from mid- and lower esophageal biopsy specimens were pooled from each patient. All samples in the study are from independent individuals.

For scATAC-seq, fresh mucosal specimens were obtained from patients undergoing endoscopic surveillance (Partners Healthcare IRB 2015P002409). BE was identified by narrow-band imaging,²⁶ and biopsy specimens were obtained from the gastric corpus and duodenum of the same individuals. Samples were frozen immediately at -80°C in Bambanker freezing medium, and parallel biopsy samples were examined by histopathology. We performed scATAC after confirming nondysplastic BE carrying goblet cells (eg, Figure 4B).

Chromatin Immunoprecipitation Sequencing

Tissue sections processed according to a published FiT-seq protocol²⁷ were precipitated using 10 μL of H3K4me2 antibody (Millipore, 07-030). Libraries were prepared using ThruPLEX DNA-seq kits (Rubicon Genomics) and sequenced on a NextSeq instrument (Illumina). Computational analysis and delineation of tissue-restricted enhancers are described in the Supplementary Methods.

Principal Coordinates and Cell-of-Origin Analyses

To obtain regional feature densities, we first divided autosomes into 1-mega base pair (Mb) regions, excluding centromeres, telomeres, and low-quality sequence tags. We calculated somatic mutation numbers per 1-Mb region and first used regional mutation densities as inputs for principal coordinates analysis to represent intersample differences in mutation frequency distribution (Figure 3A). Samples were located in 3-dimensional space using a dissimilarity matrix, calculated from Pearson correlation coefficients. For feature selection and COO analysis, we used a nonparametric machine-learning method (random forest) to predict the value of a continuous response variable by using ensemble decision trees.²⁸ Details and the approach to ascertain statistical significance are described in Supplementary Methods.

Single-Cell Assay for Transposase-Accessible Chromatin Sequencing

Nuclei were obtained from frozen biopsy specimens by suspending the tissue in 1 mL buffer containing 10 mmol/L Tris-HCl pH 7.4, 10 mmol/L NaCl, 3 mmol/L MgCl₂, 0.1% NP-40, 0.1% Tween-20, and 0.01% digitonin, followed by homogenization in a douncer using 10 loose and 10 tight pestle strokes. The resulting cell suspension was passed through a 70- μm filter, and debris was removed over an iodixanol gradient or by flow cytometry using positive selection with 4',6-diamidino-2-phenylindole, both to equal effect. Isolated nuclei were counted manually using a hemacytometer, resuspended in diluted nuclear stock solution, and processed according to the 10x Genomics protocol (<https://www.10xgenomics.com/solutions/single-cell-atac/>). Computational analyses and delineation of tissue-restricted enhancers are described in the Supplementary Methods.

Immunohistochemistry

Formalin-fixed, paraffin-embedded tissue sections (4 mm thick) were baked at 37°C overnight, deparaffinized, and rehydrated (100% xylene $\times 4$ for 3 minutes each, 100% ethanol $\times 4$ for 3 minutes each, and running water for 5 minutes). Sections were treated with 1.5% hydrogen peroxide in methanol for 10 minutes, washed under running water for 5 minutes, and placed in a pressure cooker (Biocare Medical) at 120°C in Target Retrieval Solution (pH 6.1 citrate buffer) (DAKO). After cooling and transfer to Tris-buffered saline, slides were incubated for 40 minutes at room temperature with anti-CDX2 mouse monoclonal antibody (mAb) (BioGenex clone CDX2-88, 1:100) or ANXA10 (Novus, NBP1-90156, 1:1,000) anti-serum, followed by secondary antibody (Envision+ mouse [DAKO] or PowerVision AP rabbit [Leica Biosystems]) for 30 minutes. Stains were developed using 3,3'-diaminobenzidine (brown product). For 2-color immunohistochemistry, sections were then incubated for 40 minutes at room temperature with mouse MUC5AC (Abcam clone 45M1, 1:10,000) or MUC2 (Vector Laboratories clone Ccp58, 1:400) mAb, rabbit TFF1 mAb (Cell Signaling Technology clone D2Y1J, 1:1000), or rabbit CLDN18 anti-serum (Sigma, 1:500), followed by secondary antibody (PowerVision AP mouse or rabbit, Leica Biosystems) for 30 minutes. Slides were incubated in a humid chamber

Figure 1. BE specimens carry intestinal and gastric, but not esophageal, enhancers. (A) Principal component analysis of the highest quintile of variable enhancers. Esophageal (red, $n = 4$) and gastric (blue, $n = 5$) epithelia form discrete clusters. Enhancers marked in BE (yellow, $n = 5$) are distinct from those in the native esophageal mucosa and most similar to those in the intestine (black, $n = 4$). (B) Pearson correlation coefficients among active enhancers (H3K4me2 peaks > 2 kb from transcription start sites). In unsupervised hierarchical clustering, the BE enhancer profile ($n = 5$) is most similar to intestinal (INT) mucosa ($n = 4$) and related to stomach (STO) ($n = 5$), but not to esophageal (ESO), squamous epithelium ($n = 4$). (C) H3K4me2⁺ enhancers marked uniquely in stratified esophageal (4600 sites), stomach corpus (1287 sites) or intestinal (874 sites) epithelium. Five original (discovery set) and 6 additional (validation set) BE samples show H3K4me2 at intestinal and gastric, but not at esophageal, enhancers. Heatmap scale, 0–16 units. (D) Quantitation of the H3K4me2 FiT-seq data in C, represented as violin plots and cumulative frequencies. Because the large volume of data ensures significant P values for differences across all sample pairs, we applied the Kolmogorov-Smirnov test to measure similarities and differences across enhancer signatures (D statistic, where lower values reflect greater similarity). Each quantitative measure reveals the BE enhancer signature as similar to those of the stomach and intestine, but not the esophagus. BE-d, BE-discovery; BE-v, BE-validation; D-stat, D statistic; GAS, gastric; PC, principal component.

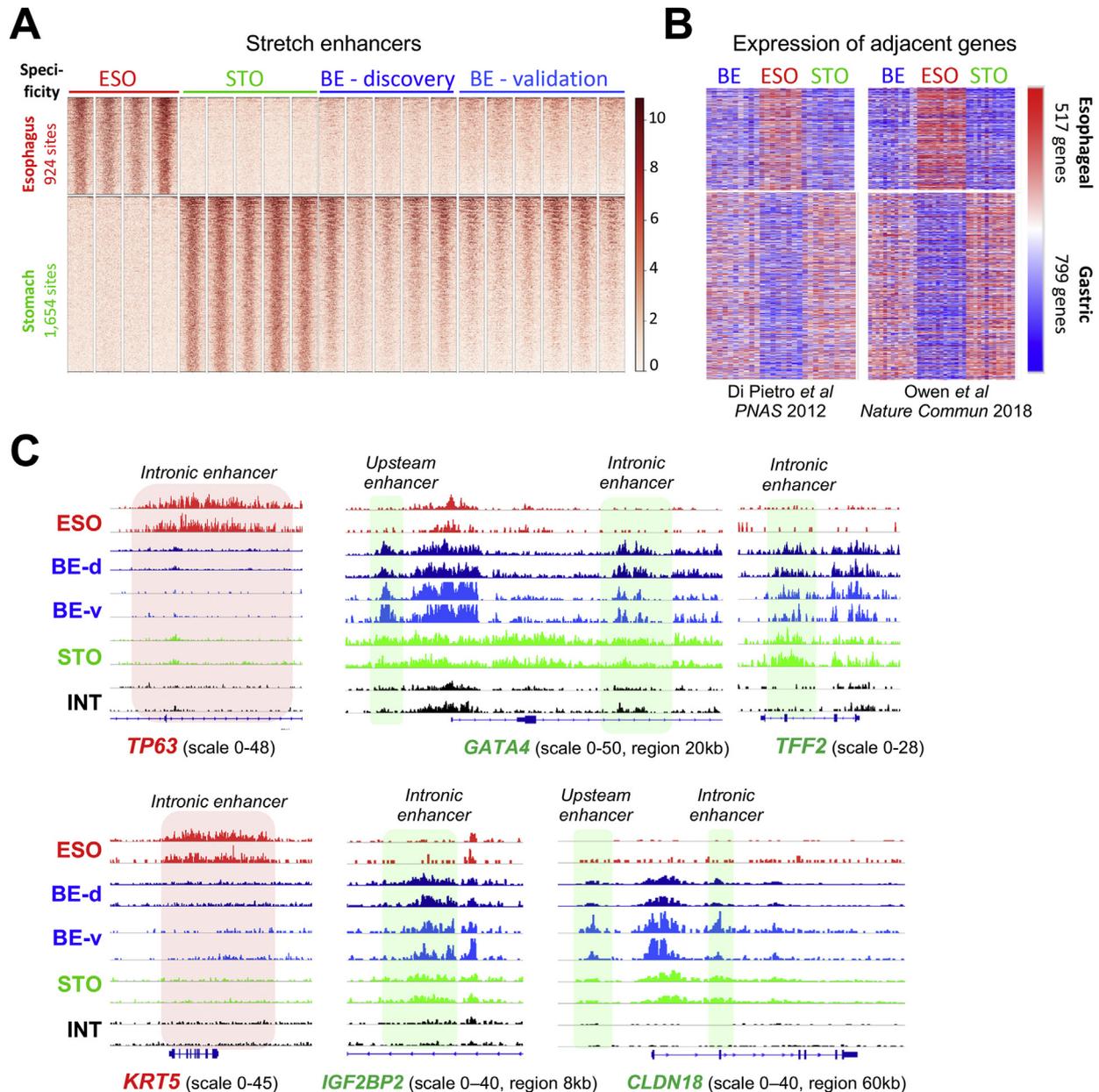


Figure 2. Stomach stretch enhancer activity and a gastric gene signature in BE. (A) H3K4me2 ChIP signals in BE specimens at 924 esophagus- (ESO) and 1654 stomach- (STO) specific stretch enhancers, showing marks at stomach but not signature esophagus regions. Esophagus, $n = 4$; stomach, $n = 5$; BE-d, $n = 5$; BE-v, $n = 6$. (B) Relative mRNA levels of 517 genes encoded <100 kb from esophagus-specific and 799 genes encoded <100 kb from stomach-specific stretch enhancers, determined from gene expression data sets reported in references 14 and 34, respectively. Blue indicates low and red indicates high mRNA expression. Enhancer marks in BE correlate with expression of nearby stomach, and not esophagus, genes. (C) Data tracks from H3K4me2 ChIP-seq at enhancers in esophagus, stomach, intestine (INT), and BE specimens. *TP63* and *KRT5* loci are selectively marked in esophagus, as expected, but not in BE. Conversely, enhancers near *GATA4*, *TFF2*, *IGF2BP2*, and *CLDN18* are marked in stomach and BE samples, but not in esophagus or intestine.

and rinsed in between with Tris-buffered saline. Sections were developed using Permanent Red and counterstained with Mayer's hematoxylin.

Transcript and Genome Profiling

All chromatin immunoprecipitation (ChIP), ATAC, and scATAC data are deposited in the Gene Expression Omnibus under accession number GSE97178.

Results

A Gastric Enhancer Fingerprint in Human Barrett's Esophagus

To discern active enhancer fingerprints in human mucosal biopsy specimens, we applied FiT-seq,²⁷ a method to map modified histones in formalin-fixed, paraffin-embedded tissues. Among active enhancer marks,^{18,29} H3K27ac resists robust immunoprecipitation in FiT-seq,

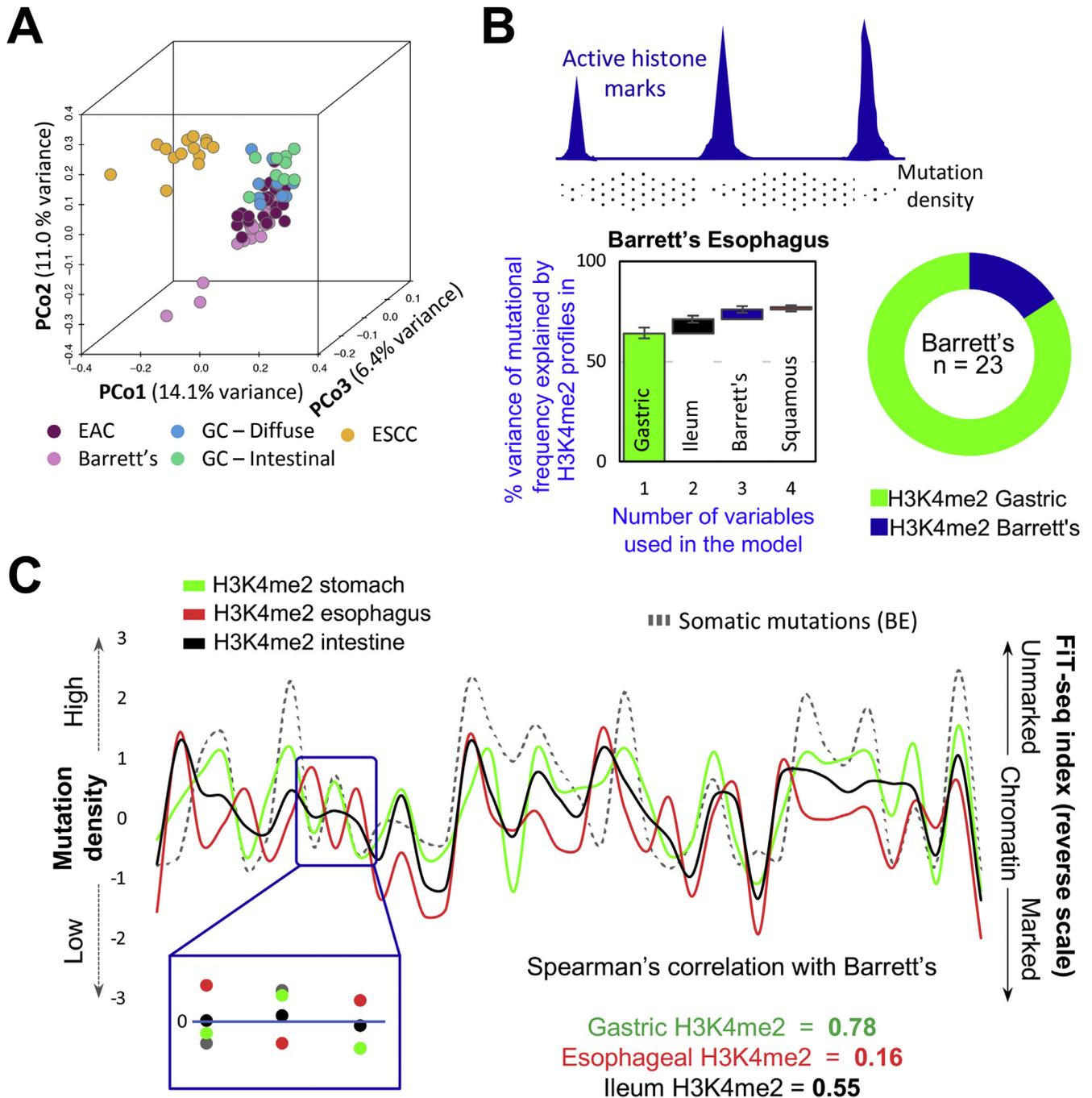


Figure 3. Mutational profiles in BE imply a gastric, not esophageal, origin. (A) Principal coordinates analysis of genome-wide somatic mutation data from BE, ESCC, EAC, and gastric cancers (GC) of the diffuse or intestinal types. BE mutational profiles overlap substantially with those in EAC and GC and differ from those in ESCC. (B) (Top) Illustration of the inverse relation of somatic mutation density and enhancer chromatin, likely reflecting differential access to DNA repair. (Bottom) A random forest machine-learning approach showed that the enhancer H3K4me2 landscape of normal gastric epithelium was the best predictor of mutational variance in 23 BE samples, with small incremental contributions from other landscapes. The algorithm was run on grouped (graph: the first bar depicts the largest contributor, and subsequent columns represent the incremental contributions of additional variables) or individual (donut) BE samples. In both analyses, the gastric signature best explained mutations found in BE. (C) A representative 38-Mb region showing correlation between mutation frequency in BE (left y-axis) and absence of enhancer marking in normal gastric chromatin (right y-axis), whereas absence of enhancer marks in stratified esophageal and intestinal epithelium correlated poorly. PCo, principal coordinate.

which identifies H3K4me2 reliably.²⁷ H3K4me2 and H3K27ac mark essentially the same enhancers in mouse intestines (Supplementary Figure 1A), and FiT-seq for

H3K4me2 largely identified the same enhancers in human pancreatic endocrine tumors as ChIP-seq for H3K27ac in frozen tumors.³⁰ FiT-seq for H3K4me2 gave robust signals

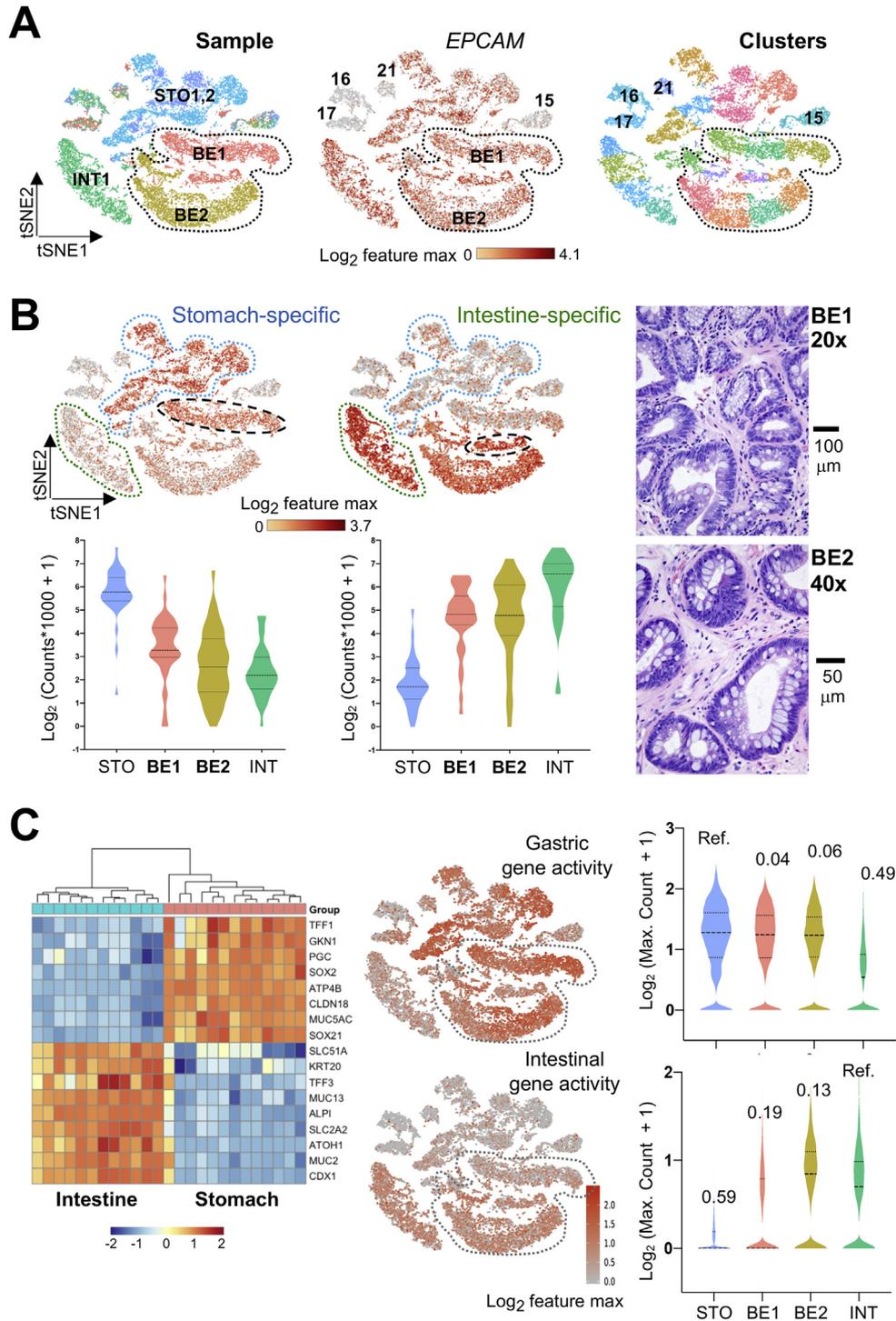


Figure 4. The sc resolution of hybrid gastric/intestinal chromatin states in BE. (A) All informative cells (after filtering by Cell Ranger criteria) projected on a tSNE plot to reduce data dimensionality. Cells are color coded by sample (left, BE1 and BE2: BE cases 1 and 2; STO1,2: stomach cases 1 and 2; INT1: intestine) or graph-based cell clusters (right). ATAC signals at the *EPCAM* promoter identify epithelial cells. BE1 and BE2 cells are demarcated. Nonepithelial cells from all samples (clusters 15–17 and 21) showed ATAC signals at *VIM* and *CD45* (Supplementary Figure 3D). (B) Distribution of single cells from BE (histology shown to the right), INT (green outline), and STO (blue outline) specimens in graph-based scATAC-seq clusters. Aggregate signals for open chromatin at stomach- (n = 37) or intestine- (n = 32) restricted enhancers are projected onto the tSNE plot, showing coactivity of tissue-specific regions throughout BE1 and BE2 (bottom right epithelial clusters), with the 2 BE1 subclusters (dashed black ovals) showing relative enrichment of stomach or intestine enhancers. Violin plots represent the average signal on each enhancer across all epithelial cells in the indicated specimens. (C) Aggregate ATAC scores at panels of classic intestinal and gastric marker genes (left, RNA-seq data from reference 14; blue indicates low and orange indicates high RNA expression) projected onto the tSNE plot. Single cells in both BE samples (gray dotted space) show extensive coactivity and signals (log-transformed maximum counts per cell) are quantified on the right, with the similarities across samples estimated using *D* statistics from the Kolmogorov-Smirnov test (noted above each violin plot; lower values reflect greater similarity, and STO and INT samples serve as references). max, maximum; Ref., reference.

in mucosal biopsy specimens from healthy esophagus ($n = 4$), gastric corpus ($n = 4$), and ileum ($n = 5$) (Supplementary Figure 1B and C and Supplementary Table 1). Considering marked regions >2 kilo base pairs (kb) from transcription start sites as putative enhancers, we used k -means clustering of the top quintile of variable regions to identify tissue-restricted enhancers (Supplementary Figure 1D). These sites were highly correlated with tissue-specific gene expression (Supplementary Figure 1E), implying that they represent active *cis*-elements. In the following, we also show high overlap between H3K4me2⁺ sites identified by FiT-seq and open chromatin identified in assays for ATAC-seq (Supplementary Figure 4A).

Principal component analysis of the top quintile of variable enhancers (Figure 1A) and inspection of tissue-specific loci (Supplementary Figure 1C) readily distinguished the 3 epithelia. By unsupervised clustering of global enhancer profiles, BE samples ($n = 5$) most resembled intestinal epithelium (median correlation coefficient, 0.73; range, 0.59–0.79) (Figure 1B), least resembled stratified esophageal epithelium (median coefficient, 0.50; range, 0.44–0.60), and showed considerable overlap with gastric enhancers (median coefficient, 0.67; range, 0.58–0.73). Among enhancers specific to normal esophagus, stomach, or intestine (Figure 1C), BE specimens carried not only the expected intestinal signature but also clear marking at stomach enhancers and barely any esophageal squamous signature (Figure 1C and D). Trace esophageal enhancer signals, higher than the background in stomach or intestine, likely reflect our manual dissection from paraffin blocks where BE abuts squamous tissue.

FiT-seq on 6 additional nondysplastic BE samples reproduced these findings (Figure 1C and D). Enhancer landscapes were similar in the discovery and validation cohorts (Supplementary Figure 1F), and sites detected in these 11 BE specimens approached saturation in a cumulative analysis (Supplementary Figure 1G). Thus, the cohort is sufficient to capture most of the active BE epigenome, showing mixed gastric and intestinal features. Stomach enhancers active in BE were enriched for GATA- and SOX-family TF motifs, which are known to regulate stomach genes^{31,32} (Supplementary Figure 1H). Because some studies suggest that H3K4me2 may reflect gene transcription,³³ we assessed stomach-selective marks that lie within introns ($n = 720$) or in intergenic regions ($n = 344$). Both groups showed comparable H3K4me2 in BE samples (Supplementary Figure 2A), indicating that the underlying signal identifies bona fide regulatory activity and not transcription per se.

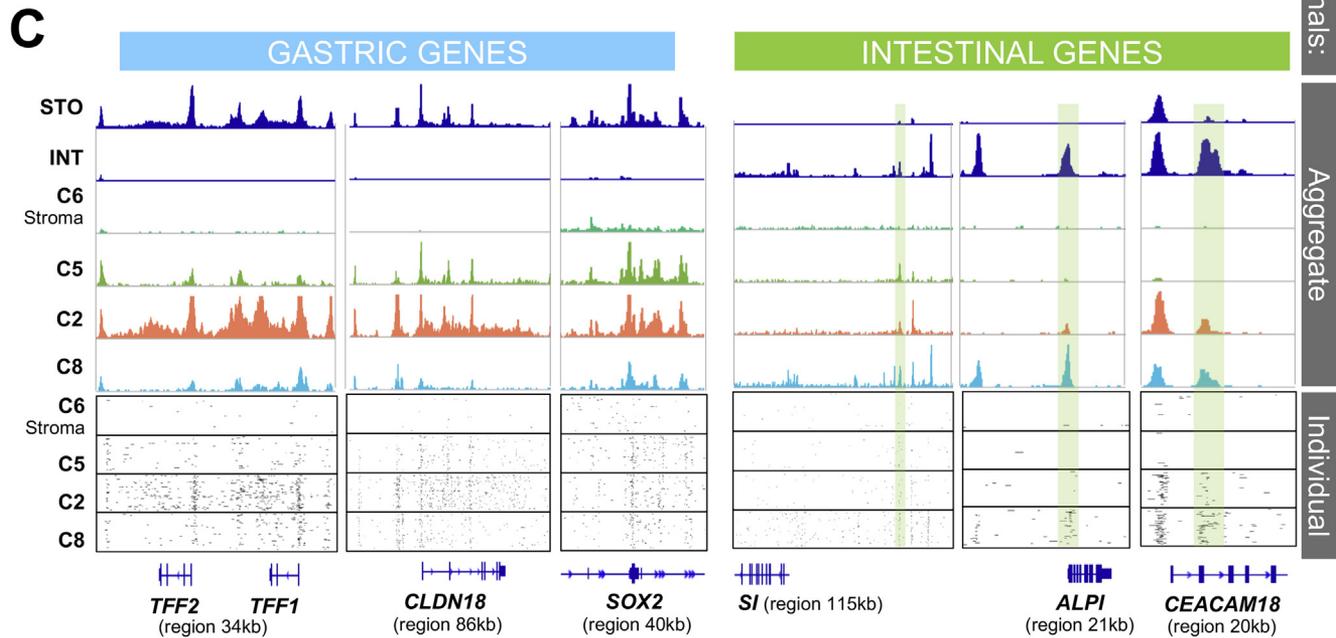
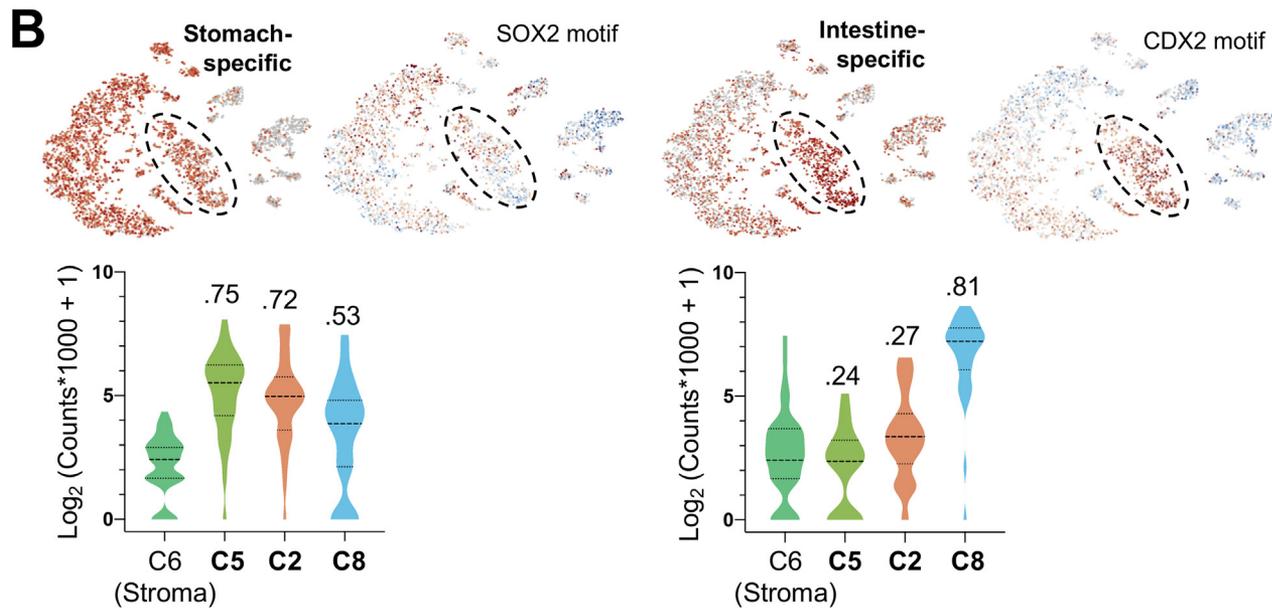
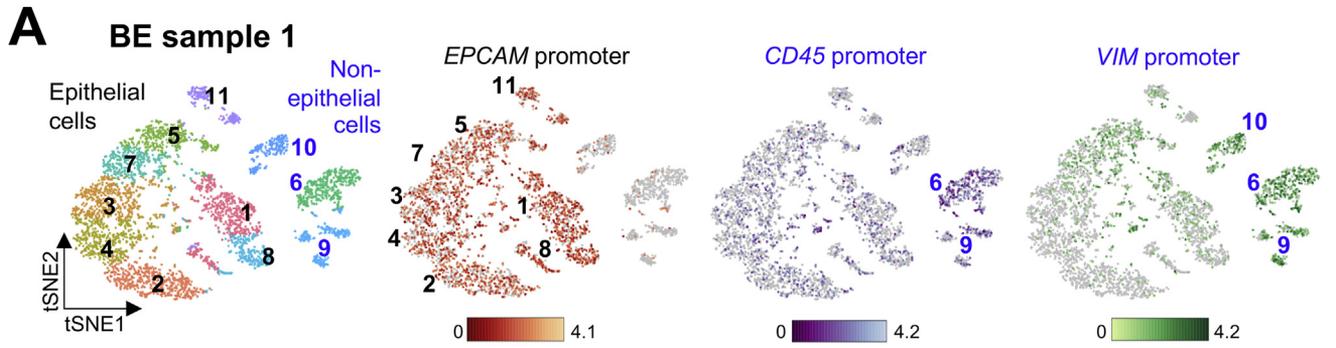
Stretch²¹ or super enhancers²² carry high levels of active histones and control lineage-defining genes; we identified H3K4me2⁺ stretch enhancers that are selective for normal gastric or esophageal mucosa and categorically absent in the other (Supplementary Table 2). All BE specimens carried H3K4me2 at stomach- but not at esophagus-restricted stretch enhancers (Figure 2A). Importantly, BE messenger RNA (mRNA) profiles^{14,34} showed transcripts located near stomach-specific sites and lacked those encoded near esophageal enhancers (Figure 2B). BE-enriched stretch

enhancers were generally marked in normal intestines (eg, *CDX1* locus [Supplementary Figure 2B]), whereas classic squamous cell loci lacked H3K4me2 (eg, *KRT5* and *TP63* [Figure 2C]), and the minimal signals at 924 esophageal stretch enhancers reinforce that trace H3K4me2 signals in Figure 1C are not consequential. Stomach enhancers were not marked in the intestine (eg, *TFF2*, *GATA4*, *CLDN18*, and *IGF2BP2* [Figure 2C]) and therefore do not represent byproducts of intestinal differentiation in BE. Rather, the data show both intestine and stomach enhancer fingerprints (eg, *TFF1/2/3* cluster [Supplementary Figure 2C]) in BE, with corresponding gene activity.

Mutational Profiles Implicate a Gastric Origin for Barrett's Esophagus

Intestinalization of reparative gastric mucosa offers one explanation for this hybrid chromatin state. Alternatively, BE could originate in squamous esophageal cells that erase their identity and activate ectopic gastric and intestinal enhancers. In this light, we note that somatic (mostly non-coding) mutations likely accumulate over decades before metaplastic conversion and propagate clonally in their original patterns. Indeed, mutations in 23 BE specimens (OCCAMS [Oesophageal Cancer Clinical and Molecular Stratification] study group³⁵) were distributed most similarly to those in EAC and gastric cancer, not those in esophageal squamous cell carcinoma (ESCC) (Figure 3A). Closed, enhancer-depleted chromatin is vulnerable to mutations, possibly owing to limited access by DNA repair enzymes,³⁶ and cells of cancer origin can be imputed from the regional frequency of somatic mutations in relation to genome-wide chromatin features²⁴ (Figure 3B). Because mutations in BE should similarly reflect the COO, we developed a random forest machine-learning algorithm to associate BE mutation frequencies with the epigenome features specific to each digestive epithelium (Figure 1). Using this algorithm, H3K4me2-based esophageal and gastric epigenomes best predicted the distributions of somatic mutations in ESCCs and gastric cancers, respectively^{37,38} (Supplementary Figure 3A).

Applying the random forest algorithm to whole-genome DNA sequence data from BE specimens,³⁵ the gastric H3K4me2 landscape predicted the mutational patterns unequivocally in 16 and strongly in another 3 cases of BE (Figure 3B). The mutation frequency in 4 samples was uninformatively low (<2.4 per Mb), and in no case did esophageal or intestinal enhancers predict the pattern better than gastric sites (Figure 3B). Removal of the stomach, but not the intestinal, epigenome significantly reduced the mutational variance explained by the model (Supplementary Figure 3B), thus confirming our conclusion. A representative region (Figure 3C) shows BE mutations concentrated in H3K4me2-depleted areas in normal gastric epithelium (correlation coefficient 0.78) and not in areas of unmarked esophageal chromatin. The consensus stomach epigenome also predicted mutation patterns in 17 of 23 EACs, with the BE epigenome predicting patterns best in the other 6 (Supplementary Figure 3C). BE precedes EAC,¹ and



these findings imply that EAC mutations mainly arise over many years in the gastric mucosa, with fewer additional mutations arising during the proportionally short interval between metaplasia and cancer.

Single-Cell Assay for Transposase-Accessible Chromatin Sequencing Shows Heterogeneous Cell and Chromatin States in Human Barrett's Esophagus

Gastric and intestinal enhancer coactivity in BE could reflect concurrent stomach- and intestine-dominant cell states in the tissue or a hybrid state within individual cells. To distinguish these possibilities, we used scATAC-seq to map accessible chromatin in endoscopic biopsy samples of human BE ($n = 2$), stomach ($n = 2$), and duodenum ($n = 1$) from 2 volunteers. All samples gave adequate cell numbers and sequencing depth; data from 19,327 informative cells were first processed from individual samples, and libraries were normalized for sequence depth, then merged to reduce dimensionality (Supplementary Table 3 and Supplementary Figure 3D). Then, t -distributed stochastic neighbor embedding (tSNE)³⁹ grouped stomach samples together, distinct from intestinal cells, whereas BE specimens formed unique clusters (Figure 4A and Supplementary Figure 3D). Open promoter chromatin identified the predominant *EPCAM*⁺ epithelial cell fraction (Figure 4A), distinct from contaminant *VIM*⁺ stromal cells or *CD45*⁺ leukocyte (clusters 15–17 and 21) (Supplementary Figure 3D). *HOXB* genes are reported to be uniquely active in BE³⁴ and aggregate (pseudo-bulk) analysis of scATAC-seq data showed BE-specific open chromatin at the *HOXB* and *HOXA* genes, showing wider *HOX* cluster dysregulation (Supplementary Figure 3E). Pseudo-bulk signals from epithelial cells in stomach samples coincided with tissue-specific enhancers identified by FiT-seq (eg, Supplementary Figure 4A); thus, both FiT-seq and scATAC accurately identify *cis*-regulatory activity.

Because sparse scATAC-seq signals limit robust interrogation of individual enhancers, we aggregated chromatin accessibility signals at 37 stomach- and 32 intestine-restricted enhancers derived from FiT-seq (Figure 1B) to generate reliable tissue-specific signatures (Supplementary

Methods and Supplementary Table 4). These enhancer-defined tissue states accurately identified graph-based stomach and intestinal epithelial cell clusters (Figure 4B). Open chromatin at intestinal enhancers dominated sample BE2, which nevertheless clustered separately from the intestine, likely owing to accessibility of stomach enhancers. In contrast, BE1 partitioned into a small tSNE cluster dominated by intestinal enhancers and a larger cluster with a predominant gastric signature (dashed ovals, Figure 4B). Thus, independent BE specimens with comparable, abundant goblet cell differentiation and accessible intestinal enhancers showed concomitantly open chromatin at many gastric enhancers (Figure 4B).

We used panels of classical gastric and intestinal transcripts (Figure 4C) to derive a second index of tissue-specific *cis*-regulatory activity: chromatin access at *cis*-elements that overlap with the respective gene bodies. This tissue signature confirmed concomitant access at gastric and intestinal loci, with BE1 again showing greater stomach character (Figure 4C). Despite intestinal dominance in BE2, we readily detected stomach-restricted *cis*-element access. Of note, although *KRT5* is reported to mark cells that initiate BE,¹⁶ scATAC showed closed chromatin at the *KRT5* promoter, compared, for example, to BE-specific *KRT7*^{16,40} (Supplementary Figure 4B).

Enhancer Coaccessibility in Individual Cells

The foregoing analyses show simultaneously open chromatin at gastric and intestinal *cis*-elements at the resolution of tSNE clusters. The relative dominance of gastric or intestinal features in BE1 subclusters (Figure 4B, dashed ovals) allowed us to ask whether stomach and intestinal elements are coaccessible in discrete cells. To this end, we clustered cells from BE1 separately and used *EPCAM* promoter signals to distinguish the majority epithelial fraction from nonepithelial *CD45*⁺ or *VIM*⁺ cells (Figure 5A). Open chromatin at intestinal enhancers dominated in clusters 1 and 8, where accessible sites were globally enriched for the *CDX2* consensus motif, whereas stomach enhancers enriched for the *SOX2* motif were distributed widely (Figure 5B). Cells in cluster 2 showed concomitant intestinal

Figure 5. Heterogeneous gastric and intestinal *cis*-element repertoires in BE. (A) Independent clustering and tSNE projection of cells from BE case 1, with ATAC signals at the *EPCAM* promoter distinguishing epithelial (black numbers) from nonepithelial (blue numbers, open chromatin at the *VIM* promoter) populations. Clusters 6 and 9 likely represent blood-derived cells (open *CD45* promoter chromatin); *CD45*-negative cluster 10 likely represents mesenchymal cells. (B) Graph-based (tSNE) clustering of single BE1 cells, showing focal increase of intestinal *cis*-element signal ($n = 37$) in specific clusters, with abundant stomach-specific enhancer signals ($n = 36$) throughout the epithelial component. Clusters enriched for intestinal *cis*-elements were also enriched for intestinal (*CDX2*) and depleted of stomach (*SOX2*) TF sequence motifs. Cells within clusters 5, 2, and 8, for example, show different degrees of stomach- and intestine-restricted enhancer coaccessibility. Average signals on each enhancer are plotted for all cells in the indicated clusters and quantified in violin plots. Cluster 2 shows a stomach-dominant pattern, whereas cluster 8 shows substantial intestinal enhancer activity alongside stomach enhancers. The D -statistics (Kolmogorov-Smirnov test) noted above each violin estimate the similarity between cell populations, using stromal cell cluster 6 as the reference. (C) Chromatin accessibility at classic gastric and intestinal gene loci in cells from BE1 epithelial clusters 5, 2, and 8, showing different degrees of stomach (STO) and intestinal (INT) differentiation; cluster 6 (nonepithelial stromal cells) serves as a control. Aggregate (pseudo-bulk) signals from stomach and intestine samples and BE1 cell clusters are displayed at the top. Cluster 8 shows notable *cis*-element coaccessibility near intestinal and gastric genes, whereas cluster 2 has few open intestinal sites (green shaded boxes) and prominent signals at gastric loci. Below, chromatin accessibility at these loci is displayed for 100 random single cells from each cluster. C, cluster.

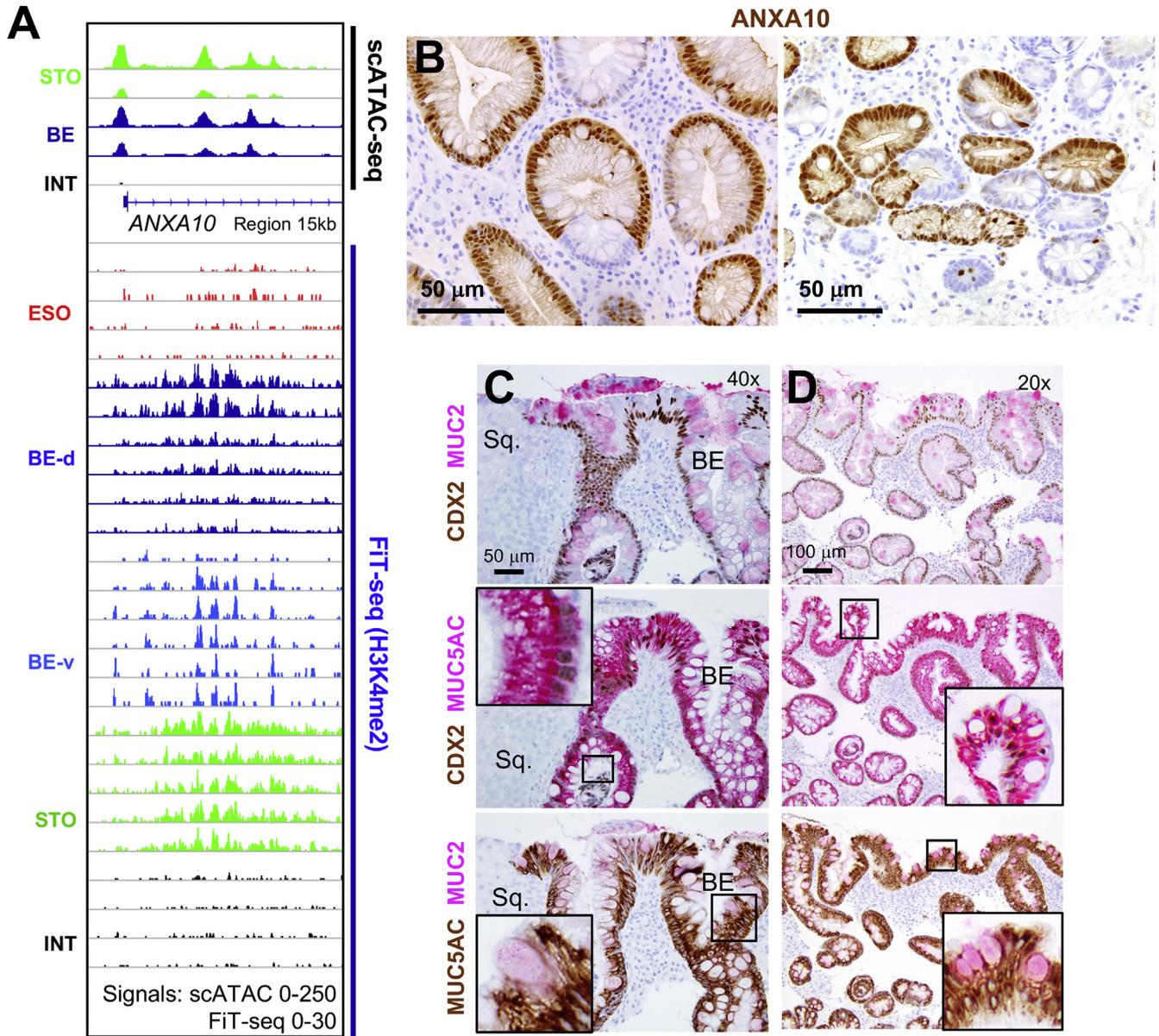


Figure 6. Novel BE marker ANXA10 and coexpression of gastric and intestinal features. (A) scATAC-seq identified open chromatin at *ANXA10*, a stomach-restricted locus, in BE. *Cis*-element activity is confirmed by H3K4me2 FiT-seq on 11 independent BE samples. (B) Representative ANXA10 immunostaining of 5 independent BE samples shows regions of prominent expression varying within and between metaplastic glands. (C, D) Two-color immunohistochemistry for CDX2 and stomach- (MUC5AC) or intestine- (MUC2) specific genes in sequential tissue sections from 5 different representative BE samples. (C and D represent specimens from different patients.) MUC5AC is coexpressed with MUC2 and CDX2 in hundreds, but not all, single cells. Adjoining squamous (Sq.) mucosa lacks all 3 markers. Insets show enlarged views of the boxed regions, highlighting marker coexpression.

and gastric *cis*-element accessibility, and gastric features were evident even in the most intestinalized group, cluster 8. For quantitative comparison and in the following examples, stromal cell cluster 6 provides an internal specificity control. At canonical tissue-restricted loci, ATAC-seq plots in Figure 5C show aggregate signals in selected cell clusters, and the boxes below display signals from 100 discrete cells. Clusters with low (no. 5), intermediate (no. 2), or high (no. 8) intestinal character show unambiguous coaccessibility of gastric sites. Thus, BE harbors a hybrid stomach-intestinal

chromatin state, with variable dominance of one state over the other among cells from the same biopsy specimens.

Pairwise comparisons identified 294 genes near chromatin that are accessible in BE and stomach, but not intestinal, specimens (Supplementary Table 5). These genes are largely expressed in stomach and BE, but not in intestines (Supplementary Figure 4C, public mRNA data^{14,41,42}) and include classical gastric loci at which FiT-seq showed H3K4me2 marking in BE: *CLDN18* (Figure 2C and Supplementary Figure 4D), *TFF1* (Supplementary

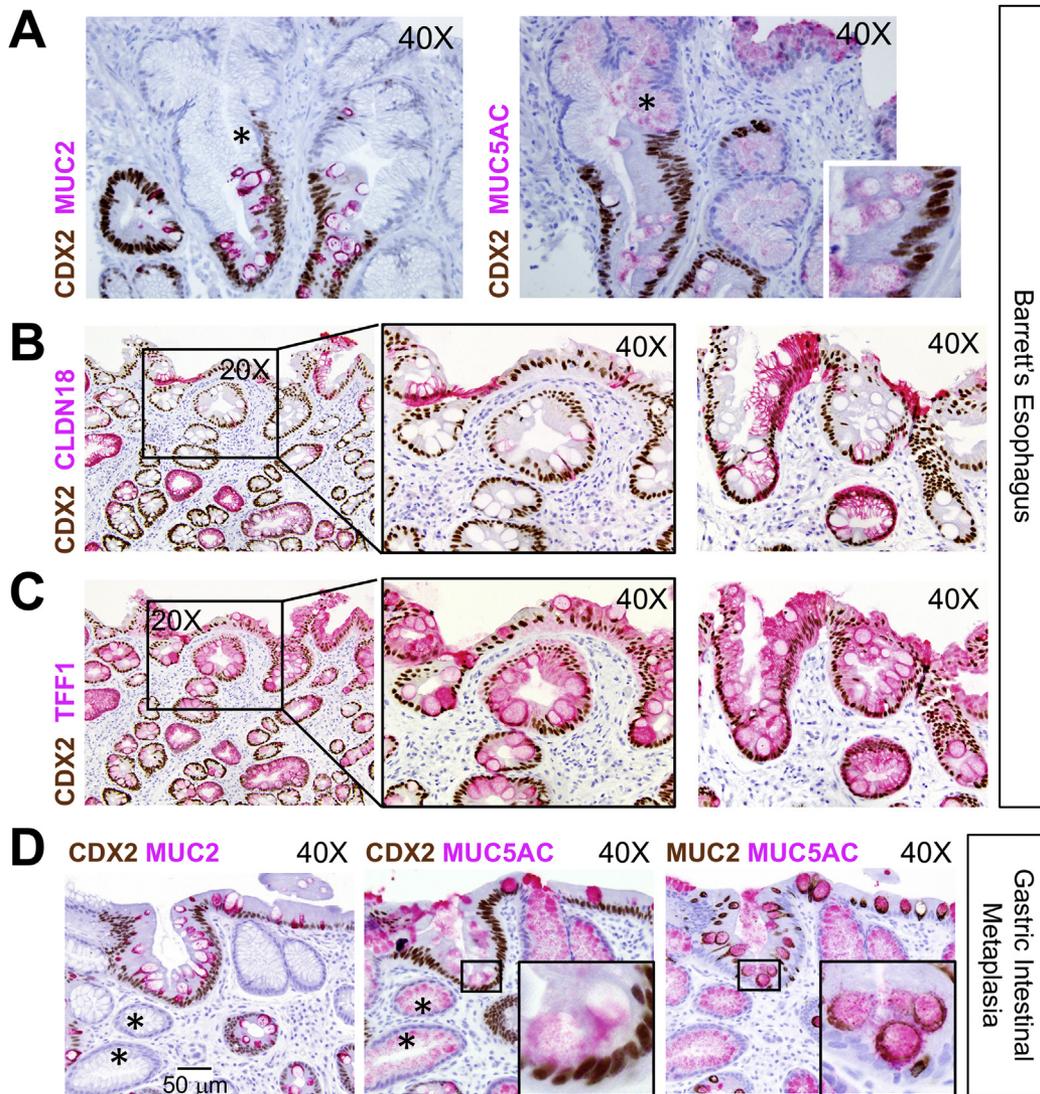


Figure 7. Heterogeneous cell identities in BE and gastric intestinal metaplasia. (A–C) Intrasample heterogeneity of coexisting stomach and intestinal states in BE. (A) Representative areas of CDX2⁻ (asterisks) and CDX2⁺ areas within presumptive clonal BE glands. CDX2 coexpression with additional stomach markers, CLDN18 (B) and TFF1 (C), is evident in specimens (n = 5) also used for ANXA10 immunostaining (Figure 6B). Two representative specimens are shown; the middle panels in B and C show boxed areas in the left panels at higher magnification. Both markers show extensive intra- and intergland heterogeneity. Images were captured through 20× or 40× objectives. Scale bar is shown in D, left panel. (D) Gastric intestinal metaplasia also harbors dual and heterogeneous gastric and intestinal cell states. In these representative images (n = 7; additional examples in Supplementary Figure 6C and D), the superficial epithelium expresses CDX2, but many deep glands do not (asterisks). MUC5AC is present broadly, but only CDX2⁺ cells express MUC2. Insets show boxed areas at higher magnification.

Figure 2C), MUC5AC, and MUC6 (Supplementary Figure 4E). Additional genes such as ANXA10 (Figure 6A), SLC45A3, and others (Supplementary Figure 5A) gave strong scATAC signals in stomach and BE, but weak to no signals in intestinal cells, and FiT-seq showed corresponding H3K4me2 marks. ANXA10 transcripts are exquisitely stomach restricted (Supplementary Figure 5B). Immunostaining confirmed ANXA10 expression in stomach, but not colonic, epithelium (Supplementary Figure 5C) and showed abundant expression in 5 of 5 BE specimens, with striking inter- and intragland heterogeneity (Figure 6B). Thus, scATAC accurately captured active cis-elements, showed extensive

stomach-specific gene activity in BE, and identified ANXA10 as a novel marker of mosaic gastric identity.

Elaboration of Dual Stomach-Intestinal Identity in Discrete Barrett's Esophagus Cells

To ask whether concurrent chromatin states yield truly hybrid cells, we immunostained stomach- (MUC5AC) and intestine- (MUC2 and CDX2) restricted proteins in 5 BE biopsy samples other than those used for FiT-seq, scATAC, or ANXA10 immunostaining. As expected, CDX2 was widespread and coexpressed with

MUC2 in goblet-like cells; MUC2 was expressed only in CDX2⁺, but CDX2 was present in many MUC2⁻ cells. All BE samples carried large numbers of cells that coexpressed MUC5AC and intestine-specific markers. (Three samples are shown in Figures 6C and D and 7A; another is shown in Supplementary Figure 6A.) Regions within every sample showed abundant MUC5AC⁺ foveolar cells, affirming their gastric character, whereas innumerable goblet-like cells coexpressed MUC2 and CDX2 along with MUC5AC (Figure 6C and D and Supplementary Figure 6A). Mixed glands were sometimes abundant and sometimes scattered, with field-to-field variation. Figure 6 shows representative coexpression and because other areas showed less or greater overlap, focal quantitation would be misleading.

Intragland heterogeneity was extensive: cells in some areas expressed all 3 markers, whereas others in the same glands lacked CDX2 and expressed only MUC5AC in cells with foveolar morphology. (An example is shown in Figure 7A.) This diversity, also seen with ANXA10 (Figure 6B), was striking and surprising because each BE gland,^{6,43} as in normal stomach,^{44,45} derives clonally from resident stem cells.^{46,47} Moreover, although MUC2⁺ goblet cells were present in all specimens, MUC5AC coexpression varied among glands and cases, with some areas showing both markers in every cell (Supplementary Figure 6A) or only in a minority. This heterogeneity matches our scATAC-seq findings, where cell subpopulations from the same BE biopsy sample showed limited, moderate, or extensive gastric character on an intestinal background.

Breadth of Dual Stomach-Intestinal Identity in Human Barrett's Esophagus and Gastric Intestinal Metaplasia

To assess the breadth of gastric properties, we examined additional stomach-specific genes identified by scATAC. In all 5 BE specimens that we first used for ANXA10 immunostaining, sizable patches of CDX2⁺ cells coexpressed CLDN18, and goblet-like cells coexpressed TFF1 (Figure 7B and C), hence verifying hybrid cell states. We again observed the heterogeneous expression of stomach markers within and across glands; moreover, sequential sections from the same specimen showed incomplete concordance among different gastric markers (Figure 7B and C, left panels), even in BE cells with gastric morphology. These findings reflect the substantial heterogeneity in mixed chromatin states evident in scATAC analysis (Figure 5).

The overt resemblance between gastric IM (GIM) and BE⁴⁸ prompted us to examine cellular and intragland features in GIM. All specimens (n = 7) showed coexpression of stomach- and intestine-specific products (Figure 7D and Supplementary Figure 6B–D), with substantial intragland and intersample variability and close proximity of extensively intestinalized glands with others that appeared wholly gastric. Regional CDX2 expression was associated in some areas with high MUC2 and exclusion of MUC5AC and in adjoining areas with high MUC5AC and absence of MUC2

(Supplementary Figure 6C and D, arrows). Again, MUC2 appeared only in CDX2⁺ cells, but the converse was not always true; for example, Supplementary Figure 6C shows superficial foveolar epithelium with low CDX2 and no MUC2 (magnified in the inset). Deeper glands in the same sample showed higher CDX2 levels and concomitant MUC2. Thus, both BE and GIM harbor a spectrum of discrete cell states, ranging from predominantly gastric to largely intestinal.

Discussion

EAC incidence is rising faster in the West than other cancers.² BE, which develops in response to esophageal injury from chronic gastric and biliary reflux, is the principal risk factor.^{1,3} A traditional view is that BE represents conversion of squamous to intestinal epithelium,^{5–7,49} but recent opinion has shifted to suggest that embryonic remnants¹⁵ or specialized transitional cells¹⁶ from the gastroesophageal junction, Lgr5⁺ stem cells from the proximal stomach,¹⁷ or ESMG cells are the source.^{14,50} These candidate sources can account for columnar BE histomorphology with gastric features and were nominated largely on the basis of animal experiments, in some cases with strong human correlates. Our examination of histone marks and open chromatin in primary human BE specimens shows, at both the population and sc scales, a hybrid stomach-intestinal chromatin state associated with extensive coexpression of gastric and intestinal marker genes.

Somatic mutations likely acquired and clonally propagated before IM also correlated better with the epigenome signature of gastric columnar than of esophageal squamous mucosa. The phenotype and long-term stability of BE implicates an IM source with stem cell activity.^{8–10} Our work does not implicate a specific COO, nor is it currently possible to isolate for epigenome analysis any of the rare candidate sources, such as embryonic rest,¹⁵ transitional junction,¹⁶ or ESMG^{13,14} cells. It is unknown if the gastric epithelial epigenome sufficiently resembles these rare cells to explain the correlation of BE mutational signature with the stomach epigenome. Our findings do, however, argue against stratified esophageal mucosa as a BE source.

We suggest that esophageal surface injury from gastric reflux is usually repaired by the native stratified epithelium but that extensive or certain forms of injury elicit repair from a nearby columnar source. Stem cells in the ectopic reparative epithelium preserve substantial elements of their original identity—expressing abundant MUC5AC, TFF1, CLDN18, ANXA10, and other stomach-specific genes—and, in addition, manifest intestinal features; to our knowledge, ANXA10 is a novel marker of the gastric phenotype. This model agrees with all known features of BE, and future preventive strategies might aspire to promote esophageal repair by native squamous cells at the expense of salvage by heterologous cells.

The active histone mark detected at stomach enhancers in our bulk FiT-seq data represents the cellular average. One striking observation from scATAC-seq, corroborated by immunostaining of multiple independent cases, is that the degree of stomach identity varies widely: some cells appear

largely intestinal, with modest gastric signals, whereas others carry a predominant gastric *cis*-regulatory signature. Notably, gastric marker expression is heterogeneous even in BE cells with clear dual identity. This heterogeneity appears not to be clonally determined because areas within the same gland show extremes of gastric or intestinal character. One untested possibility is that local factors such as pH, inflammation, or stromal signals determine cellular identities on the background of a plastic—and likely dynamic—chromatin state. Further epigenome characterization at sc resolution might clarify the degree of underlying chromatin plasticity.

Gastric cancers of the intestinal type and EAC share genomic and biological properties.¹² It is therefore worth considering the morphologic and molecular features common to their respective precursor lesions, GIM and BE,⁴⁸ which may reflect similarly partial intestinalization. We report that GIM also harbors gland heterogeneity—ranging from largely foveolar to mostly goblet-like cells—with innumerable cells coexpressing stomach and intestinal markers. Our study was limited to nondysplastic BE and GIM; it will be interesting in the future to determine whether the fractional representation of gastric and intestinal fates is a driver or biomarker for progression to dysplasia and invasive cancer. In light of the substantial clonal and subclonal heterogeneity, studies that address this question will need large cohorts, sc resolution, and cautious interpretation of data.

Hybrid and heterogeneous chromatin states provide a compelling explanation for dual, nonuniform, and variable presence of gastric and intestinal features. Many metaplastic glands lacked uniform CDX2 expression, and its presence in BE or GIM was not always associated with other intestinal products in the same cell. These findings imply that even intestinal TF expression is not hardwired in clonal metaplastic cells or guaranteed to activate intestinal genes. Rather, they suggest significant cell plasticity and localized stromal or inflammatory influence over TF expression and activities.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at <https://doi.org/10.1053/j.gastro.2021.05.057>.

References

1. Spechler SJ, Souza RF. Barrett's esophagus. *N Engl J Med* 2014;371:836–845.
2. Pohl H, Welch HG. The role of overdiagnosis and reclassification in the marked increase of esophageal adenocarcinoma incidence. *J Natl Cancer Inst* 2005;97:142–146.
3. Rustgi AK, El-Serag HB. Esophageal carcinoma. *N Engl J Med* 2014;371:2499–2509.
4. Giroux V, Rustgi AK. Metaplasia: tissue injury adaptation and a precursor to the dysplasia-cancer sequence. *Nat Rev Cancer* 2017;17:594–604.
5. Stairs DB, Nakagawa H, Klein-Szanto A, et al. Cdx1 and c-Myc foster the initiation of transdifferentiation of the normal esophageal squamous epithelium toward Barrett's esophagus. *PLoS One* 2008;3(10):e3534.
6. Nicholson AM, Graham TA, Simpson A, et al. Barrett's metaplasia glands are clonal, contain multiple stem cells and share a common squamous progenitor. *Gut* 2012;61:1380–1389.
7. Vercauteren Drubbel A, Pirard S, Kin S, et al. Reactivation of the Hedgehog pathway in esophageal progenitors turns on an embryonic-like program to initiate columnar metaplasia. *Cell Stem Cell*. Published online April 20, 2021. <https://doi.org/10.1016/j.stem.2021.03.019>
8. Arul GS, Moorghen M, Myerscough N, et al. Mucin gene expression in Barrett's oesophagus: an in situ hybridisation and immunohistochemical study. *Gut* 2000;47:753–761.
9. Lavery DL, Nicholson AM, Poulosom R, et al. The stem cell organisation, and the proliferative and gene expression profile of Barrett's epithelium, replicates pyloric-type gastric glands. *Gut* 2014;63:1854–1863.
10. McDonald SA, Graham TA, Lavery DL, et al. The Barrett's gland in phenotype space. *Cell Mol Gastroenterol Hepatol* 2015;1:41–54.
11. Van De Bovenkamp JH, Korteland-Van Male AM, Warson C, et al. Gastric-type mucin and TFF-peptide expression in Barrett's oesophagus is disturbed during increased expression of MUC2. *Histopathology* 2003;42:555–565.
12. The Cancer Genome Atlas Research Network, Analysis Working Group; Asan University, BC Cancer Agency, et al. Integrated genomic characterization of oesophageal carcinoma. *Nature* 2017;541(7636):169–175.
13. Kruger L, Gonzalez LM, Pridgen TA, et al. Ductular and proliferative response of esophageal submucosal glands in a porcine model of esophageal injury and repair. *Am J Physiol Gastrointest Liver Physiol* 2017;313(3):G180–G191.
14. Owen RP, White MJ, Severson DT, et al. Single cell RNA-seq reveals profound transcriptional similarity between Barrett's oesophagus and oesophageal submucosal glands. *Nat Commun* 2018;9(1):4261.
15. Wang X, Ouyang H, Yamamoto Y, et al. Residual embryonic cells as precursors of a Barrett's-like metaplasia. *Cell* 2011;145:1023–1035.
16. Jiang M, Li H, Zhang Y, et al. Transitional basal cells at the squamous-columnar junction generate Barrett's oesophagus. *Nature* 2017;550(7677):529–533.
17. Quante M, Bhagat G, Abrams JA, et al. Bile acid and inflammation activate gastric cardia stem cells in a mouse model of Barrett-like metaplasia. *Cancer Cell* 2012;21:36–51.
18. Barski A, Cuddapah S, Cui K, et al. High-resolution profiling of histone methylations in the human genome. *Cell* 2007;129:823–837.
19. Heintzman ND, Hon GC, Hawkins RD, et al. Histone modifications at human enhancers reflect global

- cell-type-specific gene expression. *Nature* 2009; 459(7243):108–112.
20. Roadmap Epigenomics Consortium, **Kundaje A, Meuleman W**, et al. Integrative analysis of 111 reference human epigenomes. *Nature* 2015;518(7539):317–330.
 21. **Parker SC, Stitzel ML**, Taylor DL, et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc Natl Acad Sci U S A* 2013;110:17921–17926.
 22. **Whyte WA, Orlando DA, Hnisz D**, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 2013; 153:307–319.
 23. Blokzijl F, de Ligt J, Jager M, et al. Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* 2016;538(7624):260–264.
 24. **Polak P, Karlic R**, Koren A, et al. Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature* 2015;518(7539):360–364.
 25. Shaheen NJ, Falk GW, Iyer PG, et al. ACG clinical guideline: diagnosis and management of Barrett's esophagus. *Am J Gastroenterol* 2016;111:30–50.
 26. Sharma P, Bansal A, Mathur S, et al. The utility of a novel narrow band imaging endoscopy system in patients with Barrett's esophagus. *Gastrointest Endosc* 2006;64:167–175.
 27. **Cejas P, Li L, O'Neill NK**, et al. Chromatin immunoprecipitation from fixed clinical tissues reveals tumor-specific enhancer profiles. *Nat Med* 2016;22:685–691.
 28. Breiman L. Random forests. *Machine Learning* 2001; 45:5–32.
 29. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* 2012;9:215–216.
 30. **Cejas P, Drier Y, Dreijerink KMA**, et al. Enhancer signatures stratify and predict outcomes of non-functional pancreatic neuroendocrine tumors. *Nat Med* 2019; 25:1260–1265.
 31. Arnold K, Sarkar A, Yram MA, et al. Sox2⁺ adult stem and progenitor cells are important for tissue regeneration and survival of mice. *Cell Stem Cell* 2011;9:317–329.
 32. Jacobsen CM, Narita N, Bielinska M, et al. Genetic mosaic analysis reveals that GATA-4 is required for proper differentiation of mouse gastric epithelium. *Dev Biol* 2002;241:34–46.
 33. Hyun K, Jeon J, Park K, Kim J. Writing, erasing and reading histone lysine methylations. *Exp Mol Med* 2017; 49(4):e324.
 34. di Pietro M, Lao-Sirieix P, Boyle S, et al. Evidence for a functional role of epigenetically regulated midcluster *HOXB* genes in the development of Barrett esophagus. *Proc Natl Acad Sci U S A* 2012;109(23):9077–9082.
 35. **Ross-Innes CS, Becq J**, Warren A, et al. Whole-genome sequencing provides new insights into the clonal architecture of Barrett's esophagus and esophageal adenocarcinoma. *Nat Genet* 2015;47:1038–1046.
 36. **Polak P, Lawrence MS**, Haugen E, et al. Reduced local mutation density in regulatory DNA of cancer genomes is linked to DNA repair. *Nat Biotechnol* 2014; 32:71–75.
 37. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 2014;513(7517):202–209.
 38. **Zhang L, Zhou Y, Cheng C**, et al. Genomic analyses reveal mutational signatures and frequently altered genes in esophageal squamous cell carcinoma. *Am J Hum Genet* 2015;96:597–611.
 39. **Satpathy AT, Granja JM**, Yost KE, et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat Biotechnol* 2019;37:925–936.
 40. Cabibi D, Fiorentino E, Pantuso G, et al. Keratin 7 expression as an early marker of reflux-related columnar mucosa without intestinal metaplasia in the esophagus. *Med Sci Monit* 2009;15(5):CR203–CR210.
 41. **Maag JLV, Fisher OM**, Levert-Mignon A, et al. Novel aberrations uncovered in Barrett's esophagus and esophageal adenocarcinoma using whole transcriptome sequencing. *Mol Cancer Res* 2017;15:1558–1569.
 42. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013;45:580–585.
 43. Raskind WH, Norwood T, Levine DS, et al. Persistent clonal areas and clonal expansion in Barrett's esophagus. *Cancer Res* 1992;52:2946–2950.
 44. Bjerknes M, Cheng H. Multipotential stem cells in adult mouse gastric epithelium. *Am J Physiol Gastrointest Liver Physiol* 2002;283(3):G767–G777.
 45. Nomura S, Esumi H, Job C, et al. Lineage and clonal development of gastric glands. *Dev Biol* 1998;204:124–135.
 46. **Barker N, Huch M**, Kujala P, et al. Lgr5⁺ stem cells drive self-renewal in the stomach and build long-lived gastric units in vitro. *Cell Stem Cell* 2010;6:25–36.
 47. Leushacke M, Tan SH, Wong A, et al. Lgr5-expressing chief cells drive epithelial regeneration and cancer in the oxyntic stomach. *Nat Cell Biol* 2017;19:774–786.
 48. Piazzuelo MB, Haque S, Delgado A, et al. Phenotypic differences between esophageal and gastric intestinal metaplasia. *Mod Pathol* 2004;17:62–74.
 49. **Que J, Garman KS, Souza RF, Spechler SJ**. Pathogenesis and cells of origin of Barrett's esophagus. *Gastroenterology* 2019;157:349–364.
 50. Leedham SJ, Preston SL, McDonald SA, et al. Individual crypt genetic heterogeneity and the origin of metaplastic glandular epithelium in human Barrett's oesophagus. *Gut* 2008;57:1041–1048.

Author names in bold designate shared co-first authorship.

Received November 5, 2020. Accepted May 24, 2021.

Correspondence

Address correspondence to: Ramesh A. Shivdasani, MD, PhD, Dana-Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02215. e-mail: ramesh_shivdasani@dfci.harvard.edu; or Hwajin Lee, PhD, Biomedical Knowledge Engineering Laboratory, Seoul National University, Seoul 08826, South Korea. e-mail: hwajin2k@gmail.com.

Acknowledgments

The authors thank M.R. Corces for helpful discussions.

CRedit Authorship Contributions

Harshbad Singh, MBBS (Conceptualization: Equal; Data curation: Lead; Formal analysis: Lead; Investigation: Lead; Methodology: Lead; Writing – original draft: Equal; Writing – review & editing: Equal); Kyungsik Ha, PhD

(Formal analysis: Equal; Investigation: Equal); Jason L. Hornick, MD, PhD (Formal analysis: Supporting; Investigation: Supporting); Shariq Madha, BS (Formal analysis: Supporting); Paloma Cejas, PhD (Methodology: Supporting); Kunal Jajoo, MD (Methodology: Supporting; Resources: Supporting); Pratik Singh, PhD (Methodology: Supporting); Paz Polak, PhD (Formal analysis: Supporting); Hwajin Lee, PhD (Formal analysis: Equal; Writing – original draft: Supporting; Writing – review & editing: Supporting); Ramesh A. Shivdasani, MD, PhD (Conceptualization: Equal; Formal analysis: Equal; Funding acquisition: Lead; Investigation: Equal; Supervision: Lead; Writing – original draft: Equal; Writing – review & editing: Lead).

Conflicts of interest

These authors disclose the following: Kyungsik Ha is currently employed by Alopax-Algo Co, Ltd. Jason L. Hornick is a consultant to Epizyme, Aadi Biosciences, and TRACON Pharmaceuticals. Hwajin Lee is currently

employed by UPPTthera, Inc. Ramesh A. Shivdasani receives grant support from Novartis Oncology unrelated to this work. The remaining authors disclose no conflicts.

Funding

Supported by National Institutes of Health grants R01DK082889 (to Ramesh A. Shivdasani) and P50CA127003; the Dana-Farber/Novartis Drug Discovery Program (DDP17026); a generous gift from the Sarah Rhodes Fund for Cancer Research (to Ramesh A. Shivdasani); an endowment from the William Randolph Hearst Foundation (to Ramesh A. Shivdasani); and grants to Hwajin Lee from the Korean National Research Foundation, South Korea (NRF-2018M3A9H) and the Korean Institute for Information and Communications Technology (2017-0-00398). Harshabad Singh is a William Raveis Charitable Fund Physician-Scientist of the Damon Runyon Cancer Research Foundation (PST-15-18).

Supplementary Methods

Computational Analysis of Chromatin Immunoprecipitation Sequencing (Fixed-Tissue Chromatin Immunoprecipitation Sequencing) Data

DNA sequence tags were aligned against reference genome Hg19 using Bowtie2, version 2.3.4.3,¹ and H3K4me2-marked regions were identified using MACS, version 2.1.1 (<https://github.com/taoliu/MACS>) with the default cutoff of $q < 0.01$.² H3K4me2-marked regions located >2 kb from transcription start sites were designated as enhancers, and bigwig files were generated by align2rawsignal (<https://github.com/akundaje/align2rawsignal>). Data were visualized using the Integrative Genomics Viewer³; heatmaps were prepared using Deeptools,⁴ version 2.5.0; and correlation matrices were plotted using complexHeatmap and corrPlot⁵ in the R package (<https://CRAN.R-project.org/package=corrplot>). For principal component analysis, enhancer signals (normalized reads per kb per million sequence tags, [RPKM]) were quantified using Bamliquidator, version 1.3.8 (<https://github.com/BradnerLab/pipeline/wiki/bamliquidator>), analyzed using base R, and plotted using package rgl (<https://r-forge-project.org/projects/rgl/>). Regions were filtered for those that gave an RPKM of >30 in 3 samples from the full set, and the top quintile of most variable regions was identified using the coefficient of variation. Gene Set Enrichment Analysis (GSEA)⁶ was conducted using GSEA algorithms (<http://software.broadinstitute.org/gsea/index.jsp>). H3K4me2-marked regions were annotated using GREAT⁷ to identify the nearest gene <100 kb from the enhancer center. Gene sets generated in this manner were subsequently used in GSEA analysis (Supplementary Figure 1E); we used 500 random genes from any set containing >500 genes.

Delineation and Analysis of Tissue-Restricted Enhancers

We used k -means clustering to derive esophageal, gastric, and intestinal epithelium-specific enhancer sets in R. Enhancer signals were quantified and normalized using RPKM (reads per kb per million sequence tags). Enhancers with at least 4 RPKM in at least two samples were retained and the top 20% most variable enhancers were used for k -means clustering using coefficient of variation. The gap statistic to calculate cluster number was done in R using the base R command clusGap. The FirstSEmax and Tibs2001-SEmax methods identified 5 as an optimal k value, but the 3 esophageal enhancer subgroups were essentially similar and were merged into 1 cluster. H3K4me2 ChIP signals at tissue-specific enhancers were quantified in 560-bp regions flanking the center of each, using the compute matrix command in Deeptools⁴ on quantile-normalized bigwig files. The log-transformed average signal strength for each tissue was plotted as violin plots using ggplot2,⁸ version 3.1.0, in R (Figure 1C). A nonparametric Kolmogorov-Smirnov test was used to determine the D statistic and

plot cumulative frequency graphs (Figure 1C) using GraphPad Prism, version 8, for Mac (www.graphpad.com). Lower D statistics mean that samples are similar to each other. To identify enhancers marked differently in different epithelia, we used diffReps,⁹ version 1.55.6, applying the negative binomial method with a window size of 1 kb and a step size of 100 base pairs (bp). Because H3K4me2 peaks are broad, the “nsd” parameter in diffReps was set to 2. Stretch enhancers were defined as those >3 kb in length.¹⁰ To determine whether 11 samples adequately capture the BE enhancer landscape, we performed a saturation analysis (Supplementary Figure 1H), as described previously.¹¹ For each sample size (x -axis), random samples were picked from the cohort of 11 BM specimens, and the number of enhancer peaks in the union set was determined. This process was iterated 1000 times for each sample size number to generate means and standard errors for the number of enhancer peaks.

Gene Expression and Somatic Mutation Data

Data set GSE34619 (Figure 2B) was extracted from the Gene Expression Omnibus, and expression values were RMA normalized in R (<https://CRAN.R-project.org/>). Data on normal tissues (Supplementary Figure 1E) were obtained from the human protein atlas (<https://www.proteinatlas.org/>).¹² Bulk RNA-seq data (BE) (Figure 2B)¹³ were obtained from the European Genome-Phenome Archive (EGAS00001003144), and heatmaps (Figure 2B) were created using Morpheus (<https://software.broadinstitute.org/morpheus/>). To study somatic mutations, we used data from 23 pairs of BM and matched EAC samples¹⁴ (International Cancer Genome Consortium, <http://icgc.org>), 14 ESCC samples,¹⁵ and 9 samples each of diffuse- and intestinal-type gastric cancers (from BGI, <http://www.genomics.cn/>), or for samples with average mutation rates of >20 /Mb, from the International Cancer Genome Consortium data portal¹⁶). Data were analyzed using the GATK pipeline¹⁷ (<https://www.broadinstitute.org/gatk/>), and mutations were called using MuTect¹⁸ (<https://www.broadinstitute.org/cancer/cga/mutect>).

Feature Selection (Based on Random-Forest Machine Learning) and Cell-of-Origin Analysis

For feature selection and COO analysis, we applied a nonparametric machine-learning method (random forest) to predict the value of a continuous response variable by using ensemble decision trees.¹⁹ For additional input data, we calculated numbers of H3K4me2 FiT-seq reads per 1-Mb region (local H3K4me2 densities). To combine BE-discovery and BE-validation data from different FiT-seq experiments, we normalized all read counts to \log_{10} , generated a regression model between enhancers identified in each cohort, and excluded 10% of all regions (213 of 2128) with the largest differences in FiT-seq counts between the 2 cohorts. We then generated 1000 random forest trees to predict mutation density from H3K4me2 features in each of 4 tissue types and used greedy backward elimination to determine important rankings among H3K4me2 variables.

We thus calculated the variance-explained value for each variable in every step, then removed the variable ranked lowest in all steps. Random forest modeling was repeated 1000 times, both on individual samples and on groups of the same tissue type: BE, EAC, ESCC, and gastric cancer subtypes. To calculate tissue-specific mutation densities for the group analyses, we summed data on all samples from each cancer type or BE across each 1-Mb region.

To ascertain the statistical significance of the gastric epigenome contribution to BE mutational variance (Supplementary Figure 3B), we first calculated the variance in explained values using H3K4me2 chromatin features of each tissue and then recalculated the variance by removing H3K4me2 chromatin features from stomach or intestine. We repeated this cycle 1000 times and then assessed the statistical differences in the variance-explained value distributions for each group using 1-way analysis of variance followed by the Tukey honest significant difference post hoc test.

Single-Cell Assay for Transposase-Accessible Chromatin Sequencing

Libraries were sequenced to obtain adequate coverage, and data were processed using the Cell Ranger ATAC Pipeline, version 1.1.10.²⁰ Loupe Browser (10x Genomics) was used to visualize tSNE plots, distribution of cis-element accessibility, and enrichment of DNA sequence motifs. Signals at gastric and intestinal signature genes (Figure 4C) were assessed using the Signac function in Seurat (<https://github.com/timoast/signac>)²¹ by collating chromatin accessibility over gene promoters and bodies.

Signals at stomach and intestinal enhancers (Figure 4B) were extracted for peaks differentially marked with H3K4me2 in FiT-Seq data from normal human stomach and intestinal epithelia. Owing to the sparsity of scATAC, not all differential H3K4me2 regions are detected by scATAC. Therefore, the most differential H3K4me2 peaks detected by FiT-seq were intersected with scATAC peaks called by CellRanger. We intersected varying numbers of top stomach and intestine specific FiT-seq enhancers ($q < 0.01$; top 50, 100, 150, 200, 300, and 400 sites, ordered by decreasing \log_2 fold-difference) to identify the peaks represented in the scATAC data. To consider similar numbers of stomach- and intestine-specific enhancers for the analysis of scATAC-seq data, we found that among the top 150 stomach- and top 200 intestine-specific FiT-seq enhancers, 37 stomach- and 32 intestine-specific sites had good representation in the scATAC-seq data sets (Supplementary Table 4).

These signals were quantified by summing fragment counts from each enhancer over all cells in a sample (Figure 4B) or cluster (Figure 5B), followed by log-transformation and depiction of the distribution using violin plots (GraphPad, version 8). Cis-element signals were projected onto tSNE plots (Figures 4 and 5) using the Accessibility function in Loupe Browser (10x Genomics). ATAC-seq signals over specified regions in 100 random single cells from selected clusters (Figure 5C)

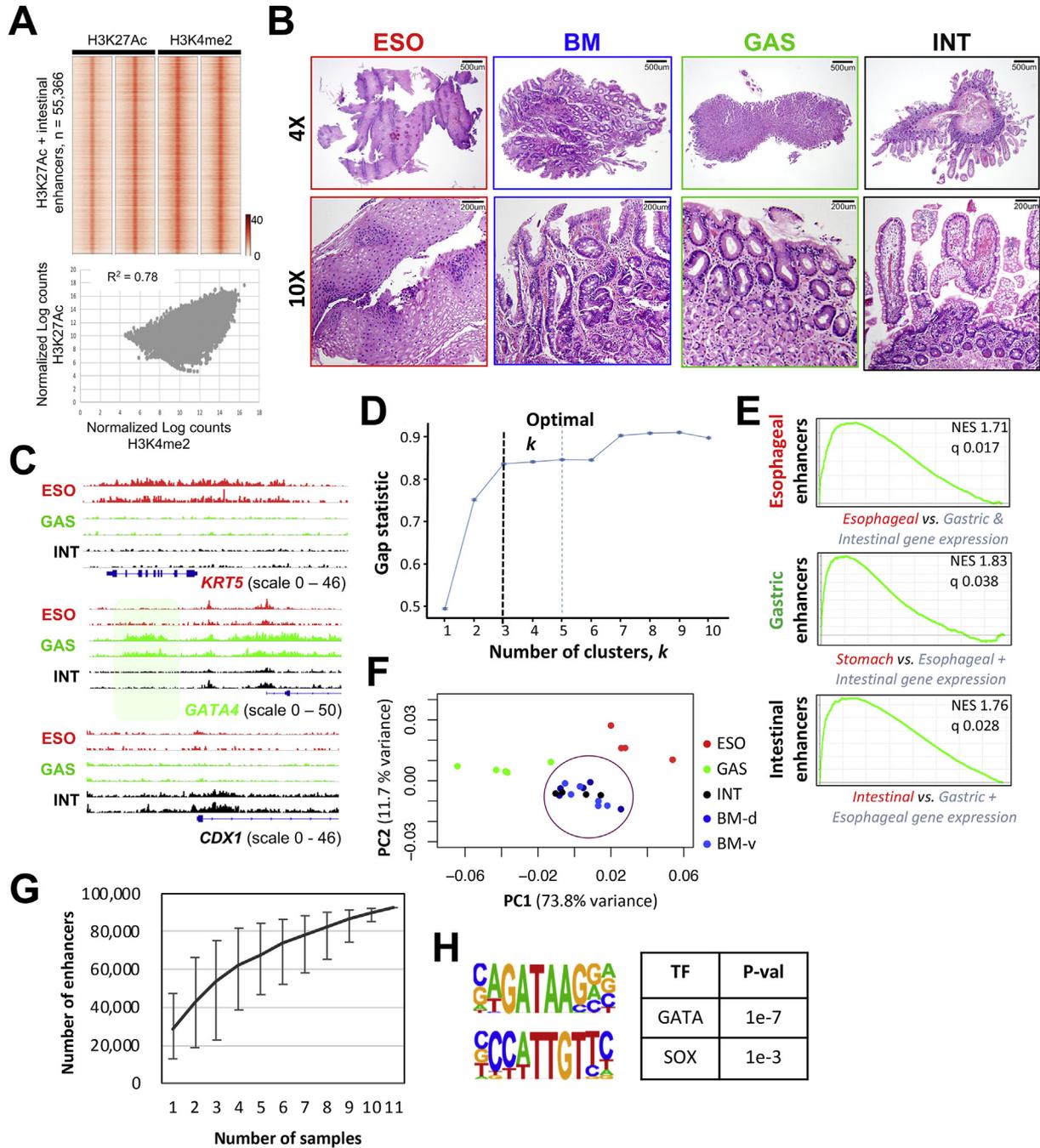
were calculated as a binarized matrix of cut counts per 100 bp, plotted using ggplot2,⁸ and depicted as described previously.²⁰ The same 100 cells are plotted for all sites.

To identify tissue markers shared between stomach and BE in scATAC-seq, we used Seurat (version 3.1.4)²¹ to obtain gene activity scores based on reads within the gene body or 2 kb upstream of the transcription start site. Data were normalized and clustered, and known markers were used to filter out mesenchymal cell clusters. We then used the FindMarkers function within Seurat to identify markers for each cell type, using default thresholds of 0.25 for log-fold change and $\geq 10\%$ of cells showing open chromatin. We compared intestine with either stomach or BE to identify genes showing significantly more open chromatin in the latter tissues (Supplementary Table 5).

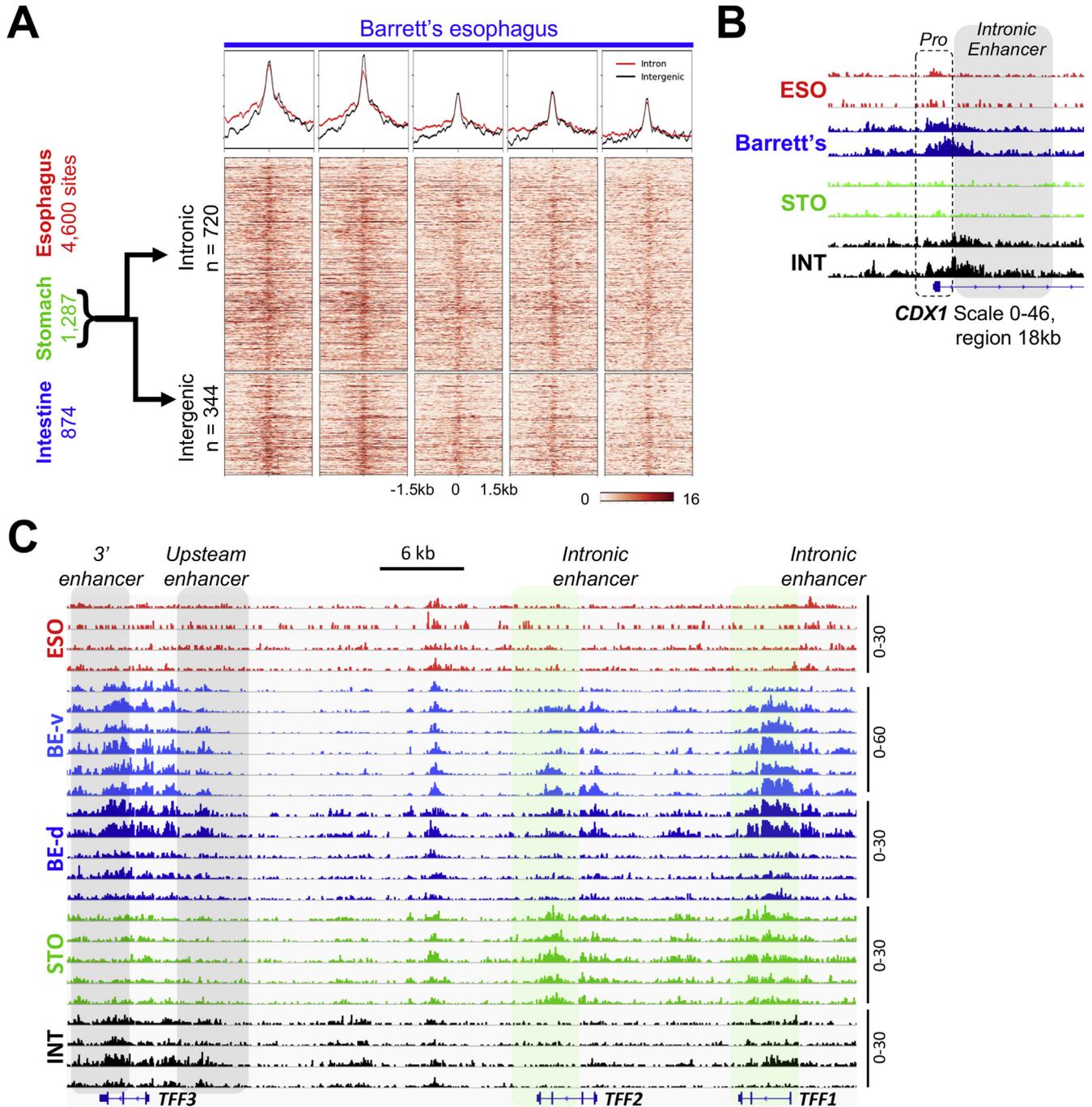
Supplementary References

- Langmead B, Trapnell C, Pop M, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009;10(3):R25.
- Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008; 9(9):R137.
- Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol* 2011; 29:24–26.
- Ramirez F, Ryan DP, Gruning B, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 2016;44(W1):W160–W165.
- Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 2016;32:2847–2849.
- Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545–15550.
- McLean CY, Bristor D, Hiller M, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 2010;28:495–501.
- Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag, 2016.
- Shen L, Shao NY, Liu X, et al. diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates. *PLoS One* 2013;8(6):e65598.
- Parker SC, Stitzel ML, Taylor DL, et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc Natl Acad Sci U S A* 2013;110:17921–17926.
- Pomerantz MM, Li F, Takeda DY, et al. The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nat Genet* 2015;47:1346–1351.
- Uhlén M, Fagerberg L, Hallström BM, et al. Tissue-based map of the human proteome. *Science* 2015; 347(6220):1260419.
- Owen RP, White MJ, Severson DT, et al. Single cell RNA-seq reveals profound transcriptional similarity

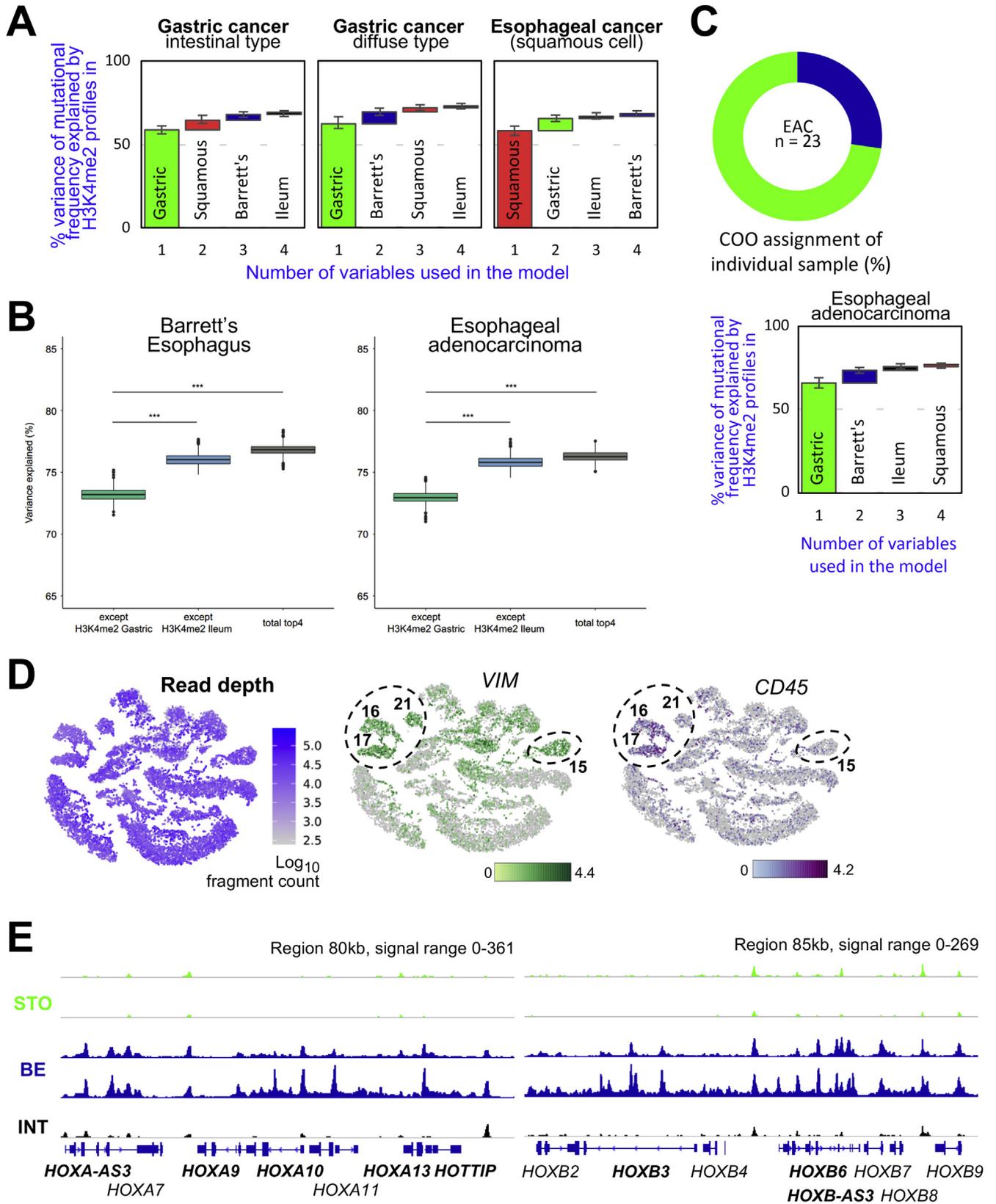
- between Barrett's oesophagus and oesophageal sub-mucosal glands. *Nat Commun* 2018;9(1):4261.
14. Ross-Innes CS, Becq J, Warren A, et al. Whole-genome sequencing provides new insights into the clonal architecture of Barrett's esophagus and esophageal adenocarcinoma. *Nat Genet* 2015;47:1038–1046.
 15. Zhang L, Zhou Y, Cheng C, et al. Genomic analyses reveal mutational signatures and frequently altered genes in esophageal squamous cell carcinoma. *Am J Hum Genet* 2015;96:597–611.
 16. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 2014;513(7517):202–209.
 17. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–1303.
 18. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013;31:213–219.
 19. Breiman L. Random forests. *Machine Learning* 2001;45:5–32.
 20. Satpathy AT, Granja JM, Yost KE, et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat Biotechnol* 2019;37:925–936.
 21. Stuart T, Butler A, Hoffman P, et al. Comprehensive integration of single-cell data. *Cell* 2019;177:1888–1902.



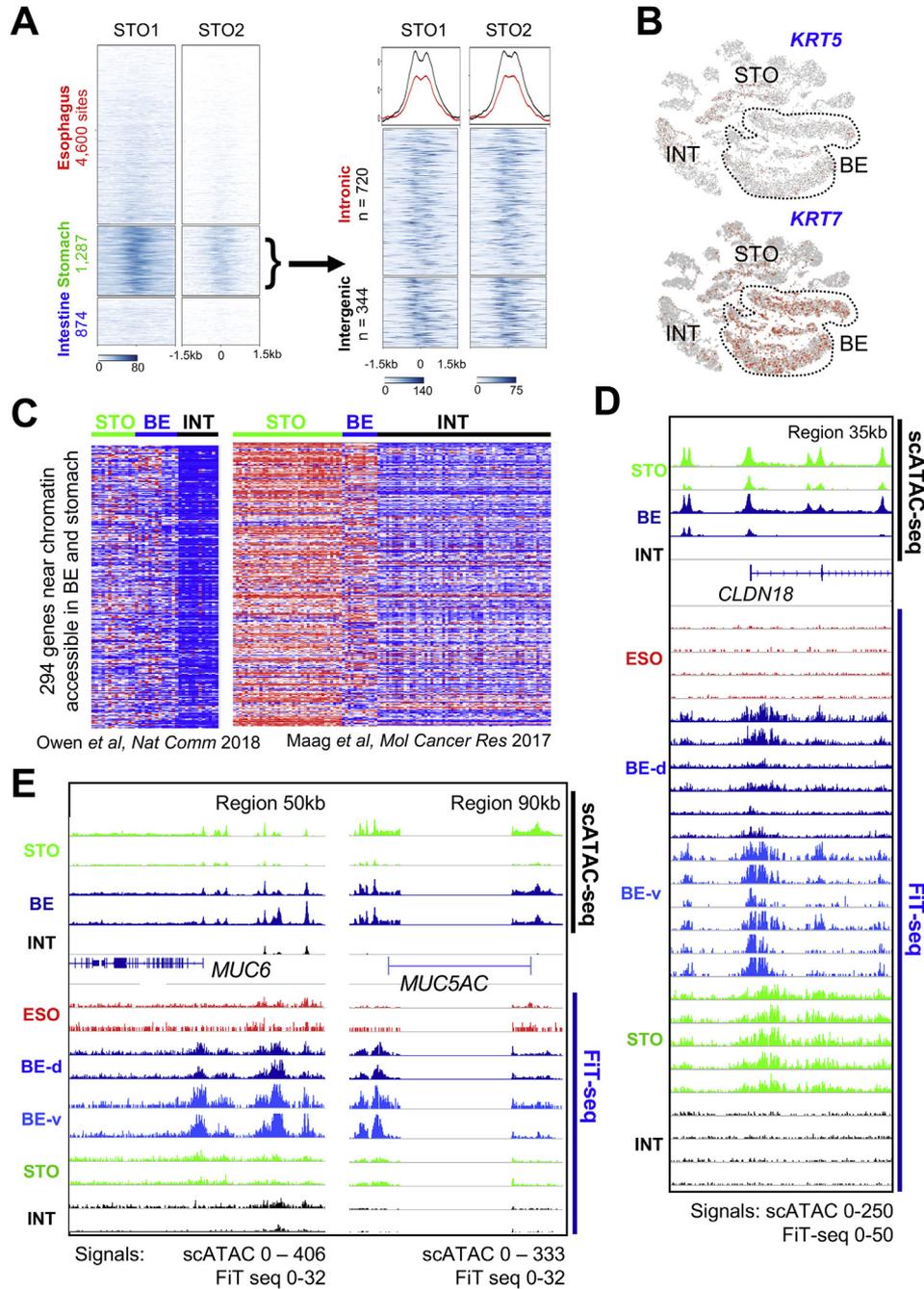
Supplementary Figure 1. *Cis*-elements derived from FiT-seq discern normal gut epithelia and BE. (A) Mouse intestinal villus enhancers identified by H3K27ac ChIP-seq show the presence of H3K4me2 (top). Scatterplot of log normalized counts (DESeq2) on a union set of enhancer (nonpromoter) peaks identified by H3K27Ac (y-axis) or H3K4me2 (x-axis) ChIP-seq (bottom). The 2 active histone marks are highly correlated at enhancers. (B) Representative photomicrographs of H&E-stained tissue sections, showing characteristics and high histologic purity of the specimens we chose to study. (C) Data tracks from H3K4me2 FiT-seq at representative tissue-restricted loci illustrate the specificity of enhancer marks at tissue-restricted genes. (D) Gap statistic showing 3–5 as the optimal number of clusters for *k*-means clustering of the top quintile of variable enhancers in normal esophageal, gastric, and intestinal mucosae. Visualization showed that beyond 3 clusters, only the esophageal group was further partitioned; we therefore considered the 3 esophageal enhancer subclusters as a single group. (E) GSEA of 500 genes (the default upper limit for GSEA) encoded <100 kb from each of the 3 tissue-specific enhancer sets. Enriched expression in the corresponding epithelium confirms function of the signature *cis*-element groups. (F) Principal component (PC) analysis of global enhancer landscapes. BE-discovery (BE-d) and -validation (BE-v) cohorts are similar and overlap most with the intestinal profiles. (G) Cumulative analysis of enhancers identified by H3K4me2 FiT-seq in 11 BE samples, showing slopes of 9733 sites across the first 5 samples and 3545 additional sites across the last 5. The sum of BE-d and BE-v encompasses most enhancers identified by FiT-seq. (H) DNA sequence motifs recognized by GATA- and SOX-family TFs are enriched among stomach-specific enhancers active in BE. NES, normalized enrichment score; P-val, *P* value.



Supplementary Figure 2. BE carries distinct stomach- but not esophagus-specific enhancers. (A) BE samples show enrichment of H3K4me2 at stomach-specific enhancers (Figure 1B). Here, 1287 gastric enhancers are divided into those overlapping with introns (n = 720) and those that lie in intergenic regions (n = 344). Both groups show comparable H3K4me2 enrichment. (B) Data tracks for H3K4me2 FiT-seq signal in the *CDX1* locus. In addition to the intestine, the enhancer locus is marked in BE but not in the esophagus or stomach, illustrating one of many intestinal properties acquired in BE. (C) Data for H3K4me2 FiT-seq in the locus encoding stomach-specific *TFF1* and *TFF2* and intestine-specific *TFF3*. Tissue-restricted enhancers are highlighted. *TFF1/2* enhancer marks in BE and their absence in intestine reflect the stomach enhancer fingerprint in BE, with additional acquisition of intestinal (*TFF3*) enhancer activity.

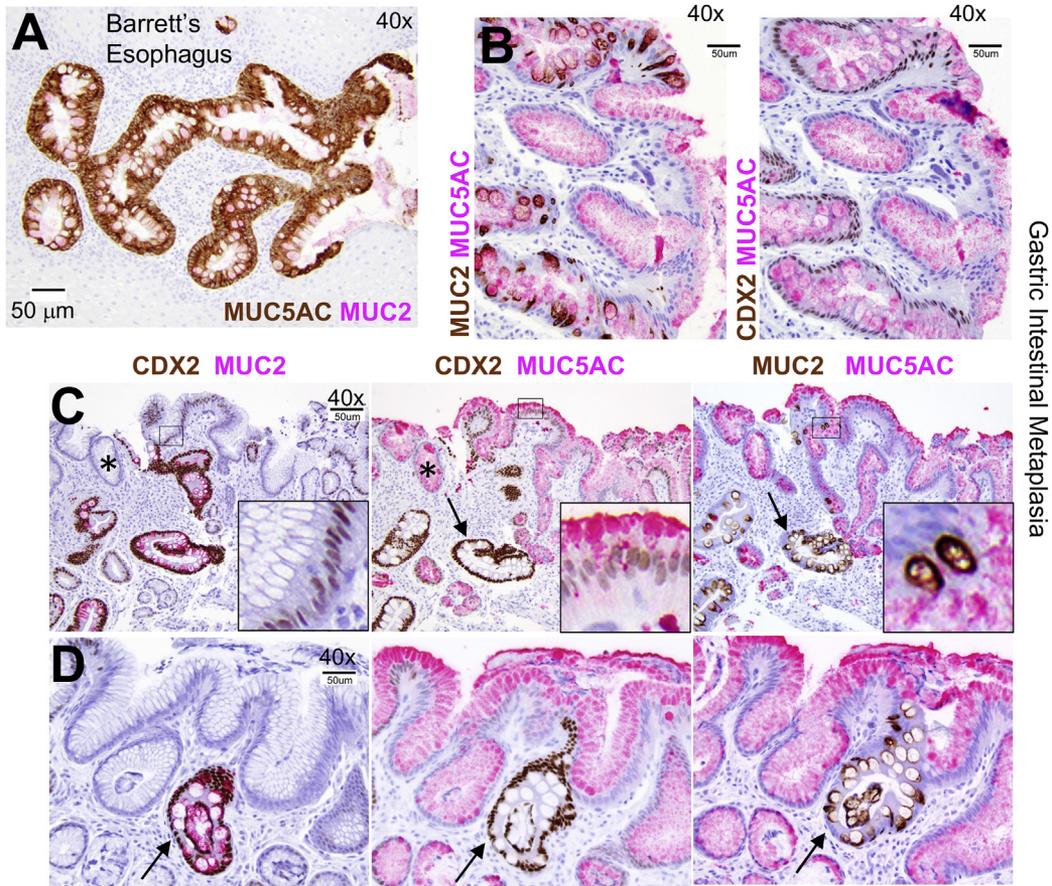


Supplementary Figure 3. Gastric epigenetic landscapes predict mutational profiles in esophageal and gastric adenocarcinomas. (A) A random forest machine-learning approach showed that enhancer landscapes in normal gastric and stratified esophageal epithelia best predict somatic mutation densities in gastric and squamous esophageal cancers, respectively, establishing the merits of the approach to define the tissue of origin. The x-axis shows the number of variables (tissue-specific enhancer profiles), and the y-axis shows the fraction of variation in cancer mutational profiles explained by each variable. The first column in each graph depicts the largest contributor; subsequent columns represent the incremental contribution of additional variables. (B) Boxplot representing the variance-explained values derived by using either all 4 or subsets of H3K4me2 chromatin features along with mutational profiles of BE or EAC samples. $***P < .001$. Removal of gastric epigenome but not intestinal epigenome led to a significant reduction in the mutational variance explained by the random forest model. (C) The stomach enhancer landscape is the best predictor of mutational frequencies in most EACs, similar to the BE specimens in Figure 6B. Of note, the BE enhancer landscape predicts EAC mutation frequency in a fraction of cases, likely reflecting that the majority of mutations in EAC arise before the development of metaplasia in most cases. (D) All informative single cells (after filtering by Cell Ranger criteria) from the tSNE plot shown in Figure 4. Cells are color-coded by ATAC-seq read depth (fragment counts, left). Promoter ATAC signals at mesenchymal gene *VIM* (center) and blood-cell marker *CD45* (right) distinguish nonepithelial clusters (nos. 15, 16, 17, and 21—dashed ovals) from epithelial cells. (E) scATAC-seq shows wide chromatin accessibility at *HOXA* and *HOXB* clusters selectively in BE, compared to stomach (STO) or intestine (INT). Aggregate (pseudo-bulk) scATAC signal from all epithelial cells in each sample are shown.



Supplementary Figure 4. scATAC-seq reflects activity of canonical stomach *cis*-elements in BE. (A) (Left) Pseudo-bulk ATAC-seq from the 2 stomach (STO) samples used for scATAC-seq. Signals from all epithelial cells in each sample were merged to create a pseudo-bulk ATAC signal that is projected on the esophagus- (ESO), stomach-, and intestine- (INT) specific enhancers defined by H3K4me2 FiT-seq (sites delineated in Figure 1B). Regions identified as H3K4me2⁺ enhancers thus show accessible chromatin in independent samples by an independent method, supporting their *cis*-regulatory roles. Normal stomach cells show lack of access at esophageal or intestinal enhancers. (Right) Stomach-specific enhancers were divided into those overlapping with introns (n = 720) and those lying in intergenic regions (n = 344). Both groups show open chromatin. Because global scATAC-seq signals differed in the 2 stomach samples, heatmaps are scaled separately (scales are shown). (B) Projection of scATAC signals at the *KRT5* and *KRT7* promoters onto the tSNE plot from Figure 4, showing extensive *KRT7* activity in BE. There is little *KRT5* activity in small fractions of the stomach and intestine (INT) and hardly any in BE cells. (C) Genes with selective chromatin accessibility in stomach and BE (not in intestine) by scATAC (n = 294) are expressed in the respective tissues, as seen in 2 public RNA data sets. Owen et al¹⁴: n = 13 each for BE, STO, and INT. Maag et al⁴¹: STO, n = 37; BE, n = 12; and INT, n = 59 from the Genotype-Tissue Expression consortium.⁴² In the heatmaps, blue indicates low and red indicates high RNA expression. (D, E) Stomach-restricted tight junction (*CLDN18*) (D) and mucin (*MUC5AC* and *MUC6*) (E) genes share open chromatin sites in stomach and BE cells, as detected by scATAC seq. (Pseudo-bulk signals from epithelial compartments are shown.) In the same regions, FiT-seq data show selective H3K4me2 marking in stomach and BE cells.

Supplementary Figure 5. scATAC-seq identifies novel gastric markers present in BE. (A) Examples of marker genes shared between stomach and BE (absent in intestine) based on scATAC-seq signals. Aggregate scATAC signals (pseudo-bulk) from epithelial cells are shown for *SLC45A3*, *CAPN9*, and *TMPRSS3*; the full gene list is provided in [Supplementary Table 5](#). Also shown are FIT-seq data for H3K4me2 from all 11 BE samples (*SLC45A3*) or, for economy of space, from 4 representative BE samples (d, discovery set; v, validation set) (see [Figure 1C](#)). (B) *ANXA10* mRNA levels in normal human tissues, showing unique gastric expression. Data are from the Genotype-Tissue Expression consortium.⁴² A tissue key is provided. (C) *ANXA10* immunostaining on human stomach and colon tissue sections. Only the gastric epithelium is stained.



Supplementary Figure 6. Dual stomach-intestinal cell state in GIM, similar to BE, with extensive heterogeneity. (A) Two-color immunohistochemistry on an additional BE case, showing extensive coexpression of MUC2 (intestinal) and MUC5AC (gastric) mucins in goblet-like BE cells. (B–D) CDX2 and stomach- (MUC5AC) or intestine- (MUC2) specific mucin staining in sequential tissue sections from 3 of 7 representative cases of human GIM examined by 2-color immunohistochemistry; B–D represent 3 independent samples, each showing extensive cell state heterogeneity among adjoining superficial epithelium and glands. (B) Adjacent CDX2⁺ and CDX2⁻ areas both express MUC5AC, and goblet-like cells in CDX2⁺ glands coexpress MUC2. (C) Surface foveolar MUC5AC⁺ cells have intermediate CDX2 levels and no MUC2⁺ cells (*inset*), whereas contiguous deep glands carry high CDX2 and MUC2⁺ MUC5AC^{low} goblet-like cells (*arrows*). (D) Extreme intestinalization (CDX2⁺ MUC2⁺ MUC5AC^{low} goblet-like cells) of a discrete GIM gland (*arrows*) surrounded by MUC2⁻ MUC5AC^{hi} gastric glands.