Gastrointestinal transcription factors drive lineage-specific developmental programs in organ specification and cancer

Roshane Francis 1,2*, Haiyang Guo 3*, Catherine Streutker 4, Musaddeque Ahmed 3, Theodora Yung 1,2, Peter B. Dirks 1,2, Housheng Hansen He 3,5†, Tae-Hee Kim 1,2†

Transcription factors (TFs) are spatially and temporally regulated during gut organ specification. Although accumulating evidence shows aberrant reactivation of developmental programs in cancer, little is known about how TFs drive lineage specification in development and cancer. We first defined gastrointestinal tissue–specific chromatin accessibility and gene expression during development, identifying the dynamic epigenetic regulation of SOX family of TFs. We revealed that Sox2 is not only essential for gastric specification, by maintaining chromatin accessibility at forestomach lineage loci, but also sufficient to promote forestomach/esophageal transformation upon Cdx2 deletion. By comparing our gastrointestinal lineage-specific transcriptome to human gastrointestinal cancer data, we found that stomach and intestinal lineage-specific programs are reactivated in Sox2 high/Sox9 high and Cdx2 high cancers, respectively. By analyzing mice deleted for both Sox2 and Sox9, we revealed their potentially redundant roles in both gastric development and cancer, highlighting the importance of developmental lineage programs reactivated by gastrointestinal TFs in cancer.

INTRODUCTION

The gastrointestinal (GI) tract is one long tube consisting of the esophagus, the stomach, the small intestine, and the large intestine (i.e., colon), and its functions are to digest food, absorb nutrients, and excrete waste. Because of genetic inadequacies and/or environmental insults, cancers of the GI tract are among the most common and deadliest in the world (1).

In select GI cancers, epithelial metaplasia or transdifferentiation into adjacent organ identities precedes malignant transformation and is associated with the expression of lineage-specifying factors. For example, a subtype of gastric and esophageal cancers exhibits intestinal metaplasia, characterized by the presence of ectopic intestinal cell types in the proximal organs (2). Moreover, the recent temporal transcriptome and proteomic analyses of the mouse stomach showed that many developmentally regulated factors are abnormally reactivated in gastric cancer (3). Since these factors are likely regulated by transcription factors (TFs), a better understanding of GI lineage-specifying developmental programs directed by TFs would provide important mechanistic insight into GI tumorigenesis.

During GI development, the stomach, the small intestine, and the colon are derived from an elongating tube of endodermal cells, termed the primitive gut tube (PGT). In mice, the PGT is first formed at embryonic day 8.5 (E8.5), when the restricted expression of TFs localizes to distinct regions of the PGT. Sex-determining region Y-box 2 (SOX2), pancreatic and duodenal homeobox 1 (PDX1), and caudal type homeobox 2 (CDX2) are expressed in the foregut, midgut, and hindgut, respectively (4). As development proceeds, the foregut will give rise to the esophagus, the lung, the stomach, and the proximal intestine, while the mid- and hindguts will develop into the pancreas, the liver, the remaining small intestine, and the colon. In more detail, the murine stomach will first regionalize into the squamous forestomach and the glandular hindstomach. Further hindstomach maturation yields two subregions: the corpus, which houses H+ K+ adenosine triphosphatase–positive (ATPase +) parietal cells, and the antrum, which contains PDX1+ epithelial cells (5).

In the pancreas, the lung, and the liver, tissue lineage decisions are initiated by organ-specific master regulators, which activate organ-specific transcriptional programs (6). While CDX2 has been shown to be critical for specifying an intestinal fate through the regulation of intestinal enhancers (7–9), it remains unclear which factors are necessary for gastric specification and maturation (5). Cdx2 deletion led to a notable forestomach/esophageal transformation and ectopic expression of Sox2, suggesting their antagonistic relationship. Previous studies also have provided evidence for the role of Sox2 in gastric development and homeostasis. Analyses of mice with various Sox2 hypomorphic alleles revealed foregut defects and altered expression of GI cell type markers (10). However, the overall gastric morphology and specification in these mice were still maintained. In addition, although Sox2 labels gastric stem cells, it is dispensable during adult homeostasis; unexpectedly, its deletion in Wnt signaling–activated gastric cancer promoted gastric tumorigenesis, implying potential redundancy (11, 12). While these studies have implicated Sox2 in gastric development and cancer, its role in gastric lineage specification and the underlying mechanisms are still unclear.

To define regulation of GI specification mediated by GI TFs, we analyzed enhancers and gene expression during mouse stomach development, identifying the dynamic expression of SOX family TFs in gastric development. Subsequently, we elucidated the critical role of Sox2 in gastric lineage specification through the regulation of forestomach enhancers and defined its genetic relationship with Cdx2. Moreover, by comparing our GI developmental gene expression

1Program in Developmental and Stem Cell Biology, The Hospital for Sick Children, Toronto, Ontario M5G 0A4, Canada. 2Department of Molecular Genetics, University of Toronto, Toronto, Ontario MSS 1A8, Canada. 3Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario MSS 1L7, Canada. 4Department of Laboratory Medicine, St. Michael’s Hospital, Toronto, Ontario M5B 1W8, Canada. 5Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 2M9, Canada. 6These authors contributed equally to this work.

†Corresponding author. Email: tae-hee.kim@sickkids.ca (T.-H.K.); hansenhe@uhnresearch.ca (H.H.H.)
data to the human GI cancer data, we have shown that stomach and intestinal lineage programs are enriched in Sox2\textsuperscript{high}/Sox9\textsuperscript{high} and Cdx2\textsuperscript{high} GI cancers, respectively. By analyzing mice deleted for both Sox2 and Sox9 in our gastric development and cancer models, we have revealed their potential redundancy. Together, our study highlights the importance of tissue-specific programs reactivated by GI lineage TFs in cancer.

**RESULTS**

**Analyses of chromatin accessibility and gene expression during GI development identify dynamic epigenetic regulation of SOX family of TFs in gastric development**

To define changes in the chromatin landscape as GI organs specify into their associated regions, we isolated epithelial cells from E13.5 and E16.5 murine guts and preformed assay for transposase-accessible chromatin sequencing (ATAC-seq). On the basis of the E16.5 forestomach, hindstomach, small intestine, and colon ATAC-seq profiles, we identified peaks enriched within the gastric and intestinal regions and those overlapping between them (Fig. 1A). These peaks primarily localized to promoter and distal intergenic regions, suggesting enrichment of DNA regulatory elements such as enhancers (fig. S1A). We found that region-enriched chromatin became accessible at E16.5 upon completion of regionalization (Fig. 1A). For example, gastric and intestinal lineage-specific genes such as gastronin-1 (Gkn1) and fatty acid–binding protein-2 (Fabp2) were inaccessible in E13.5 but became accessible in E16.5 (Fig. 1B). In contrast, the chromatin near the FoxA1 (a pan-endodermal marker) locus was accessible in both stomach and intestinal tissues at both stages (Fig. 1B). To identify transcriptionally regulated genes critical for regionalization, we also conducted RNA sequencing (RNA-seq) using epithelium isolated from the same E16.5 GI regions. Comparison of all stomach genes versus all intestinal genes showed unique gene signatures between the two organs, indicating organ-specific transcriptional programs (fig. S1B). Unsupervised hierarchical clustering of the 2000 most differentially regulated genes showed similarity between the hindstomach and the intestinal organs (Fig. 1C), and genes up-regulated in each region were significantly enriched within the matched ATAC-seq profiles, suggesting epigenetic regulation of GI region–specific genes (Fig. 1D). To further define epigenetically active genes, we also analyzed H3K27ac modification, a histone mark signifying active DNA regulatory elements such as enhancers and gene promoters (13, 14), in the four E16.5 GI regions. Peak distribution confirms that most H3K27ac chromatin immunoprecipitation sequencing (ChIP-seq) peaks were found at promoter and distal intergenic regions, conferring regulation via promoters and enhancers (fig. S1C). In addition, we observed that while the chromatin near developmental and region-enriched genes remained open in all GI regions at E16.5, genes marked by H3K27ac conferred region specificity. For example, the chromatin near the sox4 (an endodermal marker gene) locus was open and epigenetically active in all GI organs, while Sox2 and Cdx2 enhancers were active only in the stomach and intestinal tissues, respectively (Fig. 1E). To identify key transcriptional regulators of GI development, we conducted TF motif analysis of E13.5 and E16.5 region-specific ATAC-seq profiles and identified the SOX TFs as one of the most enriched TF families in both E13.5 stomach peaks and E16.5 forestomach/hindstomach common peaks (Fig. 1F). Immunohistochemistry analysis of E13.5 stomachs showed an enrichment of SOX2 and SOX9 (fig. S2A).

**Sox2 is critical for gastric specification and regionalization**

To determine the role for Sox2 in gastric development, we generated Sox2 knockout (KO) mice (Sox2\textsuperscript{CreERT2/flox}) and administered tamoxifen at E8.5 to bypass early embryonic lethality observed in Sox2 KO embryos (15). Upon complete deletion of Sox2, whole-mount images of Sox2 KO mice and hematoxylin and eosin (H&E) analysis revealed a marked reduction in stomach size and loss of squamous epithelium compared to controls (Fig. 2, A and B). Immunohistochemistry analysis of the mutant epithelium demonstrated an absence of transformation-related protein 63 (TP63), a forestomach marker, but properly localized expression of H\textsuperscript{+}K\textsuperscript{+} ATPase and PDX1 (i.e., hindstomach markers) (Fig. 2C) and CDX2, an intestinal marker (fig. S2B). These results suggest that Sox2 is critical for gastric growth and forestomach specification, while it is dispensable for hindstomach regionalization. We next examined the lineage trajectory of endodermal progenitors upon loss of Sox2. To conduct cell lineage–tracing experiments, we crossed the Sox2 KO mutants with R26\textsuperscript{ERT2} reporter mice in which Sox2\textsuperscript{CreERT2} and their descendant cells were permanently labeled with green fluorescent protein (GFP) upon tamoxifen induction at E8.5, while Sox2\textsuperscript{CreERT2} cells expressed TdTomato (Fig. 2D). We observed more GFP\textsuperscript{+} cells in the distal region of the small intestine in Sox2-deleted mice compared to the controls (Fig. 2E). These Sox2\textsuperscript{+}/GFP\textsuperscript{+} cells in the mutant intestines were negative for intestinal (i.e., Cdx2 and Tff3) and gastric markers (Tp63 and H\textsuperscript{+}K\textsuperscript{+} ATPase), unlike Sox2\textsuperscript{+}/GFP\textsuperscript{+} cells in the control intestines (fig. S2C). Collectively, these results suggest that Sox2-deleted endodermal progenitors lose the ability to respond to niche signals, and thus, fail to acquire intestinal identity properly.

**Sox9 compensates for the loss of Sox2 and maintains residual gastric development**

Residual gastric development maintained upon the loss of Sox2 implies potential redundancy of additional factors responsible for preserving gastric specification (Fig. 2, A to C). Since we identified the expression of Sox9 in the developing stomach (fig. S2A), we examined its expression in the Sox2 KO embryos and found that Sox9 became strongly activated in the proximal mutant stomach (Fig. 3A). Analysis of forestomach and hindstomach ATAC-seq peaks showed that Sox2 and Sox9 are broadly accessible despite their region-specific expression (Fig. 3B). These data suggest that the gastric chromatin is primed and can support either SOX TF expression. To determine their potential redundancy in gastric development, we generated mice conditionally deleted for both Sox2 and Sox9 [double knockout (DKO)] (Sox2\textsuperscript{CreERT2/flox};Sox9\textsuperscript{floxed}) at E8.5, the onset of gut organogenesis. Despite efficient deletion of Sox2, these DKO embryos still maintained gastric specification due to the Cre escapers expressing Sox9 (Fig. 3C and fig. S3A). Accordingly, staining for region-specific markers showed complete loss of the forestomach marker TP63, while the hindstomach/antral marker PDX1 was maintained (fig. S3A). To circumvent the temporally restricted activity of Sox2\textsuperscript{CreERT2}, we used a constitutively active, gut epithelial Cre, Shh\textsuperscript{Cre}, to replicate the Sox2;Sox9 DKO experiments. While we observed a drastic reduction in organ size by H&E staining, we similarly observed the presence of Sox9’ gastric epithelial cells in the DKO mutants, masking the true DKO phenotype (fig. S3B).

In addition, when we generated Sox9 KO embryos using an alternative hindstomach epithelial Cre, Pdx1\textsuperscript{Cre}, we observed the same Cre escapers (Fig. 3D) and maintenance of hindstomach differentiation by H\textsuperscript{+}K\textsuperscript{+} ATPase staining (fig. S3C). This result
Fig. 1. SOX TFs are enriched in the stomach as the chromatin landscape acquires accessibility during development. (A) ATAC-seq profiles of E13.5 stomach and intestine compared to E16.5 forestomach, hindstomach, intestine, and colon. Region-enriched peaks are determined on the basis of E16.5 profiles, and their corresponding accessibility is analyzed in the E13.5 tissues. (B) Representative examples of ATAC-seq tracks show acquired chromatin accessibility in E16.5 stomach-specific (Gkn1) and intestinal-specific (Fabp2) loci as the GI organs become regionalized, while common, endodermal loci such as FoxA1 are accessible throughout gut development. (C) Unsupervised hierarchical clustering of E16.5 forestomach, hindstomach, small intestine, and colon transcriptomes based on the top 2000 most variable genes. (D) Binding and Expression Target Analysis (BETA) comparing association of region-specific ATAC-seq peaks to genes differentially regulated in each region at E16.5. (E) Representative examples of H3K27ac ChIP-seq and ATAC-seq peaks of common (Sox4), stomach-specific (Sox2), and intestinal-specific (Cdx2) active DNA regulatory elements such as promoters and enhancers are shown. (F) TF motif analysis of all E13.5 stomach and E16.5 stomach common ATAC-seq peaks shows enrichment of SOX TF motif. Rank indicates the ranking of the motif enrichment in all 637 mouse TF motifs. For ATAC-seq (n = 1 per group), six to eight embryonic guts were pooled. For H3K27ac ChIP-seq (n = 1 per group), 25 to 30 embryonic guts were pooled. For RNA-seq (n = 2 per group), 13 to 15 embryonic guts were pooled. ST, stomach; INT, intestine; FST, forestomach; HST, hindstomach; SI, small intestine; COL, colon.
Fig. 2. Sox2 is essential for gastric specification and regionalization. (A) Whole-mount images and subsequent histological analyses using H&E staining of control and Sox2-deleted guts at E18.5 (n ≥ 3 each). Tamoxifen was administered at E8.5. Scale bars, 200 μm. (B) Higher magnification images of gastric regions highlight the loss of squamous, forestomach and epithelium and the maintenance of glandular, hindstomach morphology in Sox2 KO mutants. Scale bars, 100 μm. (C) Immunohistochemistry of regional gastric markers reiterates the loss of TP63+ forestomach epithelium but the proper expression of hindstomach markers, H+ K+ ATPase and PDX1, in Sox2 KO embryos (n ≥ 3 each). Scale bars, 100 μm. (D) Schematic of alleles used to generate conditional Sox2-deleted embryos using tamoxifen-induced Cre-LoxP recombination for cell lineage–tracing experiments is shown. (E) Schematic (left) and fluorescence images (right) of cell lineage tracing of Sox2-deleted endodermal progenitors throughout gut development and their allocation into the different regions of the GI tract (n = 3 each). Scale bars, 100 μm.
was consistent with the incomplete recombination of the Sox9 floxed allele previously reported during pancreatic development (Pdx1Cre;Sox9flx/flx) (16). Therefore, our findings imply that Sox9 is critical for gastric epithelial survival during development, as the Cre escapers expressing Sox9 have a selective advantage and are able to maintain gastric specification in the absence of Sox2.

**SOX2 directly regulates forestomach lineage-specific genes by maintaining chromatin accessibility at their loci**

Since our data demonstrated that Sox2 deletion leads to the complete loss of forestomach identity, we hypothesized that it regulates the expression of genes critical for forestomach specification. To address this hypothesis, we performed SOX2 ChIP-seq experiment using E16.5 forestomach epithelium. We compared Sox2 binding sites to the E16.5 forestomach ATAC-seq profile (i.e., the tissue highly expressing Sox2) and the E16.5 small intestine ATAC-seq profile (i.e., the tissue lacking Sox2 expression). We found the enrichment of SOX2 + peaks in the forestomach profile compared to the intestinal profile (Fig. 4A). Approximately 66% of all Sox2 binding sites localized to promoters and distal intergenic regions, suggesting that Sox2-directed regulation occurs at promoters and enhancers, respectively (Fig. 4B and fig. S1D).

We then compared the SOX2 targets to the 2100 most up-regulated genes in the forestomach transcriptome. Notably, we found that more than 10% of highly expressed forestomach genes were direct targets of Sox2 (Fig. 4C). These targets were enriched for Gene Ontology (GO) terms such as keratinocyte differentiation and epidermis development, which are features of the squamous epithelium found in the forestomach and the esophagus (Fig. 4D). In addition, co-TF motif analysis of all SOX2 + peaks identified TP63, a critical forestomach stem cell marker, as the top-ranked co-binding TF, suggesting their transcriptional cooperation (Fig. 4E) (17).

Since Sox2 is known to act as a pioneering factor in reprogramming (18), we examined whether Sox2 loss changes the chromatin landscape during gastric development. To address this question, we performed ATAC-seq using E16.5 Sox2-deleted gastric epithelium. Analyzing wild-type E16.5 forestomach, E16.5 hindstomach, and Sox2 KO whole-stomach ATAC-seq profiles, we found the complete loss of forestomach-enriched peaks that were originally identified in Fig. 1A, suggesting lack of chromatin changes associated with regionalization (Fig. 4F). For example, ATAC-seq traces near the locus of Krt8, a SOX2-regulated gene expressed in adult gastric glands (19), showed a drastically reduced level of chromatin accessibility (Fig. 4G). Corroborating these data, regression analysis showed that the E16.5 Sox2 KO whole-stomach ATAC-seq profile is more similar to the E13.5 stomach (i.e., less differentiated organ) ATAC-seq profile than the E16.5 fore- or hindstomach profiles (fig. S1E).

However, the common peaks accessible in all of the developing GI organs were still maintained in the Sox2-deleted gut. For instance, the chromatin near the loci of SOX2 + endodermal targets such as Sox21 and Sox4 was still accessible, suggesting that additional co-factors are able to preserve chromatin accessibility despite Sox2 deletion (Fig. 4G). Nevertheless, these results demonstrate the importance of Sox2 in gastric specification by maintaining chromatin accessibility at forestomach lineage-specific loci.

**Loss of Sox2 rescues forestomach/esophageal transformation induced by Cdx2 deletion**

A previous study showed that the deletion of the intestinal-specific factor Cdx2 in the developing endoderm transformed the intestinal,
Fig. 4. SOX2 directly induces forestomach lineage-specific genes by maintaining chromatin accessibility at their loci. (A) SOX2 ChIP-seq peaks are enriched in the E16.5 forestomach ATAC-seq profile versus the E16.5 intestinal ATAC-seq profile, which lacks Sox2 expression. For SOX2 ChIP-seq ($n = 1$), 30 embryonic stomachs were pooled. (B) Genomic distribution of SOX2 ChIP-seq peaks. 5′UTR, 5′ untranslated region. (C) Venn diagram outlining the overlap between SOX2 target gene associations and the top 2100 most highly expressed genes in the E16.5 forestomach transcriptome ($P < 0.05$ and fold change $> 1.5$ when compared to the E16.5 small intestinal transcriptome). FC, fold change. (D) GO term analysis of SOX2-regulated forestomach genes. (E) Co-TF motif analysis shows enrichment for motifs important in squamous epithelium (TP63), endodermal development (SOX4), and embryonic stem cells (Kruppel-like factor 4 [KLF4]). (F) Comparison of E16.5 forestomach and hindstomach and E13.5 stomach ATAC-seq heat maps to E16.5 Sox2 KO stomach ATAC-seq data shows loss of forestomach-enriched peaks. The peaks are categorized on the basis of the common and region-enriched peaks defined in Fig. 1. (G) Alignment of E16.5 forestomach and hindstomach ATAC-seq and H3K27ac ChIP-seq tracks with SOX2 ChIP-seq to show examples of SOX2 directly regulated genes (the green box highlights Sox2 binding site in proximity to the gene of interest). E16.5 Sox2 KO ATAC-seq profile demonstrates that forestomach lineage-specific open chromatin regions become inaccessible in Sox2 KO embryos, but the same is not observed for Sox2-regulated endodermal genes (the red boxes outlining corresponding ATAC-seq signal).
villous epithelium into gastric epithelium (7). This mutant epithelium ectopically expressed SOX2 and TP63 and acquired a forestomach, squamous cellular morphology. However, the role of these TFs in gastric transformation is still unknown. Given our findings demonstrating the critical role of SOX2 in gastric specification, we hypothesized that the ectopic expression of SOX2 in Cdx2 KO embryos might be sufficient to drive forestomach/esophageal transformation. To address this hypothesis, we generated embryos conditionally deleted for both Sox2 and Cdx2 at E8.5 (R26CreERT2;Cdx2floxtm1Sox2floxtm1). We observed the partial restoration of villous/glandular-like epithelium in these embryos (Fig. 5A). The ectopic expression of the forestomach marker, TP63, was completely lost, while the pattern of proliferating cell nuclear antigen–positive (PCNA+) cells was partially restored (Fig. 5B). Further characterization of the Sox2;Cdx2 KO intestinal epithelium revealed the absence of intestinal markers (i.e., TTF3;Alcian blue+ goblet cells and alkaline phosphatase+ enterocytes) (fig. S4A) and the hindstomach/corpus marker H+ K+ ATPase (fig. S4B). The DKO intestinal epithelium stained positive for gastric-type surface mucous cells, evidenced by periodic acid– Schiff staining and the hindstomach/antral marker PDX1, suggesting antral transformation of DKO intestinal epithelium (fig. S4B). Furthermore, these Cdx2 and Sox2 DKO embryos ectopically activated Sox9 expression in the distal intestinal epithelium (Fig. 5B), recapitulating the Sox2–Sox9 dynamics observed in the Sox2 KO embryos (Fig. 3A). In addition to the villous-like morphology (Fig. 5A) and the intervillous restriction of PCNA+ proliferating cells (Fig. 5B), these results further support an antral/proximal intestine–like identity of Sox2;Cdx2 DKO epithelial cells.

To define the epigenetic regulation of gastric transformation induced by Cdx2 deletion, we compared our E16.5 forestomach ATAC-seq and SOX2 ChIP-seq profiles to E16 Cdx2 KO ATAC-seq data generated by Banerjee et al. (9). We found that Cdx2 KO embryos contained ectopic activation of forestomach–specific ATAC-seq peaks that were partially associated with Sox2 binding (Fig. 5C). Together, our results demonstrate that the ectopic expression of Sox2 is responsible for squamous transformation induced by Cdx2 deletion.

**Developmental stomach and intestinal programs are enriched in Sox2high- and Cdx2high-expressing GI cancers, respectively**

Multiple GI cancers show an up-regulation of SOX2, SOX9, and CDX2 (20–22). Although the role of Sox2 and its transcriptional mechanisms in squamous cell esophageal carcinomas have been elucidated (17, 23), its role in gastric cancer is still controversial: It has been proposed to act in a redundant manner to compensate for the loss of either SOX TF during cancer initiation. Notably, this Sox2–Sox9 dynamic was also observed during gastric embryonic development: The loss of Sox2 up-regulated Sox9 and subsequently maintained residual gastric specification (Fig. 3A). To examine whether Sox2 and Sox9 act in a redundant manner during cancer initiation, we generated Sox2 and Sox9 DKO mice in the adenoma background. These mice exhibited more organized glandular structures and reduced Alcian blue staining in comparison to single-KO adenoma mice (Fig. 7B). Histopathology scoring confirmed significant reductions in disease severity and the presence of dysplastic epithelium when comparing DKO adenoma mice to their single-KO counterparts (fig. S5, A and B). These data demonstrate that both SOX TFs cooperatively regulate genes that are important during cancer initiation, and single SOX TF deletions are inadequate in preventing gastric cancer initiation.

**DISCUSSION**

As the endoderm and mesoderm assemble into the PGT and gut development proceeds, organ and regional boundaries become specified through the expression of key developmental TFs (4). These factors would likely bind to unique DNA regulatory elements such as H3K27ac-rich promoters and enhancers of fate determining genes, thereby regulating expression in a spatial and temporal manner. To understand global chromatin and gene expression changes during GI development, we analyzed chromatin accessibility and the H3K27ac-active enhancer mark throughout all regions of the GI tract, which include the colon, and found both broadly permissive and region–enriched chromatin regions during gut development. In identifying the dynamic expression of SOX family TFs in gastric development, we focused on the roles of Sox2 and Sox9. We demonstrated that Sox2 is critical for gastric specification and regionalization by directly regulating the expression of forestomach–specific genes and chromatin accessibility at their loci.

In the absence of Sox2, Sox9 was up-regulated and maintained hindstomach specification. Because of inefficient Cre-LoxP excision, Sox9 expression was still maintained in Sox9-deleted embryos, masking the true Sox9 KO, and Sox2,Sox9 DKO phenotypes. The
Fig. 5. Loss of Sox2 rescues forestomach/esophageal transformation induced by Cdx2 deletion. (A) Whole-mount images and subsequent histological analyses using H&E staining of control and Sox2;Cdx2 doubly deleted guts at E18.5 (n ≥ 3 each). Cre-LoxP recombination was induced at E8.5. Scale bars, 100 μm. (B) Immunofluorescence images show CDX2, SOX2, TP63, PCNA, and SOX9 staining in control, Sox2 KO, Cdx2 KO, and Sox2;Cdx2 DKO embryos. Scale bars, 100 μm. (C) Comparison of E16.5 common-, forestomach-, and small intestine–enriched ATAC-seq peaks in publicly available Cdx2 KO ATAC-seq data. Ectopic forestomach-specific peaks in Cdx2 KO embryos are associated with SOX2 binding (SOX2 ChIP-seq). Cdx2KO ATAC-seq was performed in E16 intestine tissue by Banerjee et al. and retrieved from GSM3181688. SOX2 ChIP-Seq was performed in E16.5 forestomach epithelium.
Fig. 6. Developmental stomach and intestinal programs are enriched in Sox2<sup>high</sup>- and Cdx2<sup>high</sup>-expressing GI cancers, respectively. (A) Top: Differential genes expressed in Sox2<sup>high</sup>, Sox9<sup>high</sup>, and Cdx2<sup>high</sup> human stomach cancer (STAD) transcriptomes are compared to our E16.5 GI RNA-seq profiles. Bottom: Gene set enrichment analysis reveals strong associations between Sox2<sup>high</sup>/Sox9<sup>high</sup> cancer genes and the E16.5 stomach profiles and between Cdx2<sup>high</sup> cancer genes and the E16.5 intestinal profiles. (B) Similar analysis conducted using human colon cancer (COAD) transcriptomes from TCGA shows the activation of stomach and intestinal lineage-specific genes in Sox2<sup>high</sup>/Sox9<sup>high</sup> and Cdx2<sup>high</sup> cancers, respectively.
Cre escapers still expressing Sox9 had a selective survival advantage over Sox9−/− cells and were able to maintain residual gastric specification.

A previous study showed that Sox2 overexpression is sufficient to induce gastric cell type markers, but CDX2 expression and villus-like glands were still maintained (30), in contrast to the complete forestomach/esophageal transformation upon Cdx2 deletion at E8.5 (7). The Sox2 overexpression study used Villin-rtTA transgenic mice, in which Sox2 expression begins at E13.5. Given the various phenotypes induced by Cdx2 deletions at different developmental stages due to chromatin-mediated developmental plasticity (8, 9), it would be interesting to examine Sox2 overexpression at earlier stages. Sox2 deletion at E8.5 followed by lineage tracing showed that despite the fact that Sox2-deleted gastric progenitors were properly incorporated into the intestine, they failed to express intestinal genes, suggesting a deficiency in responding to niche signals. Therefore, this study also highlights potentially distinct epigenetic regulation of the stomach and intestine mediated by Sox2 and CDX2, respectively.

Since developmental programs are known to be frequently reactivated in cancer, we also investigated their significance in cancer by comparing our GI developmental and available adult cancer transcriptomes. We found an enrichment of forestomach, hindstomach, and intestinal lineage-specific developmental genes in Sox2high−, Sox9high−, and Cdx2high−, expressing gastric and colorectal cancers, respectively. To determine the importance of developmental lineage programs induced by these TFs in GI cancers, we examined the roles of Sox2 and Sox9 in gastric cancer mice. Consistent with the Sox2-Sox9 dynamics observed during development, we found that a single deletion of either TF leads to activation of the reciprocal factor, while the double deletion of Sox2 and Sox9 in the adenoma background ameliorates cancer progression when compared to single-KO adenoma models. This SOX TF redundancy provides new insight into the complex roles of Sox2 proposed in gastric cancer (12, 24).

A recent study has established CDX2 as a prognostic biomarker in stage II and stage III colon cancer, suggesting that a subgroup of patients with stage II colon cancer might benefit from adjuvant chemotherapy based on the lack of CDX2 expression in cancer stem cells (31). Notably, SOX2 expression has been associated with a colon cancer stem cell state and down-regulation of CDX2 (22). This relationship between SOX2 and CDX2 in human colon cancer is consistent with our finding of their genetic relationship during mouse GI specification. Therefore, a better understanding of these developmental lineage-specific TFs and their epigenetic mechanisms would likely provide significant insight into the disease mechanisms of human GI cancers.

**Fig. 7. Single deletion of Sox2 or Sox9 in gastric adenoma mice increases cancer severity, whereas DKO rescues tumor severity when compared to singly deleted counterparts.** (A) SOX9 immunohistochemistry (top) and Alcian blue staining (bottom) show SOX9 overexpression at sites of adenomatous epithelium in early (8 weeks) and late (18 weeks) adenoma mice (n = 3 each). Scale bars, 100 μm. (B) Corresponding H&E (top) and Alcian blue (bottom) staining shows the aberrant changes in gastric epithelial structure and cell types in the various adenoma models (n ≥ 3 for Atp4bCre;R26NICD, Atp4bCre;R26NICD;Sox9flox/flox, Atp4bCre;R26NICD;Sox2flox/flox, and Atp4bCre;R26NICD;Sox2flox/flox;Sox9flox/flox mice). Scale bars, 100 μm. (C) SOX2 immunohistochemistry in Atp4bCre;R26NICD;Sox9flox/flox mice (top) and SOX9 immunohistochemistry in Atp4bCre;R26NICD;Sox2flox/flox;Sox9flox/flox mice (bottom) demonstrate reciprocal activation in the absence of either SOX TF (n = 3 each). Scale bars, 100 μm.
MATERIALS AND METHODS

Experimental animals

All animal husbandry was conducted in adherence to the standards proposed by the Canadian Council on Animal Care and approved by the institutional animal care committees of The Centre for Phenogenomics, Toronto. Shh(Cre)EgfP (32)/R26MmTr/Gc (33)/Pdx1Crem (34), Cdx2flxed (35)/R26NICE (36), R26CreERT2 (37), and Atbp4Cre (38) mice were obtained as gifts from C.-c. Hui, R. Shivdasani, J. Rossant, and J. Mills, respectively. Sox2CreERT2, Sox9flxed, and Sox9flxed mice were purchased from the Jackson Laboratory.

Tamoxifen oral gavage to pregnant dams

Murine females aged 15 weeks were used for embryonic tamoxifen administration via oral gavage. On E8.5 and E10.5, 100 μl of tamoxifen (20 mg/ml) and progesterone (10 mg/ml) in sunflower seed oil was administered to the dam on each day. Females were euthanized at the appropriate end point, and embryos were immediately collected via cesarean section for histological and high-throughput experiments.

Embryonic epithelial isolation for ATAC-seq, RNA-seq, and ChIP-seq

For E13.5 mouse experiments, ShhCre;R26YFP/+ embryos were used for downstream high-throughput experiments. After dissection, the stomach and intestinal organs were isolated, separated, washed in cold phosphate-buffered saline (PBS) and digested in 4 ml of a 3:1 solution of trypsin LE: 1× PBS at 37°C. After 10 min, tissue was vigorously pipetted until the organs broke down into a single-cell suspension, which was confirmed using bright-field microscopy. The enzymatic reaction was then inhibited with 10% fetal bovine serum. After washing cells in cold PBS, the single cells were resuspended in 400 μl of 1:5000 SYTOX Blue (Thermo Fisher Scientific, S34857) in cold 1× PBS, filtered through a 30-μm mesh and submitted for sorting. Sorted cells were washed again and processed for ATAC-seq [50,000 cells/400 μl of 1:5000 SYTOX Blue, and submitted for sorting. Sorted cells were then used for ATAC-seq using the same protocol as mentioned above.

Chromatin immunoprecipitation sequencing

After isolation of embryonic epithelium using dispase for forestomach and EDTA for the hindstomach, intestine, and colon, epithelial pellets were fixed in 1% paraformaldehyde for 10 min at room temperature, quenched with one-twentieth volume of 2.5 M glycine for 5 min, washed twice with cold 1× PBS, and flash-frozen. When enough sample was acquired (approximately 0.1 g of tissue per region), fixed pellets were pooled, dounced in cold PBS with protease inhibitor, and filtered through a 70-μm mesh. Samples were then incubated in cold cell membrane and nuclear lysis buffers containing non-denaturing detergents for 30 min each on a shaker at 4°C. Washed samples were then pelleted and resuspended in 300 μl of sonication buffer containing 0.1% SDS. Sonication was conducted using the Diagenode Bioruptor (Diagenode, B01020001) at high power for 40 cycles, 30 s ON/OFF. Sonicated samples were cleared with 1:10 dilution of 30% Triton X-100 and added to antibody bound beads (human anti-SOX2: R&D Systems, AF2018-SP; rabbit anti-H3K27ac: Abcam, ab4729). The next day, beads were washed with high- and low-salt buffers and treated with proteinase K overnight at 65°C. DNA was isolated using phenol/chloroform extraction, and 10 ng of immunoprecipitated DNA was used for cDNA library construction using the Rubicon DNA-seq ThruPLEX 48S Kit. Libraries were size-selected and submitted to be sequenced at ~50–base pair read length, single-end reads, with 15 million to 20 million reads per library. All of the -omics data can be accessed at GSE134277.

Histology, immunofluorescence, and immunohistochemistry

After dissection of either embryos or adult stomachs, tissue was immediately fixed in fresh 4% PFA in 1× PBS overnight. Whole-mount pictures were taken with a Leica microscope and ImageJ software (National Institutes of Health). After dehydration, tissues were processed, embedded, and cut into 5- to 7-μm sections. After baking overnight at 37°C, these slides were deparaffinized and rehydrated for staining. For simple dyes, Harris’s hematoxylin was used with alcoholic eosin Y (Electron Medical Sciences), Alcian blue with nuclear fast red, periodic acid–Schiff (Sigma-Aldrich, 395B-1KT) with Harris’s hematoxylin, and alkaline phosphatase (Roche, 11093274910) staining with nuclear fast red.

For protein staining, slides were antigen-retrieved using a citrate buffer (pH 6) for 30 min, blocked for 1 hour, and stained with the following antibodies: rabbit anti-Sox2 (Abcam, ab97959; 1:200), rabbit anti-Sox9 (Millipore, AB5535; 1:300), mouse anti-proton pump (H^+ K^+ ATPase beta subunit) (Cedarlane, D032-3; 1:500), mouse anti-TP63 (Santa Cruz Biotechnology, sc-8431; 1:200), mouse anti-PCNA (Santa Cruz Biotechnology, sc-56; 1:200), mouse anti-PDX1 (Developmental Studies Hybridoma Bank, F109-D12; 1:200), rabbit anti-CDX2 (R&D System, AF3665; 1:50), and chicken anti-GFP (Abcam, ab13970; 1:2000). For immunofluorescence, secondary staining with Alexa Fluor 488– and Alexa Fluor 568–conjugated (1:500; Thermo Fisher Scientific) antibodies was followed by mounting with 4′,6-diamidino-2-phenylindole–containing VECTASHIELD media (Vector Laboratories). For immunohistochemistry, we used the 3,3′-diaminobenzidine (Abcam) protocol followed by hematoxylin staining. Images were acquired using a Nikon microscope DS-Qi1Mc camera.

Francis et al., Sci. Adv. 2019; 5 : eaax8898  11 December 2019
Histopathology scoring  
H&E-stained sections of gastric samples were blindly scored on the basis of seven cancer traits, on a scale from 0 to 4 per trait; the scoring system was adapted from a previously adapted paper, where the trait “Hyalinosis Red refractile droplets and crystals” was disregarded (40). Cumulative scores were calculated, and significance was determined using an unpaired Student’s t test, comparing two groups at a time.

ATAC-seq and ChIP-seq data analysis  
After removing adaptors by Cutadapt, pair-end ATAC-seq reads were aligned to the mouse reference genome (GRCm38) using Bowtie2 with the parameter “-X 2000” (41). For ChIP-seq, the filtered single-end reads were aligned to the mouse reference genome (GRCm38) using Bowtie2 with default parameters. ATAC-seq and ChIP-seq peaks were then called by MACS2 -callpeak with parameters “-keep-dup 1” and “--SPMR” (42). ENCODE mm10 blacklist regions (www.encodeproject.org/annotations/ENCSR636HFF/) were excluded from the called peaks for downstream analysis. Resultant bedGraph files were converted to big wiggle files using the University of California, Santa Cruz (UCSC) bedGraphToBigWig tool (43). Genomic signal within 2 kb of peaks were visualized using deepTools (44). The Binding and Expression Target Analysis (BETA) tool from Cistrome was used to assess the activating or repressive function of selected peaks (45). TF motifs enriched in the selected peaks were identified using the “SeqPos” tool from Cistrome (46) and the R package “PWMEnrich” (www.biocductor.org/packages/2.13/bioc/html/PWMEnrich.html).

To determine the similarities between different tissues by ATAC-seq data, all the ATAC-seq peaks from different tissues were merged and the ATAC-seq signal of merged peaks were retrieved from .bw files. Pearson correlations were then calculated in a pairwise manner using the ATAC-seq signal.

RNA-seq analysis  
The trimmed pair-end reads were aligned to the mouse reference genome (GRCm38) with STAR (version 2.4.2a) (47), and gene expression was then quantified using the reads per kilobase per million mapped reads method. COAD (TCGA-COAD) and STAD (TCGA-STAD) RNA-seq data were downloaded from the Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/) (27, 28). Signature genes of a specific feature were selected by the threshold of fold change [fragments per kilobase of transcript, per million mapped reads (FPKM) +1] > 2 and false discovery rate < 0.05. The “getLDS” function from the R package “biomaRt” was used to align gene symbols between mouse and human (48). Enrichment plots of gene signatures were generated using the R package “clusterProfiler” (49).

SUPPLEMENTARY MATERIALS  
Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/5/12/eaax8898/DC1

Fig. S1. Quality control of genomic datasets.
Fig. S2. Analyses of gut region--specific markers during embryonic development.
Fig. S3. Hindstomach specification maintained by Sox9+ Creescapers.
Fig. S4. Molecular characterization of Sox2CreD2DKO epithelium.
Fig. S5. Disease quantification of adenoma samples.
Fig. S4. Molecular characterization of DKO epithelium.

REFERENCES AND NOTES  
1. J. Ferlay, H.-R. Shin, F. Bray, D. Forman, C. Mathers, D. M. Parkin, Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int. J. Cancer 127, 2893–2917 (2010).
2. V. Giroux, A. K. Rustgi, Metaplasia: Tissue injury adaptation and a precursor to the dysplasia-cancer sequence. Nat. Rev. Cancer 17, 594–604 (2017).
3. X. Li, C. Zhang, T. Gong, X. Ji, L. Duan, Z. M. Liu, L. Song, C. Ding, J. Xu, B. Zhen, Y. Wang, J. Qin, A time-resolved multi-omic atlas of the developing mouse stomach. Nat. Commun. 9, 4910 (2018).
4. R. I. Sherwood, T.-Y. Chen, D. A. Melton, Transcriptional dynamics of endodermal organ formation. Dev. Dyn. 238, 29–42 (2009).
5. S. G. Willet, J. C. Mills, Stomach organ and cell lineage differentiation: From embryogenesis to adult homeostasis. Cell. Mol. Gastroenterol. Hepatol. 2, 546–559 (2016).
6. A. M. Zorn, J. M. Wells, Vertebrate endoderm development and organ formation. Annu. Rev. Cell Dev. Biol. 25, 221–251 (2009).

[continued...]
27. J. Gao, B. A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S. O. Sumer, Y. Sun, A. Jacobsen, R. Sinha, E. Larsson, E. Cerami, C. Sander, N. Schultz, Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal. 6, pl1 (2013).

28. E. Cerami, J. Gao, B. E. Gross, S. O. Sumer, B. A. Aksoy, A. Jacobsen, C. J. Byrne, M. L. Heuer, E. Larsson, Y. Antipin, B. Reva, A. P. Goldberg, C. Sander, N. Schultz, The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2, 401–404 (2012).

29. T.-H. Kim, R. A. Shivdasani, Notch signaling in stomach epithelial stem cell homeostasis. J. Exp. Med. 208, 677–688 (2011).

30. L. Raghoebar, E. R. M. Bakker, J. C. Mills, S. Swagemakers, M. B. V. Kempen, A. B. D. Munck, S. Driegen, D. Meijer, F. Grossveld, D. Tibboel, R. Smits, R. J. Rottier, SOX2 redirects the developmental fate of the intestinal epithelium toward a premature gastric phenotype. J. Mol. Cell Biol. 4, 377–385 (2012).

31. P. Dalerba, D. Sahoo, S. Paik, X. Guo, G. Yothers, N. Song, N. Wilcox-Fogel, E. Forgo, P. S. Rajendran, S. P. Miranda, S. Hisamori, J. Hutchison, T. Kalisky, D. Qian, N. Wolmark, G. A. Fisher, M. van de Rijn, M. F. Clarke, CDX2 as a prognostic biomarker in stage II and stage III colon cancer. N. Engl. J. Med. 374, 211–222 (2016).

32. B. D. Harfe, P. J. Scherz, S. Nissim, H. Tian, A. P. McMahon, C. J. Tabin, Evidence for an expansion-based temporal Shh gradient in specifying vertebrate digit identities. Cell 118, 517–528 (2004).

33. M. D. Muzumdar, B. Tasic, K. Miyamichi, L. Li, L. Luo, A global double-fluorescent Cre reporter mouse. Genesis 45, 593–605 (2007).

34. S. R. Hingorani, E. F. Petricoin III, A. Maitra, V. Rajapakse, C. King, M. A. Jacobetz, S. Ross, T. P. Conrads, T. D. Veenstra, B. A. Hitt, Y. Kawaguchi, D. Johann, L. A. Liotta, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).

35. M. P. Verzi, H. Sin, L.-L. Ho, X. S. Liu, R. A. Shivdasani, Essential and redundant functions for visualizing cell morphology and clonal relationships in the mouse. Mol. Cell. Biol. 31, 2026–2039 (2011).

36. L. C. Murtaugh, B. Z. Stanger, K. M. Kwan, D. A. Melton, Notch signaling controls multiple steps of pancreatic differentiation. Proc. Natl. Acad. Sci. U.S.A. 100, 14920–14925 (2003).

37. T. C. Badea, Y. Wang, J. Nathans, A noninvasive genetic/pharmacologic strategy for visualizing cell morphology and clonal relationships in the mouse. J. Neurosci. 23, 2314–2322 (2003).

38. A. J. Syder, S. M. Karam, J. C. Mills, J. E. Ippolito, H. R. Ansari, V. Farook, J. I. Gordon, A transgenic mouse model of metastatic carcinoma involving transdifferentiation of a gastric epithelial lineage progenitor to a neuroendocrine phenotype. Proc. Natl. Acad. Sci. U.S.A. 101, 4471–4476 (2004).

39. J. D. Buenrostro, B. Wu, H. Y. Chang, W. J. Greenleaf, ATAC-seq: A method for assaying chromatin accessibility genome-wide. Curr. Protoc. Mol. Biol. 109, 21.29.1–21.29.9 (2015).

40. A. B. Rogers, N. S. Taylor, M. T. Whary, E. D. Stefanich, T. C. Wang, J. G. Fox, Helicobacter pylori but not high salt induces gastric intraepithelial neoplasia in B6129 mice. Cancer Res. 65, 10709–10715 (2005).

41. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012).

42. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nussbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-seq (MACS). Genome Biol. 9, R137 (2008).

43. W. J. Kent, A. S. Zweig, G. Barber, A. S. Hinrichs, D. Karolchik, BigWig and BigBed: Enabling browsing of large distributed datasets. Bioinformatics 26, 2204–2207 (2010).

44. F. Ramirez, D. P. Ryan, B. Grünig, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dünard, T. Manke, deepTools2: A next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44, W160–W165 (2016).

45. S. Wang, H. Sun, J. Ma, C. Zang, C. Wang, J. Wang, Q. Tang, C. A. Meyer, Y. Zhang, X. S. Liu, Target analysis by integration of transcriptome and ChIP-seq data with BETA. Nat. Protoc. 8, 2502–2515 (2013).

46. S. Mei, Q. Qin, Q. Wu, H. Sun, R. Zheng, C. Zang, M. Zhu, J. Wu, X. Shi, L. Taing, T. Liu, M. Brown, C. A. Meyer, X. S. Liu, Cistrome Data Browser: A data portal for ChIP-seq and chromatin accessibility data in human and mouse. Nucleic Acids Res. 45, D658–D662 (2017).

47. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chiasson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).

48. S. Durinck, P. T. Spellman, E. Biney, W. Huber, Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. 4, 1184–1191 (2009).

49. G. Yu, L. G. Wang, Y. Han, Q. Y. He, clusterProfiler: An R package for comparing biological themes among gene clusters. OMICS 16, 284–287 (2012).

Acknowledgments: We thank the Rossant and Huang Labs for sharing primary and secondary antibodies. Funding: The work was supported by SickKids Foundation startup, Cancer Research Society, and the University of Toronto’s Medicine by Design Initiative/the Canada First Research Excellence Fund (T.H.K.) and the NSERC discovery grant (498706), the CIHR New Investigator Award, and the OMIR Early Researcher Award (H.H.H.). Author contributions: R.F. designed and performed experiments, analyzed data, and wrote the manuscript. H.G. performed experiments; analyzed ATAC-seq, ChIP-seq, and RNA-seq data; and drafted the bioinformatics analytical sections. C.S. performed histopathology scoring. M.A. helped with RNA-seq analysis. T.Y. assisted in RNA-seq experiments. P.B.D. provided Sox2 mouse lines. H.H.H. supervised genomic and epigenomic data analyses. T.-H.K. conceived and supervised the study, analyzed data, and wrote the manuscript. Competing interests: The authors declare that they have no competing interests. Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. All of the presented -omics data are available at the NCBI GEO (accession number: GSE134277). Additional data related to this paper may be requested from the authors.

Submitted 2 May 2019
Accepted 16 October 2019
Published 11 December 2019
10.1126/sciadv.aax8989

Citation: R. Francis, H. Guo, C. Streutker, M. Ahmed, T. Yung, P. B. Dirks, H. H. He, T.-H. Kim, Gastrointestinal transcription factors drive lineage-specific developmental programs in organ specification and cancer. Sci. Adv. 5, eaax8989 (2019).