HHS Public Access

Author manuscript

Cell Stem Cell. Author manuscript; available in PMC 2023 January 06.

Published in final edited form as:

Cell Stem Cell. 2022 January 06; 29(1): 101–115.e10. doi:10.1016/j.stem.2021.09.004.

SATB2 preserves colon stem cell identity and mediates ileum-colon conversion via enhancer remodeling

Wei Gu[1], **Hua Wang**[2], **Xiaofeng Huang**[1], **Judith Kraiczy**[3,4], **Pratik N. P. Singh**[3,4], **Charles Ng**[5], **Sezin Dagdeviren**[6], **Sean Houghton**[1], **Oscar Pellon-Cardenas**[7], **Ying Lan**[1], **Yaohui Nie**[1], **Jiaoyue Zhang**[1], **Kushal K Banerjee**[3,4], **Emily J. Onufer**[8], **Brad W. Warner**[8], **Jason Spence**[9], **Ellen Scherl**[5], **Shahin Rafii**[1], **Richard T. Lee**[6], **Michael P. Verzi**[7], **David Redmond**[1], **Randy Longman**[5], **Kristian Helin**[2,10,11], **Ramesh A. Shivdasani**[3,4], **Qiao Zhou**[1,#]

[1]Division of Regenerative Medicine & Ansary Stem Cell Institute, Department of Medicine, Weill Cornell Medicine, 1300 York Avenue, New York, NY, 10065, USA

[2]Cell Biology Program and Center for Epigenetics Research, Memorial Sloan Kettering Cancer Center, 430 E 67th Street, New York, NY, 10065, USA

[3]Department of Medical Oncology, Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA, 02215, USA

[4]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, 75 Francis Street, Boston, MA, 02115, USA

[5]Jill Roberts Center for Inflammatory Bowel Disease, Weill Cornell Medicine, 1283 York Avenue, New York, NY, 10065, USA

[6]Department of Stem Cell and Regenerative Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA, 02138, USA

[7]Department of Genetics, Rutgers University, 145 Bevier Road, Piscataway, NJ, 08854, USA

[8]Division of Pediatric Surgery, Department of Surgery, Washington University School of Medicine, 660 S Euclid Avenue, St. Louis, MO, 63110, USA

[9]Department of Internal Medicine, University of Michigan, 1500 E Medical Center Drive, Ann Arbor, MI, 48109, USA

[#]Lead Contact (jqz4001@med.cornell.edu).

Author contributions

WG and QZ designed all experiments, interpreted data, and wrote the manuscript. WG performed all experiments. HW performed ChIP and enhancer analysis, with help from PNPS and SH. XH and DR performed scRNA-seq analysis. XH and WG performed Bulk RNA-seq analysis. JK and RAS provided bulk RNA-seq raw data of Eed deletion and early time course experiments. SD participated in the functional nutrient absorption studies, with contributions from EO, BWW, OPC and MPV. CN, ES, and RTL provided human samples and advised data interpretation. YL, YN and JZ contributed to independent validation of data. RAS, KH, and SR contributed to experimental design, discussed experiments, and edited the manuscript.

Declaration of interests

A patent application related to this work is pending at Weill Cornell Medicine.

[10]Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Copenhagen N 2200 Denmark

[11]The Novo Nordisk Foundation for Stem Cell Biology (Danstem), University of Copenhagen, Copenhagen N 2200, Denmark

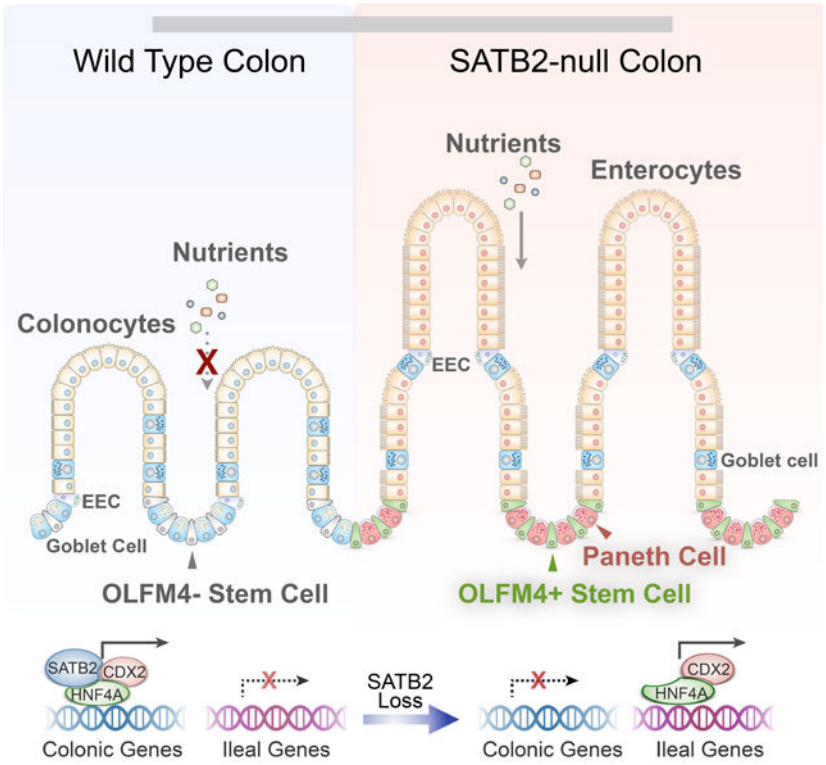
Summary

Adult stem cells maintain regenerative tissue structure and function by producing tissue-specific progeny, but factors that preserve their tissue identities are not well understood. The small and large intestines differ markedly in cell composition and function, reflecting their distinct stem cell populations. Here we show that SATB2, a colon-restricted chromatin factor, singularly preserves $LGR5^+$ adult colonic stem cell and epithelial identity in mice and humans. *Satb2* loss in adult mice leads to stable conversion of colonic stem cells into small intestine ileal-like stem cells and replacement of colonic mucosa with one that resembles the ileum. Conversely, SATB2 confers colonic properties on the mouse ileum. Human colonic organoids also adopt ileal characteristics upon SATB2 loss. SATB2 regulates colonic identity in part by modulating enhancer binding of intestinal transcription factors CDX2 and HNF4A. Our study uncovers a conserved core regulator of colonic stem cells able to mediate cross-tissue plasticity in mature intestines.

eTOC:

Zhou and colleagues showed that the adult colonic mucosa has a surprising degree of plasticity, regulated by the chromatin factor SATB2. Loss of SATB2 in mice or in human colonic organoids leads to transformation of colonic mucosa to small intestine ileum, with marked morphological, functional, and epigenetic remodeling.

Graphical Abstract



Introduction

Adult stem cells are found in many regenerative tissues, each producing a discrete set of progenies suitable for local tissue repair and replenishment (Clevers and Watt, 2018; Wells and Watt, 2018). Upon tissue injury, adult stem cells can display substantial phenotypic plasticity by activating alternative gene expression programs or altering their repertoire of progeny (Blanpain and Fuchs, 2014; Nusse et al., 2018; Page et al., 2013; Wang et al., 2019). Moreover, differentiated cells or committed progenitors may dedifferentiate to replenish the stem cell pool (Donati et al., 2017; Leushacke et al., 2017; Stange et al., 2013; Tata et al., 2013; van Es et al., 2012). Notably, these plastic behaviors are confined within the original stem cell differentiation hierarchy and do not violate the bounds of specific tissue fates. Thus, appropriate tissue-specific cell types are produced, and the adult stem cells return to their homeostatic states when conditions normalize. These observations suggest that fundamental mechanisms safeguard the core tissue identity of each adult stem cell population.

Overt conversion of one stem cell population to another will lead to replacement of the original multi-lineage tissue. Although formation of stable ectopic tissues in adults is observed in some human organs and in animal studies, molecular underpinnings for these unusual events are often unclear and may include mechanisms that do not involve switches in stem cell fate (Giroux and Rustgi, 2017; Slack, 2007). In the digestive tract, expression of the intestine regulator CDX2 in mature murine stomach leads to intestinal metaplasia in

mice (Mutoh et al., 2002). It is uncertain whether a gastric-to-intestinal stem cell conversion occurred or whether the induced intestine-like stem cells in stomach resemble their native counterparts or represent a novel cell type. There is limited understanding of how adult stem cells maintain tissue identities and, importantly, whether specific stem cell regulators can both sustain identity and convert stem cell fate in adult organs.

Small and large intestine mucosae are highly regenerative, with 4- to 7-day cycles of self-renewal powered by separate populations of $LGR5^+$ intestinal stem cells (ISCs) in the crypts of Lieberkühn (Gehart and Clevers, 2019). The progeny of small and large intestine stem cells are distinctly different (Barker et al., 2007). Colonic epithelium lacks microbicidal Paneth cells and carries abundant mucus-secreting goblet cells and electrolyte-absorbing colonocytes. In contrast, the proximal (duodenum and jejunum) and the distal small intestines (ileum) have villus structures lined with nutrient-absorbing enterocytes and fewer goblet cells, and Paneth cells intercalated with ISCs in the crypts. The Ileum has the additional specialized function of absorbing bile acids and vitamins.

Committed intestinal cells have a pronounced ability to dedifferentiate into $LGR5^+$ stem cells upon ISC loss (Murata et al., 2020; Tetteh et al., 2015). After parasitic infection or injury, $SCA1^+$ or $HOPX^+$ progenitors may temporarily replace ISCs (Nusse et al., 2018; Wang et al., 2019; Yui et al., 2018). Despite this plasticity, the fundamental demarcation between the small and large intestine remains unaltered. Moreover, organoids from the small or large intestine maintain their distinct molecular features and differentiation pattern in culture or upon transplantation (Sato et al., 2011; Sato et al., 2009; Sugimoto et al., 2018), revealing intrinsic ISC stability. Studies of numerous transcription factors and signaling pathways, including HNF4A/G, GATA4/6, WNT, and BMPs, illuminate their critical roles in intestinal homeostasis and regeneration (Beumer and Clevers, 2021; Santos et al., 2018; Thompson et al., 2018; Wells and Spence, 2014), but no known factor or pathway changes small versus large intestine stem cell fate or tissue identity. Here we identify the colon-restricted chromatin factor SATB2 as a crucial determinant of mouse and human colonic stem cell fate. SATB2 loss transforms colonic epithelium into ileum-like tissue in mice and in human colonic organoids. Single-cell RNA-sequencing demonstrated a genuine conversion of colonic to ileum-like stem cells and their differentiation into ileal cell types in the mutant colon. SATB2 also directly safeguards the identity of mature colonocytes; loss of SATB2 rapidly transforms colonocytes into enterocytes. Mechanistically, SATB2 acts to maintain intestinal transcription factors (TFs) CDX2 and HNF4A at colon-specific enhancers; in its absence, CDX2 and HNF4A relocate to ileal enhancers. Together, our studies identify a chromatin factor crucial for regulating colonic stem cell fate and uncover a surprising and conserved cross-tissue plasticity in the mature intestine.

Results

SATB2 is enriched and required in colonic epithelium

To identify genes involved in maintaining colonic identity in adult mice and humans, we interrogated our published RNA-seq data of purified murine $LGR5^+$ ISCs from the duodenum and colon for colon-enriched TFs (Fig. S1A) (Jadhav et al., 2016; Murata et al., 2020). We further cultured duodenal and colonic organoids from human biopsy samples

under high WNT conditions (WENR medium), which favor ISC growth (VanDussen et al., 2019), and used RNA-seq to identify TFs enriched in human colonic organoids (Fig. S1B). Besides posterior *Hox* genes, two TFs, *SATB2* and *FOXD2*, were enriched in both murine and human colon (Fig. S1C). To assess these TFs' requirements in regulating colonic identity, we used CRISPR (Clustered regularly interspaced short palindromic repeats), CAS9, and 3 different guide RNAs to disrupt *Satb2* or *Foxd2* in murine colonic organoids, achieving deletion efficiencies of 55% to 95% in independent experiments (Fig. S1D, E). Disrupting *Foxd2* had little impact on the colonic transcriptome, whereas *Satb2* loss altered the mRNA profile significantly, reducing colonic genes and increasing small intestine genes (Fig. S1F–I), indicating a requirement for *Satb2* in maintaining adult colonic identity.

Replacement of colonic mucosa by ileal-like mucosa in adult mice after SATB2 loss

SATB2 is a homeodomain-containing chromatin factor expressed in developing craniofacial tissues and cortical neurons (Alcamo et al., 2008; Britanova et al., 2008; Britanova et al., 2006; Dobreva et al., 2006). Human *SATB2* mutations cause craniofacial anomalies and cognitive impairment (Zarate and Fish, 2017). *SATB2* is also expressed in fetal and adult murine and human hindgut and may be used as a diagnostic marker for colorectal cancer (Munera and Wells, 2017; Perez Montiel et al., 2015), but its intestinal functions are largely unknown. Immunoblots (Fig. S2A) and immunohistochemistry (Fig. 1A) revealed prominent SATB2 expression in adult mouse cecal and colonic epithelia, including in $LGR5^{+}$ ISCs at the crypt base. Ileal villus cells showed weak nuclear SATB2 staining but the small intestine is otherwise devoid of SATB2 (Fig. 1A and Fig. S2A). Immunoblotting confirmed lack of SATB2 expression in ileal crypts (Fig. S2B). QRT-PCR of $LGR5^{+}$ ISCs isolated by FACS from different gut segments of $Lgr5^{CreERGFP}$ reporter mice showed high-level *Satb2* mRNA in colonic but not in any of small intestine ISCs (Fig. S2C).

To evaluate intestinal *Satb2* function in vivo, we deleted *Satb2* in intestinal mucosa in 2-month old $Satb2^{f/f}$ mice using the $Villin\text{-}Cre^{ER(T2)}$ strain (Fig. S2D), leading to near complete absence of SATB2 (Fig. S2E). One month after Tamoxifen (TAM) treatment, the colonic mucosa of $Vil\text{-}Cre^{ER};Satb2^{f/f}$ mice (referred thereafter as $Satb2^{cKO}$) was significantly remodeled, with the characteristic flat epithelium replaced by villus structures (mucosal depth 208 ± 24 μm vs. 92 ± 18 μm in $Satb2^{f/f}$ controls) and presence of Paneth cells at the crypt base (Fig. 1B, C, and Fig. S2E, F), resembling the small intestine. Goblet cells stained with Alcian blue were significantly decreased (9% ± 2% of all epithelial cells), to levels comparable to ileum (8% ± 1%) rather than colon (15% ± 1.5%; Fig. S2G). Apoptosis rates were similar in mutant and control colon (Fig. S2H). 5-ethynyl-2-deoxyuridine (EdU) pulse-chase revealed accelerated movement of epithelial cells away from the crypt base (mutant vs control colon, $P < 0.0001$, Tukey's multiple comparison test, Fig. S2I), similar to normal ileum (mutant colon vs control ileum, $P > 0.99$, Tukey's multiple comparison test), which may account for the architectural remodeling in $Satb2^{cKO}$ colon.

Whole epithelium RNA-seq revealed little difference between *Satb2*-null and control jejunum or ileum, whereas the mutant cecal and colonic transcriptomes resembled that of normal ileum (Fig. 1D). Of the 362 ileal enriched genes (control ileal vs colonic

transcriptome, Log_2 fold change (LFC) >2, adjusted P value (Padj) < 0.05), 309 (85.4%) were up-regulated in Satb2^{cKO} colon whereas 238 of 302 colon-enriched genes (78.8%) were down-regulated (Fig. 1E, and Fig. S2J). Accordingly, molecular pathways that control digestion, absorption, and solute transport, reflecting ileal functions, were activated at the expense of colonic functions such as fatty acid and xenobiotic metabolism (Fig. S2K–M). Immunohistochemistry revealed loss of colonic markers such as CA1 and AQP4 and gain of ileal markers such as OLFM4 (stem cells), FABP6 and FGF15 (enterocytes), and the Paneth cell product LYZ1 (Fig. 1F). Tissue remodeling in Satb2^{cKO} colon was accompanied by elevated immune cell presence. Six months after TAM treatment, *Satb2*-null colon was still wholly lined by an ileum-like mucosae with widespread expression of ileal genes in both proximal and distal colon, with the proximal colon displaying more prominent villi (Fig. S3A–C). These data indicate stable colonic to ileal conversion after SATB2 loss.

Conversion of LGR5$^+$ colonic stem cells to ileal-like stem cells

Given the stable colonic remodeling after SATB2 loss, we reasoned that colonic ISCs may have converted into ileum-like ISCs. To evaluate this hypothesis, we used three different approaches: single-cell transcriptome profiling of LGR5$^+$ stem cells, organoid cultures, and *Satb2* deletion from LGR5$^+$ ISCs.

First, we profiled epithelial cells (FACS-purified EPCAM$^+$ CD45$^-$CD31$^-$ cells, 30 days post-TAM) using single-cell RNA sequencing (scRNA-seq). Transcriptomes from 3,912 control ileal, 3,627 control colonic, and 4,370 Satb2^{cKO} colonic cells were integrated and partitioned into 7 populations including goblet, enterocyte, colonocyte, Paneth, tuft, and enteroendocrine (EE) cells, annotated with lineage-specific markers (Haber et al., 2017) (Fig. 2A and Fig. S4A; Table S1). Cells bearing low lineage markers but high proliferating genes Mki67 or Mcm3/6 were classified as progenitors, which include LGR5$^+$ ISCs as a subset (Fig. 2A and Fig. S4A, B). A majority of differentiated cells from the Satb2^{cKO} colon clustered with ileal cells (Fig. 2A) and expressed canonical ileal markers (Fig. 2B). We further assessed the similarity between control ileal and Satb2^{cKO} colonic transcriptomes using cohorts of genes enriched in each ileal cell type (ileal identity scores, Table S2), which similarly showed broad adoption of ileal identity by Satb2^{cKO} colonic cells (Fig. S4C). For instance, colonocytes, representing 22.4% of the control colon, were replaced by enterocytes in Satb2^{cKO} mice (21.5% of the total population) (Fig.S4B, C; Table S3).

Lgr5$^+$ stem cells within the “progenitor” groups (Fig. S4D) expressed high levels of ISC markers *Ascl2* and *Axin2* and scored significantly higher than *Lgr5*$^-$ progenitors on a stem-cell scorecard (Munoz et al., 2012) (Wilcoxon rank sum test continuity correction p-value < $2.2e^{-16}$) (Fig. S4D, E). Focusing on ISC subsets at the G1/S cell cycle phase (control ileum, 209 cells; control colon, 230 cells; mutant colon, 155 cells), which are proposed as basal stem cells (Biton et al., 2018), Satb2^{cKO} colonic cells clustered with ileal, not with colonic ISCs (Fig. 2C and Fig.S4F). Compared with control colon, the ileum-like ISCs in Satb2^{cKO} colon were enriched for GeneOntology pathways of antimicrobial and innate immune responses while depleted of sulfur and phospholipid metabolism pathways (Fig. S4G).

Next, we evaluated stem cell properties in organoid cultures. Large and small intestines ISCs differ in their ability to form organoids in 3D Matrigel cultures. Colonic crypts fail to generate organoids in standard ENR small intestine medium lacking WNT3A (Sato et al., 2011). Crypts isolated from control ileum, control colon and Satb2^{cKO} colon all produced spheroids in WENR medium containing high WNT3A (Fig. S5A). However, in ENR medium, control colonic crypts yielded only few non-branching spheroids (0.015 ± 0.013 structures per crypt) and most of these could not be passaged whereas both control ileal (0.25 ± 0.06 primary and 1.4 ± 0.6 secondary structures per crypt) and *Satb2*-null colonic crypts (0.19 ± 0.03 primary and 1.8 ± 0.5 secondary structures per crypt) formed branching organoids that could be propagated (Fig. 2D and Fig. S5A, B).

Lastly, we deleted *Satb2* directly from LGR5^{+} ISCs in *Lgr5$^{GFP\text{-}Cre(ER)}$;Satb2$^{f/f}$* mice. *Lgr5$^{GFP\text{-}Cre(ER)}$* expression is mosaic and restricted to the ISC compartment. TAM injection into *Lgr5$^{GFP\text{-}Cre(ER)}$;Satb2$^{f/f}$* mice accordingly yielded mosaic *Satb2*-null colonic crypts carrying GFP^{+} ISCs (Fig. 2E). One week after treatment, SATB2 disappeared from the lower parts of GFP^{+} crypts, where new cells reside, but persisted in higher cell tiers, which house older cells with intact *Satb2*. Activation of ileal markers OLFM4 and FABP6 and suppression of colonic marker CA1 were partial in GFP^{+} glands and LYZ1^{+} cells were absent (Fig. 2E), suggesting incomplete epithelial remodeling at this early time. Importantly, all FABP6^{+} cell lost SATB2 whereas FABP6 and CA1 mark distinct cells, indicating cell-autonomous regulation of these marker genes by SATB2 (Fig. S4H). ISC and epithelial remodeling were complete by 36 days, with OLFM4 present in most GFP^{+} cells, LYZ1^{+} cells present in GFP^{+} glands, and replacement of CA1^{+} colonocytes by FABP6^{+} enterocytes (Fig. 2E). These observations indicate a time-dependent conversion of colonic stem cells and resetting of the differentiation pattern. In aggregate, findings from single-cell profiling, organoid culture, and ISC-specific gene deletion are consistent with a fundamental conversion of colonic into ileum-like stem cells in the absence of SATB2.

SATB2 safeguards colonocyte identity

Given that SATB2 is expressed in stem and differentiated colonic cells, we evaluated whether SATB2 might directly regulate differentiated cell identity. We administered a single dose of TAM to Satb2^{cKO} mice and examined colonic gene expression 1, 2, 4, or 6 days later (Fig. 3). SATB2 signals disappeared by Day 2, with concomitant FABP6 activation and CA1 suppression in a subset of colonic cells located in the upper glands (Fig. 3A). At this early time, epithelial self-renewal has not reached the upper glands, so these cells were pre-existing mature colonocytes. By day 4, most glandular cells had activated FABP6 and lost CA1 (Fig. 3A). In contrast, OLFM4 and LYZ expression in the crypts only became prominent in Day 30 samples (Fig. 3A). These data suggest a rapid identity switch from colonocytes to enterocytes after SATB2 loss, independent of stem cell conversion.

Consistent with the immunohistochemistry data, RNA-seq showed activation of hundreds of genes involved in nutrient absorption and transport by day 4 (Fig. 3B; clusters 2 & 3, Fig. 3C; clusters 6, Fig. S4I; Table S4) and down-regulation of colon-specific pathways in immune modulation and glycosylation (clusters 4 & 5, Fig. 3C, Table S4). Significant activation of *Defensin* genes and *Olfm4* was only achieved at Day 30 (cluster 1, Fig. 3C,

Table S4). Notably, gene set enrichment analysis revealed no enrichment of fetal signature genes (Fordham et al., 2013) in Satb2^{cKO} colonic transcriptomes between day 1 and 6 (Fig. S4J). Examination of two fetal markers showed modest and transient increase in *Ly6a* (*Sca1*) but not *Anxa1* (Fig. S4K). Significant fetal gene activation was thus not associated with colonic to ileal transformation. Our data indicate that SATB2 safeguards the identity of mature colonocytes in addition to its critical role in maintaining stem cell identity.

Environmental factors influence colonic to ileal conversion

A minority of Satb2^{cKO} colonic cells retained colonic identity, including 9.2% of mature absorptive cells and 3.8% of goblet cells (Fig. S4C). We postulated that the colonic milieu may influence differentiation of the ileum-like mucosa in Satb2^{cKO} colon. Indeed, many studies illustrate the importance of microbial and niche signals in regulating intestinal gene expression and TF activity (Chen et al., 2019; Davison et al., 2017; Nichols and Davenport, 2021; Thaiss et al., 2016). For instance, the microbiota is necessary and sufficient to induce expression of major histocompatibility complex class II (MHCII) genes in small intestine (but not colon) ISCs (Biton et al., 2018; Umesaki et al., 1995). Consistently, MHCII genes were high in ileal and low in control colonic and ileum-like Satb2^{cKO} colonic ISCs (Fig. S5D). To mitigate environmental influences, we cultured ileal and Satb2^{cKO} colonic organoids in identical WENR medium for one passage, differentiated the organoids and performed RNA-Seq. Principal Component Analysis (PCA) and Pearson correlation showed that the transcriptomes of ileal and Satb2^{cKO} colonic organoids resembled each other (Pearson $r = 0.983$) more closely than the two samples harvested in vivo ($r = 0.954$; Fig. S5F, G). Nevertheless, significant differences remained. These data suggest that environmental factors contributed to, but were not the main cause for, the incomplete conversion of a subset of colonic cells to ileal identity in Satb2-null colon.

Generation of bona fide nutrient-absorbing enterocytes in the ileum-like colon

Ileal enterocytes absorb nutrients as well as bile salts and vitamins. Both ileal and Satb2^{cKO} colonic enterocytes expressed many transporters for lipids, carbohydrates, amino acids, bile salts and vitamins that were absent or low in colonocytes (Fig. 4A and Fig. S5H, I). Satb2^{cKO} colonic enterocytes were enriched for functional pathways in nutrient absorption and digestion, and notably, in genes relating to microvillus organization (Fig. 4A). Enterocytes are well known to sprout longer microvilli than colonocytes to increase their absorptive surface. Electron microscopy revealed substantially longer microvilli in *Satb2^{cKO}* colonic enterocytes than in control colonocytes, comparable to those of ileal enterocytes (Fig. 4B).

To evaluate whether the ileum-like mucosa in Satb2^{cKO} colon can more readily absorb nutrients and bile salts, we employed an in vivo absorption assay by tying both ends of a segment of the ileum or colon to create a pouch, followed by injection of [^{3}H] glucose and [^{14}C] taurocholic acid into this pouch enabling detection of trans-epithelial transport of radiolabeled materials into the portal circulation and its incorporation in the liver tissue (Fig. 4C). Both portal plasma and liver parenchyma from Satb2^{cKO} mice showed significantly higher radiotracer levels compared to controls (Fig. 4D). These findings together indicate generation of bona fide enterocytes in Satb2^{cKO} colon.

SATB2 confers colonic characteristics on the mature ileum

To evaluate whether SATB2 can confer colonic fate to the small intestine mucosa, we generated a transgenic mouse line, *CAG*$^{SATB2\text{-}GFP}$, in which CRE excision of a stop cassette activates HA epitope-tagged SATB2 and GFP(Fig. 5A). TAM treatment of 2-month old *Vil-Cre*^{ER};*CAG*$^{SATB2\text{-}GFP}$ mice (referred to as Satb2^{OE}) led to mosaic expression of the HA-tagged SATB2and GFP throughout the intestine, relatively low in ileum (approximately 10–15% of the glands – Fig. 5B), and higher in jejunum and duodenum (>50% of the glands – Fig. S5J). RNA-seq of FACS-purified GFP^{+} cells from the ileum and jejunum showed comparable *Satb2* mRNA levels to the colon (Fig. 5C, D, Fig. S5K). Compared with GFP^{-} ileal cells, 225 genes were down-regulated and 131 genes were up-regulated in ileal GFP^{+} cells (LFC >1.5, Padj < 0.1); these genes were enriched for colonic and ileal tissue signatures, respectively (Fig. 5E). Among the down-regulated genes were large numbers of enterocyte nutrient transporters and *Defensins* characteristic of Paneth cells (Fig. 5F).

Colonic epithelium absorbs electrocytes and synthesizes many glycoproteins including specific mucins for anti-microbial defense. GFP^{+} ileal cells expressed an array of key electrolyte transporters and principal glycosylation enzymes (Fig. 5F). Thus, they acquired molecular machineries necessary for colonic functions. In ileal villi marked with GFP, immunohistochemistry 30 days after TAM showed suppression of ileal marker FABP6 and activation of colonic marker CA1 (Fig. 5G, H). OLFM4 and LYZ1 also disappeared from GFP^{+} crypts, consistent with the transcriptomic data. In contrast to Satb2^{OE} ileum, qRT-PCR analysis indicated that jejunal and duodenal GFP^{+} cells down-regulated small intestine genes but failed to activate most colonic genes, with duodenum being the least responsive (Fig. S5K–M). Taken together, SATB2 is sufficient to confer colon-like characteristics on the adult ileum.

SATB2 regulates enhancer dynamics and transcription factor binding in colon

To investigate how SATB2 might control colonic fate and tissue plasticity, we mapped genome-wide SATB2 binding using chromatin immunoprecipitation-sequencing (ChIP-seq). Duplicate SATB2 ChIP data from control colonic epithelia yielded highly concordant data with 25,576 high-quality peaks ($P < 1 \times 10^{-9}$, using both input DNA and SATB2^{cKO} ChIP as controls) (Fig. S6A). These peaks were enriched for AT-rich sequences, consistent with SATB2 binding preference (Szemes et al., 2006) (Fig. 6A). Among the top enriched DNA-binding motifs identified by HOMER were those for the intestinal TFs CDX2 and HNF4A, suggesting co-localization of the two TFs with SATB2 (Fig. 6A, Table S5). Indeed, ChIP-seq for CDX2 and HNF4A revealed extensive co-localization, with 54.1% (13,843 out of 25,576) of SATB2 peaks co-bound by both TFs (Fig. 6B and Fig. S6C, D). Moreover, CDX2 and HNF4A antibodies co-precipitated SATB2 from colonic tissue (Fig. 6C), suggesting physical interactions of SATB2 with CDX2/HNF4A.

Colonic SATB2 binding occurred predominantly in intergenic regions and introns (39.1% and 53.2% of peaks, respectively; Fig. S6B), and enriched for the motif of P300, the histone H3K27 acetyltransferase and a hallmark of active enhancers (Fig. 6A; $P < 1 \times 10^{-443}$). We used Cleavage Under Targets & Release Using Nuclease (CUT&RUN) to map putative (H3K4me1) and active (H3K27ac) enhancers in control ileum, colon,

and SATB2^{cKO} colon epithelia. Peaks were called with MACS2 using duplicate samples. Tissue-specific enhancers (H3K4me1, TSS-distal regions) were identified by MAnorm. Assay for Transposase-Accessible Chromatin (ATAC-seq) was employed to further chart the chromatin landscapes in these tissues. Analysis of all H3K4me1^{+} enhancers defined 7,375 colon-specific and 5,784 ileum-specific sites (MAnorm; $P < 0.01$), with the nearby genes (<50 kb) enriched for colonic or ileal expression, respectively (Fig. 6D). In normal colon, a majority of SATB2 binding (59.5%) occurred within H3K27ac^{+} active enhancers (Fig. 6E and Fig. S6E, F), and 67% of SATB2/CDX2/HNF4A co-bound sites overlap with active enhancers.

In control colon, the colon-specific enhancers had high levels of H3K4me1 and H3K27ac, strong ATAC signals, and robust binding by CDX2 and HNF4A, all hallmarks of active enhancers (Fig. 6F, G and Fig. S6G, H). These enhancers were de-activated in Satb2^{cKO} colon and displayed low levels of H3K4me1, H3K27ac, open chromatin, and CDX2 or HNF4A occupancy, indicating a critical role for SATB2 in maintaining active colonic enhancers. Notably, ileum-specific enhancers in normal colon retained low but detectable signals of ATAC, H3K4me1, H3K27ac, CDX2 and HNF4A; after SATB2 loss each of these signals was significantly enhanced (Fig. 6F, G and Fig. S6G, H). Thus, Ileal enhancers are not constitutively silent at baseline in mature colon but retain weak enhancer features. These "primed" enhancers likely provide the necessary chromatin substrate for ileal gene activation and tissue fate plasticity.

Many developmental enhancers active in embryos and decommissioned in adult intestines retain low H3K4me1 and get reactivated after prolonged loss of polycomb repressive complex 2 (PRC2) (Jadhav et al., 2019). We asked whether the "primed" ileal enhancers in adult colon (and vice versa) were erstwhile active developmental enhancers. Analysis of published midgut ATAC profiles (Banerjee et al., 2018) and new hindgut ATAC profiles from developing (embryonic days 12, 14, 16) and newborn mice revealed no evidence that ileal enhancers had been active in developing hindgut or colonic enhancers active in developing midgut (Fig. S6J). Moreover, genetic deletion of *Eed* in adult intestine (*Villin^{CreER};Eed$^{f/f}$*), which inactivated PRC2, did not result in colonic to ileal conversion (Fig. S6K, L). Combined removal of *Eed* and *Satb2* (*Villin^{CreER};Eed$^{f/f}$;Satb2$^{f/f}$*) also did not enhance the transcriptomic shift toward ileum (Fig. S6K, L). Thus, the primed ileal enhancers in adult colon are not decommissioned fetal enhancers and PRC2 is not overtly involved in SATB2-dependent colonic identity maintenance.

In adult intestine, CDX2 and HNF4A primarily function as transcriptional activators (Verzi et al., 2011; Verzi et al., 2013). After *Satb2* loss, CDX2 levels decreased approximately one-fold while HNF4A increasing by one-fold (Fig. S6N, O). Nevertheless, the two TFs associate with each other in both normal and *Satb2^{cKO}* colon (Fig. 6C and Fig. S6M), and their co-binding switched from colonic to ileal enhancers after SATB2 loss (Fig. 6F, G and Fig. S6G–I), closely correlating with down-regulation of colonic and activation of ileal genes (Fig. 6D). These data indicate that SATB2 regulates colonic gene expression and tissue plasticity in part by modulating enhancer interactions with crucial intestinal TFs.

Lastly, we evaluated SATB2 binding in LGR5^{+} colonic ISCs versus differentiated cells, isolated respectively by FACS from LGR5^{DTRGFP} reporter mice (Fig. 6H). SATB2 CUT&RUN with approximately 2×10^5 cells each yielded 7,523 peaks in LGR5^{+} cells and 15,757 peaks in LGR5^{-} cells. Although the average signal strength of SATB2 binding events in ISCs was lower than in differentiated cells, their binding patterns on the whole genome level were highly concordant (Fig. 6I; Pearson r=0.94) and aligned well with the SATB2 binding peaks identified by ChIP-seq (Fig. S6P). Thus, SATB2 bound the same genomic sites in stem as in differentiated cells, suggesting that SATB2 “primes” the stem cells for a differentiation path to colonic progenies.

Human colonic organoids adopt ileal characteristics after SATB2 loss

SATB2 expression is restricted to colonic mucosa in adult human intestine (Fig. 7A). To evaluate SATB2 function in human, we used CRISPR-CAS9 to delete *SATB2* from 5 normal human colonic organoid lines, which expressed SATB2 at comparable levels (Fig. S7A, B). Of the four guide RNAs (gRNAs) assessed, one efficiently reduced SATB2 expression by 95–98% (Fig. 7B and Fig. S7C, D). RNA-seq analysis of the 5 isogenic control (CAS9 alone) and SATB2 knockout (SATB2^{hKO}) organoid lines showed significant suppression of colonic genes and activation of small intestinal genes (Fig. 7C, D, and Fig. S7E). The top activated KEGG pathways included nutrient and vitamin absorption and retinol metabolism (Fig. 7E). Immunohistochemistry of SATB2^{hKO} colonic organoids confirmed expression of the ileal enterocyte markers FABP6 and RBP2 and the small intestine brush-border peptide transporter SLC15A1 (Fig. 7F and Fig. S7F, G). Digestive enzyme activities of the small intestine disaccharidase and dipeptidyl peptidase were also significantly elevated in SATB2^{hKO} colonic organoids (Fig. 7G and Fig. S7H). In contrast, human colon-enriched markers CEACAM1 and MUC2 were down-regulated (Fig. S7I, J). These data indicate that SATB2 has a conserved function in preserving human colonic epithelial identity and mediating colonic to ileal plasticity.

Discussion

Adult stem cells sustain structure and function of regenerative tissues. Phenotypic plasticity of adult stem cells observed after injury in many organs (Blanpain and Fuchs, 2014; Tetteh et al., 2015) generally occurs along existing differentiation hierarchy while the core tissue identities remain intact. In principle, adult tissue fate could be enforced by distributed actions of intrinsic and extrinsic factors, with perturbation of each producing only a limited effect. Although master regulatory factors specify tissue identity in embryos, their abilities are often lost in adults, owing partly to altered epigenetic landscapes across development (Banerjee et al., 2018; Spitz and Furlong, 2012; Stergachis et al., 2013; Zaret and Mango, 2016). Loss of CDX2, for example, has dramatic effects in embryos, including homeotic-like transformations to esophagus or stomach (Gao et al., 2009; Grainger et al., 2013), but CDX2 loss from the adult intestine spares tissue identity (Banerjee et al., 2018). In contrast, we show here that a tissue-restricted chromatin factor, SATB2, uniquely maintains mouse and human colonic stem cell and tissue identity. Similarly, important fate regulators might also operate in other stem cell populations in the adult body.

Studies in the thymus and other tissues suggest that SATB1 can engage nuclear matrix, bind DNA at base-unpaired regions, regulate genomic binding of chromatin remodeling complexes and signaling molecules, and influence chromatin looping (Cai et al., 2003; Skowronska-Krawczyk et al., 2014; Yasui et al., 2002). These multi-faceted functions have led to the proposal that it acts as a hub for various protein-protein and protein-chromatin interactions. Our studies indicate that SATB2 regulates colonic transcription and cell fate in part by modulating enhancer dynamics and targeting of the intestinal TFs such as CDX2 and HNF4A, consistent with the proposed properties of SATB1/2 proteins. SATB2 binds the same genomic sites in stem and differentiated cells and directly maintains colonic identity of both stem and non-stem cells. These observations suggest a fundamental similarity of SATB2 mechanisms of function in stem vs non-stem cells.

Our data reveal a surprising degree of inherent plasticity between adult ileal and colonic mucosa, likely enabled by the presence of primed ileal enhancers in colon and vice versa. Unlike many quiescent enhancers in the adult gut (Jadhav et al., 2019), the ileal enhancers activated by *Satb2* loss are not decommissioned developmental enhancers and they are not overtly subject to PRC2 regulation. Low levels of SATB2 are present in ileum but not jejunum or duodenum and exogenous SATB2 expression elicits a gradient of response along the proximal-distal axis with ileum being the most responsive and duodenum the least. This correlation suggests a potential role for endogenous ileal SATB2 in regulating this plasticity by “priming” colonic enhancers, analogous to its role in the colon. Further investigation will be needed to evaluate this possibility.

Physiologic or pathologic factors may influence SATB2 expression and decreases or increases beyond certain threshold levels may trigger ileal or colonic plasticity. It is notable that in human patients with total colectomy, the terminal ileal pouch can sometimes take on certain colonic characteristics (de Silva et al., 1991; Donati et al., 2010). It will be interesting to determine whether this tissue plasticity reflects gain of function of SATB2 or SATB2 cofactors or effectors in the ileal pouch.

The digestive tract is one of the most ancient and conserved organs across multicellular organisms. A distinct large intestine, separated from the small intestine by an ileocaecal valve, is however only well recognized in tetrapods (Schultz et al., 1989). Colon-like structures are postulated to exist in lower vertebrates but there are uncertainties (Brugman, 2016). The SATB2 gene is highly conserved across animal phyla. Studies of SATB2 may contribute to the understanding of the evolutionary origin of colon.

Limitations of study

Due to technical challenges of purifying $LGR5^{+}$ colonic stem cells in sufficient numbers, we could perform CUT&RUN for SATB2 on only one set of stem and non-stem cell samples. These results should ideally be replicated with additional samples. SATB1, a close homolog of SATB2, has been implicated in mediating chromatin looping. It will be important in future work to determine whether SATB2 regulates local or long-distance chromatin interactions in colonic stem and differentiated cells.

STAR Methods

Resource Availability

Lead contact—Further information and requests for resources should be directed to the Lead Contact, Qiao Zhou (jqz4001@weill.cornell.edu).

Materials Availability—All the materials will be available upon request to corresponding author under material transfer agreement with Weill Cornell Medicine.

Data and Code Availability—The high-through sequencing raw and processed data in this paper have been deposited to Gene Expression Omnibus (GEO) (ATAC-Seq: GSE148690 and GSE180037. ScRNA-Seq: GSE148693. ChIP-Seq: GSE167287. CUT&RUN: GSE180029 and GSE167289. Bulk RNA-Seq: GSE148692, GSE167284, GSE180023, GSE167281, GSE167282, GSE167283, GSE167285, GSE16728 and GSE180013). We also analyzed public GEO datasets: GSE115541, GSE71713 and GSE130822. (Banerjee et al., 2018; Jadhav et al., 2016; Murata et al., 2020).

Experimental model and subject details

Mouse strains—All mouse experiments were conducted under the IACUC protocol 2018–0050 at Weill Cornell Medical College or protocol 03–132 at Dana-Farber Cancer Institute. Mice were housed in a temperature- and humidity- controlled environment with 12hr light/dark cycle and food/water ad libitum. All mouse experiments were performed with both males and females at 2 months of age. The $Satb2^{loxp/loxp}$ ($Satb2^{f/f}$) strain (Dobreva et al., 2006) was a gift from Dr. Jeff Macklis of Harvard University. The Vil-Cre^{ERT2} strain (el Marjou et al., 2004) was a gift from Sylvie Robine (Institute Pasteur). Vil-Cre^{ERT2};$Eed^{loxp/loxp}$ ($Eed^{f/f}$) mice were derived as described previously (Jadhav et al., 2016; Xie et al., 2014). $Lgr5^{DTRGFP}$ strain was from Genentech Inc. (Tian et al., 2011). $Lgr5^{GFPCreER}$ strain (Barker et al., 2007) was purchased from Jackson Lab. $CAG^{SATB2GFP}$ strain was generated in this study (details of generation in Method Details section). To confer conditional deletion of floxed alleles, 4 mg per 25 g of body weight of tamoxifen (TAM, 10 mg per ml in corn oil) was intraperitoneally injected once every 2 days for a total of 3 times.

Mouse primary intestinal organoids—Mouse primary intestinal organoid culture was performed as previously described (Sugimoto and Sato, 2017). Organoid derivation was performed on ice or at 4°C unless specified. Briefly, intestinal tissues were cut into approximately 0.5 cm size pieces and incubated in 2.5 mM EDTA for 45 minutes (mins) (small intestine) or in 10 mM EDTA for 60 mins (large intestine). After vigorous pipetting with 1% BSA pre-coated 10 ml serological pipettes, epithelium cell clumps were collected by centrifugation at 300g for 5 mins. Crypts were further isolated by filtering through a 70μm cell strainer. 50–200 Crypts per 25 μl Matrigel™ droplet were cultured in either ENR (small intestine) or WENR (large intestine) medium (Table S6) in humidified chambers containing 5% CO_2 at 37°C. The formation efficiency of primary organoids was determined by dividing the number of organoids at Day 5 by the initial Crypt numbers. To assay secondary organoids, primary organoids were dissociated with TrypLE Express (3 minutes

at 37°C), resuspended in cold DMEM with 2% FBS, and centrifuged at 300 g for 3 mins. The cell pellets were embedded in Matrigel™ in a 1:5 ratio. The formation efficiency of secondary organoids was determined by dividing the number of organoids at Day 5 by the initial crypt number.

Human primary intestinal organoids and intestinal biopsy samples—Human organoids were generated from biopsy samples collected at Weill Cornell Medicine or obtained from the In Vivo Animal and Human Studies Core at University of Michigan Center for Gastrointestinal Research (Key Resource Table). To generate organoids, human colon or ileum biopsy samples were cut into ~1 mm pieces and washed with cold DPBS by pipetting 2–3 times. Samples were treated with collagenase type IV (Worthington, 2 mg/ml in F12K medium) at 37°C for 30 mins with pipetting every 10 mins. Digestion was terminated by adding F12K with 10% FBS, followed by filtration with a 100 μm cell strainer (VWR). Pelleted crypts were resuspended in human 3D Organoid Culture Medium (HCM, Table S6) and Matrigel™ with a 1:5 volume ratio and embedded with 10–20 crypts per 10 μl droplet. Human organoids were expanded in HCM and differentiated in Human 3D Organoid Differentiation Medium (HDM, Table S6) for 72 hours.

Mouse and human colonic tissues used for experimentation were generally taken from proximal colon unless indicated otherwise.

Method Details

Generation of the CAG^{SATB2GFP} transgenic mouse line—The knock-in construct, modified from pR26CAG/GFP Dest (Addgene #74281) (Chu et al., 2016), carries a CAG promoter followed by a Neomycin-transcription stop cassette flanked by Loxp sites, HA epitope-tagged murine Satb2, an IRES element, and GFP. Donor DNA consists of a 1,083kb left arm and a 4,341bp right arm. The construct was targeted to the ROSA26 locus by pro-nuclear injection paired with purified CAS9 protein (purchased from IDT) and a validated gRNA targeting ROSA26 (ACUCCAGUCUUUCUAGAAGA). The transgenic progenies were genotyped for cassette integration into the genomic locus of ROSA26. A total of 5 double transgenic lines were established by crossing with the Vil-Cre^{ERT2} mouse line. Transgene expression in adult mice was analyzed by immunohistochemistry for GFP, the HA epitope tag, and SATB2 after TAM injection at 2 months of age. This analysis yielded very similar results from all 5 transgenic lines.

CRISPR-mediated gene knockout in colonic organoids and genomic targeting efficiency calculation—*Satb2* and *Foxd2* sgRNAs were designed with either Broad Institute online software or the Synthego CRISPR design tool (Table S7) and cloned into a LentiCRISPRv2 vector (Addgene plasmid #52961) (Sanjana et al., 2014). The lentiviruses were packaged with second-generation helper plasmids by transfection with lipofectamine 3000 (Thermo Fisher Scientific, L3000015) and titrated by counting puromycin resistant clones in HEK293T cells 5 days after infection.

To generate the colonic organoids with gene ablation, single cell suspensions of 10^5 murine or human colonic organoids were mixed with 20 μl of 10^8 $TCID_{50}$/ml of virus in 200 μl medium (either WENR for murine or HCM for human) in one well of a

non-tissue culture treated 24 well plate, and centrifuged at 1,100g at 37°C for 30 mins to facilitate infection. After centrifugation, 200 µl of culture medium was added and the plate was further incubated for 4 hours at 37°C. Cells were then resuspended, pelleted, and embedded in Matrigel™. Puromycin selection (1.0 – 2.5 µg/ml) was initiated 4 days post infection and lasted for 4 days. After puromycin selection, colonic organoids were seeded into new Matrigel drops and cultured in differentiation medium (DEM) (WENR medium without WRN conditioned medium and with the addition of 1 µg/ml RSpondin and 10 µM L-161,982). 3 days after differentiation, the organoids were either directly lysed in RLT buffer (Qiagen) for RNA exaction, or incubated with cell recovery solution on ice, to remove Matrigel, for immunofluorescence and immunoblotting analyses.

The CRISPR-mediated deletion efficiency of Satb2 was analyzed with immunofluorescence and immunoblotting, using a rabbit monoclonal anti-Satb2 antibody (Key Resource Table). For Foxd2, multiple commercially available antibodies were tested, but none was found suitable for immunofluorescence or Western Blot. Instead, the disruption efficiency at the Foxd2 genomic locus was evaluated, using a DNA mismatch detection assay with T7 endonuclease1 (NEB). Genomic DNA was extracted with an E.Z.N.A tissue DNA kit (OMEGA). Foxd2 target regions were PCR amplified with Phusion High-Fidelity DNA polymerase (NEB) plus Kapa Hifi GC buffer (ThermoFisher), according to the manufacturer's protocol. PCR products were pre-amplified with forward primer: GGCATAAGCTTTGACTTCCAGTAAC and reverse primer: GTGATGAGGGCGATGTACGAATAA (Table S7), at high annealing temperature (68°C) for 10 cycles, followed by 60°C for 30 cycles. The hetero-duplexed PCR products from Foxd2 CRISPR KO and homogeneous PCR products from the control group were incubated individually or mixed at a 1:1 ratio with T7 endonuclease 1 at 37 °C for 15 mins. The reaction was stopped by adding 1mM EDTA (final concentration) and purified with the ZYMO DNA purification Kit. DNA fragment concentration was visualized by agarose gel electrophoresis and quantified with an Agilent TapeStation (A.02.02). The gene modification percentage was calculated using the following formula:

$$\%\ \text{gene modification} = 100 \times (1 - (1 - \text{fraction cleaved})1/2)$$

For the group mixed 1:1 with the control DNA fragment, the formula used is as below:

$$\%\ \text{gene modification} = 200 \times (1 - (1 - \text{fraction cleaved})1/2)$$

Bulk RNA sequencing analysis—Bulk RNA-seq was performed as previously described with the exception of mapping to the mouse reference genome mm10 instead of mm9 (Banerjee et al., 2018). Briefly, reads alignment was performed by STAR package (Dobin et al., 2013). The raw count tables were generated by featureCounts (Liao et al., 2014). The DEseq2 package was used for differential expression analysis (Love et al., 2014). The Limma package (Ritchie et al., 2015) was used to remove donor-donor variance and batch-effect. Differentially expressed genes were generally determined using parameters of adjusted p-value < 0.05 and LFC > 2 or < −2 unless specified. The heatmaps were plotted using the R package, pheatmap. GO enrichment analysis and GSEA analysis were

conducted with the clusterProfiler package (Yu et al., 2012) and GSEA desktop software (Subramanian et al., 2005)

***In vivo* time course SATB2 deletion—**Satb2^{cKO} mice were injected once with tamoxifen at 2mg per 25g body weight. The proximal 1/3 of the colon was collected at days 1, 2, 4, and 6 post-injection. Non-injected Satb2^{cKO} mice (day 0) and injected Satb2$^{f/f}$ littermates served as controls. For RNA-seq, epithelial cells were isolated by three subsequent incubations with 10mM EDTA and 1mM DL-Dithiothreitol (DTT) in cold DMEM (Gibco) for 10 min on a rotator, vigorous shaking, and collection of supernatants. All three supernatants were combined, centrifuged for 2 min at 400g and lysed in TRIZOL (Life Technologies). RNA was extracted with TRIZOL Plus RNA purification kit (Life Technologies) according to manufacturer's instructions combined with on-column DNAse-treatment (Qiagen) and sent out to Novogene Corp. inc. (CA, USA) for quality control, library preparation, and sequencing. The DEseq2 R package was used to normalize the raw feature counts. The likelihood Ratio Test (LRT) was used to identify Differentially Expressed Genes (DEGs) across time-points with the threshold adjusted p-value < 0.01. The normalized counts of DEGs were transformed by vst function and scaled by scale function. The distance of DEGs was calculated by the dist function. Gene expression patterns across time points were then clustered by hierarchical clustering. The tree was then cut by the cutree function with the parameter k=9. The mean values were used in the data visualization.

ATAC-seq—ATAC experiments were performed following the Omni-ATAC protocol (Buenrostro et al., 2015) as previously described (Banerjee et al., 2018). Briefly, 50K intestinal epithelial cells were purified by FACS and pelleted by centrifugation at 500g at 4°C for 5 mins. Nuclei were exacted in ATAC-Resuspension Buffer (RSB, 10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$) with 0.1% NP40, 0.1% Tween-20, and 0.01% Digitonin. DNA was fragmented by Nextera Tn5 Transposase (Illumina, 2003419) and immediately purified with a MiniElute PCR Purification Kit (Qiagen). For ATAC-seq library building, NEBNext 2x MasterMix (New England Biolabs) was used to pre-Amplify for 5 cycles and determine the required number of additional cycles by qPCR amplification. The final libraries were size selected (200bp to 800bp, including index) with AMPure XP beads (Beckman), purified, and loaded for sequencing.

The purified libraries were sequenced by Novogene on Illumina HiSeq-2000, to obtain paired-end 150bp reads. Alignment BAM files for the ATAC-seq were generated with the mm10 reference genome using nf-core pipelines. Narrow peaks were called using standard MACS2 (Feng et al, 2012). ATAC-seq peak files that had regions of less than 1kb from transcriptional start site (TSS) were removed using bedtools (Quinlan and Hall, 2010).. Biological replicates were concatenated and sorted, and peaks merged within a maximum distance of 500 bp.

Cell Sorting and 10X genomic sample preparation for single cell RNA sequencing—To purify intestinal cells for scRNA-seq, a published protocol was followed.(Haber et al., 2017) Briefly, murine proximal colon from TAM injected Vil-Cre^{ER};Satb2$^{f/f}$(Satb2^{cKO}) mice, or proximal colon and entire ileum from TAM injected Satb2$^{f/f}$(Controls) mice were harvested and rinsed in cold PBS. The tissues were incubated

in 20mM EDTA-PBS while rocking in a cold room for 90 mins. Every 30 mins, the tubes containing intestinal tissues were shaken vigorously for one minute, and dissociated epithelial fractions were collected. After 90 mins, the 3 collections from each tissue were combined and dissociated into single cells with TrypLE (one minute at 37°C). The single cell suspensions in FACS buffer (1% glucose, 10 mM HEPES, 10 μM Y-27632, 1 mM *N*-acetyl-l-cysteine and 2% FBS in DPBS) were passed through a 40 μm filter and stained with anti-mouse CD326 (Epcam), anti-mouse CD31 and anti-mouse CD45 (Key Resource Table). Live $Epcam^{+}$, $CD45^{-}$, and $CD31^{-}$ epithelial cells, sorted by SONY MA900 in FACS buffer, were washed with 0.4% BSA in PBS and processed with 10X genomics single cell droplet sample preparation workflow at the Genomics Core facility at Weill Cornell Medicine.

10,000 cells in Master Mix were loaded into each channel of the Chromium Controller cartridge to produce droplets. Beads-in-Emulsion (GEMs) were transferred, and GEMs-RT was undertaken in droplets by PCR incubation. After purification of first-strand cDNA from the post GEM-RT reaction mixture, barcoded and full-length cDNAs were amplified via PCR for library construction. Enzymatic fragmentation and size selection were used to optimize the cDNA amplicon size. TruSeq Read 1 (read 1 primer sequence) was added during GEM incubation. A sample index and TruSeq Read 2 (read 2 primer sequence) were added via end-repair, A-tailing, adaptor ligation, and PCR. The final libraries were assessed by an Agilent Technology 2100 Bioanalyzer and sequenced on an Illumina NovaSeq sequencer.

scRNA-seq analysis with Seurat—Sequencing data from the Illumina NovaSeq were aligned to mouse mm10 in CellRanger 3.1.0. Seurat version 3.2.0 was used to perform quality control, count normalization, and clustering on the single cell transcriptomic data using standard methods as follows: unique molecular identifiers (UMIs) which barcode each individual mRNA molecule within a cell during reverse transcription were used to remove PCR duplicates. Cells expressing fewer than 300, or greater than 5,000 genes were removed to exclude non-cells or cell aggregates. Cells expressing greater than 18 percent mitochondrial related genes were also removed.

After quality control, the objects of wild-type (WT) ileum, colon and $Satb2^{cKO}$ colon were merged (hereafter named “the combined object”) and the CellCycleScoring function was used to calculate a cell cycle score and assign a cell cycle status for each cell. Normalization of the combined object was conducted by SCTransform method in Seurat with regression out of confounding sources including mitochondrial mapping percentage and cell cycle scores (S score and G2M score). To perform linear dimensional reduction, RunPCA function was implemented with default parameters. To construct a K-nearest neighbor (KNN) graph, the FinderNeighbours function was used and took first 22 principal components as input. The FinderClusters function with default parameters a resolution of 1.5 implements modularity optimization technique to iteratively group cells together in order to cluster the cells. Non-linear dimensional reduction techniques FIt-SNE and UMAP were used to visualize the results.

The enriched transcripts in each cell cluster or groups of clusters were identified by the FindAllMarkers function using the following parameters: only.pos = TRUE, min.pct = 0.3, logfc.threshold = 0.3, and referred to as “global markers”. Each cluster was then annotated based on the markers. To refine the annotation, wild-type colon object was additionally subseted and re-normalization performed. The cells from colonic object were re-clustered using first 30 PCs and a resolution of 1. The markers of WT colonic clusters were identified by the FindAllMarkers with the same parameters mentioned above. Each colonic cell cluster was annotated and re-assigned back to the combined object to refine cluster annotation.

For integrated analysis, per best practice suggestions in the Seurat package (Butler et al., 2018), the SelectIntegrationFeatures function was used to select the genes that were taken as input in the anchors identification procedure by the PrepSCTIntegration and the FindIntegrationAnchors functions using the merged control colon and ileum object as reference and the knockout object as query. The integration of the KO and WT objects were implemented by the IntegratedData function with the anchors identified previously. Dimension reduction, clustering and visualization were performed using the same methods mentioned above.

For ileal cell type scoring, the control wild type objects were merged and SCTransformed using the same parameters mentioned above with the annotation retained. The scoring gene list for each cell type consists of the top 20 (avg_logFC) global cluster markers and all the differentially expressed genes between each ileal cell type and its colonic counterpart (e.g. enterocytes vs colonocytes) identified by the FindAllMarkers function and the FindMarkers function respectively with the following parameters: only.pos =TRUE, min.pct = 0.3, logfc.threshold = 0.3. The MHC genes were among the most differentially expressed between small intestine and colon. Their expression is strongly influenced by the microbiome and they do not constitute an intrinsic feature of the ileum. MHCII genes were removed from the gene lists. The gene lists were then applied as inputs in the AddModuleScore function with default parameters, which calculates module scores for feature expression programs on single cell level.

In order to identify stem cells in the samples, we firstly subset the combined object for the progenitor group and divide it into two groups based on the detection of Lgr5 gene expression. The stem cell gene list was taken as input for the AddModuleScore with default parameters to calculate stem cell signature score between these two groups. Eventually, the combined object was subset for the Lgr5 positive “Progenitor” in the G1 or S cell cycle. The subset went through the standard analysis procedure as mentioned above for normalization, clustering and visualization with the first 20 PCs and a resolution of 0.8. Four clusters were identified, and their makers were found by the FindAllMarkers function. According to the markers, cluster 1 appeared to be Goblet progenitors and thus was removed.

Dual Cross-linking ChIP-Seq and Cut & Run Enhancer-Seq—ChIP for Transcription Factors (TFs) SATB2, HNF4A, and CDX2, was performed as described (Saxena et al., 2017). EDTA stripped primary intestinal glands were cross-linked with 2 mM disuccinimidyl glutarate (DSG, Thermo Fisher Scientific, 20593) at room temperature (RT) for 45 mins, followed by 1% formaldehyde (Sigma, F8775) fixation for 10 mins. For

each experiment, 50 μl of pelleted cross-linked cells were resuspended in 350 μl sarkosyl lysis buffer (0.25% sarkosyl, 1 mM DTT and protease inhibitor in RIPA buffer (0.1% SDS, 1% Triton X-100, 10 mM Tris HCl, 1 mM EDTA, 0.1% sodium deoxycholate, 0.3 M sodium chloride, PH 7.5)) and sonicated at 15% amplification by a tip sonicator (Qsonica, Q125) to obtain 200bp to 800bp chromatin fragments. Lysates were spun down at 20,000g at 4°C to remove insoluble fractions, then diluted in RIPA buffer with protease inhibitor in a final 2 ml volume. Diluted lysates were incubated with TFs antibodies (Key Resource Table) at 4°C overnight and were additionally incubated with 30 μl protein A/G magnetic beads (Thermo Fisher Scientific, 88803) for 90 mins the next day. This was followed by 6 washes with cold RIPA buffer beads. Cross-links were reversed overnight by incubating at 65°C in 1% SDS and 0.1 M $NaHCO_3$. Any remaining proteins were digested by Proteinase K (Thermo Fisher Scientific, 26160) for 1 hour at 37°C. DNA was purified with a MinElute purification kit (Qiagen, 28004). Libraries were prepared using the ThruPLEX DNA-Seq Kit (Takara bio, R400428 and R400427).

Cut & Run was performed by the Center for Epigenetics Research (CER) in Memorial Sloan Kettering Cancer Center. Briefly. Single cell suspensions were collected as described in the single cell RNA sequencing section. Dead cells were removed using a Dead Cell Removal Kit (Miltenyi Biotec). 10^5 cells were attached to Concanavalin A conjugated magnetic beads, permeabilized, and incubated with histone enhancer maker antibodies (Key Resource Table) at RT for 20 mins. pAG-MNase (1:1000) was added in digitonin buffer (5% digitonin, 60 mM HEPES, 0.5 M sodium chloride, 1.5 mM spermidine hydrochloride, protease inhibitor, PH 7.5) to bind with antibodies. Finally, targeted chromatins were digested and released into the supernatant. DNA was purified with a MinElute purification kit. Libraries were prepared using the ThruPLEX DNA-Seq Kit (Takara bio, R400665). All the libraries were size selected (200–800 bp) by AMPure XP beads and loaded for sequencing.

SATB2 Cut & Run experiment with FACS purified $LGR5^+$ stem cells was performed with the same method as Cut & Run described above, using polyclonal anti-SATB2 or control Rabbit IgG antibodies (Key Resource Table) incubated overnight at 4°C.

ChIP-Seq analyses for transcription factors—All reads (CDX2, HNF4A and SATB2) were trimmed with trim_galore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), and subject to quality control with FastQC before and after adapter trimming. For ChIP-Seq, Bowtie2 (Langmead and Salzberg, 2012) was used to align the two independent ChIP-Seq analyses to the mouse (mm10) genome with default parameters. Aligned ChIP-Seq data in SAM format were transformed to BAM files and non-uniquely mapped reads were filtered-out. Duplicate alignments were then marked and removed using Sambamba (Tarasov et al., 2015). The merge function in samtools (Li et al., 2009) was used to merge the BAM files of different replicates and filter out non-uniquely mapped reads. Deeptools (Ramirez et al., 2014) bamCoverage (duplicate reads ignored, RPKM normalized) was used to generate bigWig files from BAM files. Reads that overlapped with the Broad Institute sequencing blacklist (ENCODE Project Consortium, http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/mm10-mouse/mm10.blacklist.bed.gz) were discarded. The mapped reads from the biological replicates were combined for each factor and then peak calling was performed using the ChIP-Seq (macs2) (Feng et al., 2012) peak caller

(v2.2.7) with parameters callpeak -f BAMPE -g mm -p 0.0000000001, and was controlled by KO/input. Heat maps of ChIP-Seq were created by quantile normalized bigWigs using computeMatrix, plotHeatmap, and plotProfile from deeptools.

MAnorm(Shao et al., 2012), software designed for quantitative comparisons of ChIP-Seq datasets, was applied to compare ChIP-Seq signal intensities between samples. The window size was 1 kb, which matched the average width of the identified ChIP-Seq peaks. Tissue specific peaks were defined using the following criteria: (1) defined as ‘unique’ by the MAnorm algorithm, (2) P value < 0.01, (3) raw counts of unique reads > 10. Peaks common to two samples were defined using the following criteria: (1) defined as ‘common’ by the MAnorm algorithm and (2) raw read counts of both samples > 10.

The annotatePeaks function in HOMER (Heinz et al., 2010) was used to annotate the peaks. To identify the distribution of the binding sites of ChIP-Seq data, peak sites were mapped to TSS (transcription start site), TTS (transcription termination site), Exon (Coding), 5’ UTR Exon, 3’ UTR Exon, Intronic, or Intergenic, which are common annotations defined by HOMER. A promoter region was defined as a region within ± 2 Kb from the TSS. Enriched motifs were identified within 200 bp regions centered on SATB2 ChIP-seq peak summits using findMotifsGenome.pl with options ‘-length -len “8,10,12”‘ and ‘-size 200’ on the repeat-masked mouse genome (mm10r) from HOMER.

Cut & Run analyses for histone modifications—Reads (H3K4me1 and H3K27ac) were trimmed with trim_galore. Paired-end reads were then mapped to the mm10 genome using Bowtie2, with parameters as previously detailed,(Skene et al., 2018) using --local --very-sensitive-local --no-unal --no-mixed --no-discordant --phred33 -I 10 -X 700. Only uniquely mapped reads were retained with samtools. Peaks were called from macs2 with pooled reads and with both replicate samples by merge broad and narrow peak files. For the enhancer analysis, only peaks ± 2 Kb outside TSSs were permitted, using ‘distal peaks’ as enhancer peaks. Tissue specific enhancers were identified between ileum and colon using MAnorm.

Immunohistochemistry, Edu labeling, and Western Blot—Intestinal tissues were processed as previously described.(Ariyachet et al., 2016) Organoids were removed from Matrigel with Cell Recovery Solution (Corning 354253), fixed with 4% paraformaldehyde in PBS on ice for 30 mins, and then processed with the same procedure as the intestinal tissues. Immunohistochemistry was performed using a standard procedure, incubating with primary antibodies (Key Resource Table) at 4°C overnight, followed with secondary antibodies (Key Resource Table) at room temperature for 45 mins. A Click-iT™ EDU Cell Proliferation Kit with Alexa Fluor® 555 (C10338) was used to evaluate proliferation. The images were captured using either a confocal microscope (710 Meta) or a Nikon fluorescence microscope. For Western blot analysis, a monoclonal rabbit anti-SATB2 antibody was used to bind SATB2 protein, followed by an incubation with a secondary anti-Rabbit Peroxidase (HRP). Protein bands were visualized using enhanced chemiluminescent substrate (Pico from Thermo fisher) and recorded by a Li-COR C-Digit or li-COR odyssey clx blot scanner. The relative signal intensity was quantified by Image J (v1.51 (100)).

For immunohistochemistry, samples were processed through heat mediated antigen retrieval in Citric Acid buffer (pH 6.0) except for the samples that stained for monoclonal anti-SATB2 antibodies, which were processed in Tris-EDTA (pH 9.0). Samples were then stained with primary rabbit antibodies (Key Resource Table), followed by Goat anti-Rabbit HRP polymer (Vector Laboratories, MP-7451) incubation, and finally, developed with AP (Magenta color, Vector Laboratories, MP-7724) or DAB (Brown color, Vector Laboratories, SK-4103) HRP Substrate. The images of swiss rolled colon were taken by a confocal digital slide scanner in MSKCC image core and processed by Caseviewer (v2.4). An Alcian Blue Stain Kit (Vector Laboratories, H-3501) was used to stain goblet cells.

Quantitative Analysis of Histological Staining and Fluorescence in ImageJ— All sections were evaluated by multiple people, including a clinical pathologist. All image quantifications were done in ImageJ (Fiji, Version: 2.1.0/1.53c) as previously described (Fuhrich et al., 2013; Jensen, 2013). Briefly, immunofluorescence images were split into individual channel by click image > color > split channel. Relative signal intensity was calculated by comparison to the average density of the controls.

For immunohistochemistry images, the hematoxylin and specific antibody staining were separated into 3 different panels with the function of color deconvolution for PAS (AP development) or H-DAB (DAB development). Next, epithelial area was overlaid on the AP/DAB signal channel image. The final epithelial AP/DAB intensity (*f*) was calculated according the formula: $f = 255 -$ mean intensity (obtained from the software analysis, range from 0 – 255, zero = deep brown, highest expression, 255 = total white). Relative signal intensity was calculated by comparison to the average density of the controls.

Immunoprecipitation—EDTA stripped colonic grand epithelium cells from control and $Satb2^{cKO}$ mice were cross-linked with DSP (Thermo Fisher Scientific, PG82081) at RT for 45 mins. Pellets of epithelial cells were incubated with RIPA buffer and sonicated at 15% amplification for 20 seconds. After 10 mins (maximum speed down), supernatants were incubated with anti-CDX2 and anti-HNF4A (Key Resource Table) overnight in a cold room with a rotation speed of 10 RPM. After adding 30 μl protein A/G magnetic beads for 90 mins on the next day, the protein and beads complex was pulled down by a magnetic stander. Next, 6 cold RIPA buffer washes were performed. Then cross-links were cleaved by 50 mM DTT boiling for 5 mins. Immunoblots were used to visualize the interaction between target proteins.

[^{14}C]-Taurocholic acid and [^{3}H]-Glucose *In vivo* absorption study—^{14}C-Taurocholic acid and ^{3}H-Glucose were purchased from American Radiolabeled Chemicals, Inc. To perform the absorption study, mice were fasted overnight for ~16 hours. Following deep anesthetization, a 2cm section of the distal ileum or proximal colon was cleaned of luminal content by repeated flushing with saline and was tied on both ends with sutures to create a sealed pouch. 2 μCi ^{3}H-glucose and 0.6 μCi ^{14}C-Taurocholic acid dissolved in 100 μl 10% dextrose solution were injected into the pouch. After 5 or 20 mins, blood was collected from the hepatic portal vein with a 27-gauge needle. Plasma was harvested after centrifugation at 12,000 rpm for 4 mins at 4°C. Plasma protein was precipitated with the addition of $Ba(OH)_2$ and $ZnSO_4$ to 20 μl plasma. The supernatant was dissolved in Ultima

Gold Scintillation fluid (PerkinElmer). A liquid scintillation counter with dual channels for ^{3}H and ^{14}C was used to measure radioactivity in all samples.

For liver sampling, the right upper lobe from each mouse was removed and stored at −20°C. Frozen tissues were homogenized in dH_2O (50 mg of tissue in 500 μl of dH_2O) with a Dounce homogenizer. Following homogenization, the glass tubes were placed in a heat block for 10 mins at 100°C, vortexed, and cooled to room temperature. The homogenized samples were centrifuged at 16,000g for 5 mins. Supernatants were collected. 500 μl of supernatant per sample was added to scintillation vials containing the scintillation cocktail for counting.

Disaccharidase and Dipeptidyl Peptidase IV (DPP4) Assay—After Matrigel removal, differentiated human organoids were transferred to BSA-pre-coated 1.5 ml Eppendorf tubes and washed three times in PBS. For the disaccharidase enzyme activity assay, 5 mg of an organoid pellet was incubated with 100 μl of 56 mM sucrose in PBS or PBS only at 37°C for 45 mins. Aliquots of the supernatant were sampled for glucose detection using the Glucose Colorimetric Assay Kit (Cayman), according to the manufacturer's protocol. Briefly, the samples were diluted with PBS in a 1:1 and 1:2 ratio to ensure glucose concentration levels in the standard range (0–25 mg/dl). The enzyme and samples mixtures were incubated at 37°C for 10 mins. The absorbance (510 nm) was measured with a plate reader (SpectraMax M2). Glucose concentration was determined by comparison to a glucose standard curve. For the DPP4 assay, Gly-Pro-p-nitroanilide hydrochloride (Sigma, G0513) in PBS was added to an organoid pellet at a final concentration of 1.5 mM. The organoid tubes were incubated at 37°C in a tissue culture incubator with the lip open for 30 mins and were mixed every 10 mins. The supernatants were collected and absorbance was measured at 410 nm with a plate reader (SpectraMax M2). Released nitroanilide concentration was determined by comparison to a 4-nitroanilide (Sigma, 185310) standard curve (0 – 200 μg/ml). The concentration was finally normalized to a total cell lysate protein amount of 1 mg.

Quantification and statistical analysis

Quantification methods and statistical analysis are described in the figure legends. The exact biological or technical replicates were indicated within individual figure legends. The statistical results were presented as the means, individual values and error bars represent SD. GraphPad Prism 9 or R was used to determine statistical significance by unpaired/paired Student t test or Mann-Whitney U test if the data do not meet t test requirements (Normal distribution and similar variance). The exact p values were reported in each figure or indicated as ***, $p \leq 0.001$; **, $p \leq 0.01$; *, $p \leq 0.05$, ns = not significant.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank the WCM Epigenetics Core (Jenny Xiang), WCM Flow Cytometry Core (Jason McCormick and Tomas Baumgartner), Tri-Institutional Stem cell Derivation Core (Raphael Lis and Tyler Lu), MSKCC Center

for Epigenetics Research (Ruifang Li), and Harvard Genome Modification Facility (Lin Wu) for excellent technical assistance. We are grateful to Shariq Madha for assistance with computational analyses, Dr. Frederick J de Sauvage for providing Lgr5^{DTRGFP} mice, Drs. James Wells and Todd Evans for discussion, and Jonathan Colarusso for editing the manuscript. This work was supported by Weill Cornell Medicine funds to Q.Z, a Tri-Institutional Stem Cell grant to Q.Z and K.H, NIH/NIDDK grants to R.T.L. (1R01DK107396) and to M.P.V. (R01DK121915), a NNF Center grant for Stem Cell Biology (no. NNF17CC0027852) to K.H, and through the MSKCC Support Grant (no. NIH P30 CA008748). J.K. is grateful to support from the Deutsche Forschungsgemeinschaft (DFG) (KR 5451/1-1).

References

Alcamo EA, Chirivella L, Dautzenberg M, Dobreva G, Farinas I, Grosschedl R, and McConnell SK (2008). Satb2 regulates callosal projection neuron identity in the developing cerebral cortex. Neuron 57, 364–377. [PubMed: 18255030]

Ariyachet C, Tovaglieri A, Xiang G, Lu J, Shah MS, Richmond CA, Verbeke C, Melton DA, Stanger BZ, Mooney D, et al. (2016). Reprogrammed Stomach Tissue as a Renewable Source of Functional beta Cells for Blood Glucose Regulation. Cell Stem Cell 18, 410–421. [PubMed: 26908146]

Banerjee KK, Saxena M, Kumar N, Chen L, Cavazza A, Toke NH, O'Neill NK, Madha S, Jadhav U, Verzi MP, et al. (2018). Enhancer, transcriptional, and cell fate plasticity precedes intestinal determination during endoderm development. Genes Dev 32, 1430–1442. [PubMed: 30366903]

Barker N, van Es JH, Kuipers J, Kujala P, van den Born M, Cozijnsen M, Haegebarth A, Korving J, Begthel H, Peters PJ, et al. (2007). Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature 449, 1003–1007. [PubMed: 17934449]

Beumer J, and Clevers H (2021). Cell fate specification and differentiation in the adult mammalian intestine. Nat Rev Mol Cell Biol 22, 39–53. [PubMed: 32958874]

Biton M, Haber AL, Rogel N, Burgin G, Beyaz S, Schnell A, Ashenberg O, Su CW, Smillie C, Shekhar K, et al. (2018). T Helper Cell Cytokines Modulate Intestinal Stem Cell Renewal and Differentiation. Cell 175, 1307–1320 e1322. [PubMed: 30392957]

Blanpain C, and Fuchs E (2014). Stem cell plasticity. Plasticity of epithelial stem cells in tissue regeneration. Science 344, 1242281. [PubMed: 24926024]

Britanova O, de Juan Romero C, Cheung A, Kwan KY, Schwark M, Gyorgy A, Vogel T, Akopov S, Mitkovski M, Agoston D, et al. (2008). Satb2 is a postmitotic determinant for upper-layer neuron specification in the neocortex. Neuron 57, 378–392. [PubMed: 18255031]

Britanova O, Depew MJ, Schwark M, Thomas BL, Miletich I, Sharpe P, and Tarabykin V (2006). Satb2 haploinsufficiency phenocopies 2q32-q33 deletions, whereas loss suggests a fundamental role in the coordination of jaw development. Am J Hum Genet 79, 668–678. [PubMed: 16960803]

Brugman S (2016). The zebrafish as a model to study intestinal inflammation. Dev Comp Immunol 64, 82–92. [PubMed: 26902932]

Buenrostro JD, Wu B, Chang HY, and Greenleaf WJ (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol 109, 21 29 21–21 29 29.

Butler A, Hoffman P, Smibert P, Papalexi E, and Satija R (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 36, 411–420. [PubMed: 29608179]

Cai S, Han HJ, and Kohwi-Shigematsu T (2003). Tissue-specific nuclear architecture and gene expression regulated by SATB1. Nat Genet 34, 42–51. [PubMed: 12692553]

Chen L, Toke NH, Luo S, Vasoya RP, Fullem RL, Parthasarathy A, Perekatt AO, and Verzi MP (2019). A reinforcing HNF4-SMAD4 feed-forward module stabilizes enterocyte identity. Nat Genet 51, 777–785. [PubMed: 30988513]

Chu VT, Weber T, Graf R, Sommermann T, Petsch K, Sack U, Volchkov P, Rajewsky K, and Kuhn R (2016). Efficient generation of Rosa26 knock-in mice using CRISPR/Cas9 in C57BL/6 zygotes. BMC Biotechnol 16, 4. [PubMed: 26772810]

Clevers H, and Watt FM (2018). Defining Adult Stem Cells by Function, not by Phenotype. Annu Rev Biochem 87, 1015–1027. [PubMed: 29494240]

Davison JM, Lickwar CR, Song L, Breton G, Crawford GE, and Rawls JF (2017). Microbiota regulate intestinal epithelial gene expression by suppressing the transcription factor Hepatocyte nuclear factor 4 alpha. Genome Res 27, 1195–1206. [PubMed: 28385711]

de Silva HJ, Millard PR, Kettlewell M, Mortensen NJ, Prince C, and Jewell DP (1991). Mucosal characteristics of pelvic ileal pouches. Gut 32, 61–65. [PubMed: 1846839]

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. [PubMed: 23104886]

Dobreva G, Chahrour M, Dautzenberg M, Chirivella L, Kanzler B, Farinas I, Karsenty G, and Grosschedl R (2006). SATB2 is a multifunctional determinant of craniofacial patterning and osteoblast differentiation. Cell 125, 971–986. [PubMed: 16751105]

Donati G, Rognoni E, Hiratsuka T, Liakath-Ali K, Hoste E, Kar G, Kayikci M, Russell R, Kretzschmar K, Mulder KW, et al. (2017). Wounding induces dedifferentiation of epidermal Gata6(+) cells and acquisition of stem cell properties. Nat Cell Biol 19, 603–613. [PubMed: 28504705]

Donati OF, Weishaupt D, Weber A, and Hahnloser D (2010). Colonic transformation of ileal pouch-anal anastomosis and of the distal ileum: MRI findings. Br J Radiol 83, e185–187. [PubMed: 20739339]

el Marjou F, Janssen KP, Chang BH, Li M, Hindie V, Chan L, Louvard D, Chambon P, Metzger D, and Robine S (2004). Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. Genesis 39, 186–193. [PubMed: 15282745]

Feng J, Liu T, Qin B, Zhang Y, and Liu XS (2012). Identifying ChIP-seq enrichment using MACS. Nat Protoc 7, 1728–1740. [PubMed: 22936215]

Fordham RP, Yui S, Hannan NR, Soendergaard C, Madgwick A, Schweiger PJ, Nielsen OH, Vallier L, Pedersen RA, Nakamura T, et al. (2013). Transplantation of expanded fetal intestinal progenitors contributes to colon regeneration after injury. Cell Stem Cell 13, 734–744. [PubMed: 24139758]

Fuhrich DG, Lessey BA, and Savaris RF (2013). Comparison of HSCORE assessment of endometrial beta3 integrin subunit expression with digital HSCORE using computerized image analysis (ImageJ). Anal Quant Cytopathol Histpathol 35, 210–216. [PubMed: 24341124]

Gao N, White P, and Kaestner KH (2009). Establishment of intestinal identity and epithelial-mesenchymal signaling by Cdx2. Dev Cell 16, 588–599. [PubMed: 19386267]

Gehart H, and Clevers H (2019). Tales from the crypt: new insights into intestinal stem cells. Nat Rev Gastroenterol Hepatol 16, 19–34. [PubMed: 30429586]

Giroux V, and Rustgi AK (2017). Metaplasia: tissue injury adaptation and a precursor to the dysplasia-cancer sequence. Nat Rev Cancer 17, 594–604. [PubMed: 28860646]

Grainger S, Hryniuk A, and Lohnes D (2013). Cdx1 and Cdx2 exhibit transcriptional specificity in the intestine. PLoS One 8, e54757. [PubMed: 23382958]

Haber AL, Biton M, Rogel N, Herbst RH, Shekhar K, Smillie C, Burgin G, Delorey TM, Howitt MR, Katz Y, et al. (2017). A single-cell survey of the small intestinal epithelium. Nature 551, 333–339. [PubMed: 29144463]

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, and Glass CK (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38, 576–589. [PubMed: 20513432]

Jadhav U, Cavazza A, Banerjee KK, Xie H, O'Neill NK, Saenz-Vash V, Herbert Z, Madha S, Orkin SH, Zhai H, et al. (2019). Extensive Recovery of Embryonic Enhancer and Gene Memory Stored in Hypomethylated Enhancer DNA. Mol Cell 74, 542–554 e545. [PubMed: 30905509]

Jadhav U, Nalapareddy K, Saxena M, O'Neill NK, Pinello L, Yuan GC, Orkin SH, and Shivdasani RA (2016). Acquired Tissue-Specific Promoter Bivalency Is a Basis for PRC2 Necessity in Adult Cells. Cell 165, 1389–1400. [PubMed: 27212235]

Jensen EC (2013). Quantitative analysis of histological staining and fluorescence using ImageJ. Anat Rec (Hoboken) 296, 378–381. [PubMed: 23382140]

Langmead B, and Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–359. [PubMed: 22388286]

Leushacke M, Tan SH, Wong A, Swathi Y, Hajamohideen A, Tan LT, Goh J, Wong E, Denil S, Murakami K, et al. (2017). Lgr5-expressing chief cells drive epithelial regeneration and cancer in the oxyntic stomach. Nat Cell Biol 19, 774–786. [PubMed: 28581476]

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. [PubMed: 19505943]

Liao Y, Smyth GK, and Shi W (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930. [PubMed: 24227677]

Love MI, Huber W, and Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550. [PubMed: 25516281]

Munera JO, and Wells JM (2017). Generation of Gastrointestinal Organoids from Human Pluripotent Stem Cells. Methods Mol Biol 1597, 167–177. [PubMed: 28361317]

Munoz J, Stange DE, Schepers AG, van de Wetering M, Koo BK, Itzkovitz S, Volckmann R, Kung KS, Koster J, Radulescu S, et al. (2012). The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent '+4' cell markers. EMBO J 31, 3079–3091. [PubMed: 22692129]

Murata K, Jadhav U, Madha S, van Es J, Dean J, Cavazza A, Wucherpfennig K, Michor F, Clevers H, and Shivdasani RA (2020). Ascl2-Dependent Cell Dedifferentiation Drives Regeneration of Ablated Intestinal Stem Cells. Cell Stem Cell 26, 377–390 e376. [PubMed: 32084390]

Mutoh H, Hakamata Y, Sato K, Eda A, Yanaka I, Honda S, Osawa H, Kaneko Y, and Sugano K (2002). Conversion of gastric mucosa to intestinal metaplasia in Cdx2-expressing transgenic mice. Biochem Biophys Res Commun 294, 470–479. [PubMed: 12051735]

Nichols RG, and Davenport ER (2021). The relationship between the gut microbiome and host gene expression: a review. Hum Genet 140, 747–760. [PubMed: 33221945]

Nusse YM, Savage AK, Marangoni P, Rosendahl-Huber AKM, Landman TA, de Sauvage FJ, Locksley RM, and Klein OD (2018). Parasitic helminths induce fetal-like reversion in the intestinal stem cell niche. Nature 559, 109–113. [PubMed: 29950724]

Page ME, Lombard P, Ng F, Gottgens B, and Jensen KB (2013). The epidermis comprises autonomous compartments maintained by distinct stem cell populations. Cell Stem Cell 13, 471–482. [PubMed: 23954751]

Perez Montiel D, Arispe Angulo K, Cantu-de Leon D, Bornstein Quevedo L, Chanona Vilchis J, and Herrera Montalvo L (2015). The value of SATB2 in the differential diagnosis of intestinal-type mucinous tumors of the ovary: primary vs metastatic. Ann Diagn Pathol 19, 249–252. [PubMed: 26059401]

Quinlan AR, and Hall IM (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. [PubMed: 20110278]

Ramirez F, Dundar F, Diehl S, Gruning BA, and Manke T (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res 42, W187–191. [PubMed: 24799436]

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, and Smyth GK (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43, e47. [PubMed: 25605792]

Sanjana NE, Shalem O, and Zhang F (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat Methods 11, 783–784. [PubMed: 25075903]

Santos AJM, Lo YH, Mah AT, and Kuo CJ (2018). The Intestinal Stem Cell Niche: Homeostasis and Adaptations. Trends Cell Biol 28, 1062–1078. [PubMed: 30195922]

Sato T, Stange DE, Ferrante M, Vries RG, Van Es JH, Van den Brink S, Van Houdt WJ, Pronk A, Van Gorp J, Siersema PD, et al. (2011). Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. Gastroenterology 141, 1762–1772. [PubMed: 21889923]

Sato T, Vries RG, Snippert HJ, van de Wetering M, Barker N, Stange DE, van Es JH, Abo A, Kujala P, Peters PJ, et al. (2009). Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature 459, 262–265. [PubMed: 19329995]

Saxena M, Roman AKS, O'Neill NK, Sulahian R, Jadhav U, and Shivdasani RA (2017). Transcription factor-dependent 'anti-repressive' mammalian enhancers exclude H3K27me3 from extended genomic domains. Genes Dev 31, 2391–2404. [PubMed: 29321178]

Schultz SG, Rauner BB, Wood JD, and American Physiological Society (1887-) (1989). The Gastrointestinal system. (Bethesda, Md. New York, N.Y.: American Physiological Society; Distributed by Oxford University Press).

Shao Z, Zhang Y, Yuan GC, Orkin SH, and Waxman DJ (2012). MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. Genome Biol 13, R16. [PubMed: 22424423]

Skene PJ, Henikoff JG, and Henikoff S (2018). Targeted in situ genome-wide profiling with high efficiency for low cell numbers. Nat Protoc 13, 1006–1019. [PubMed: 29651053]

Skowronska-Krawczyk D, Ma Q, Schwartz M, Scully K, Li W, Liu Z, Taylor H, Tollkuhn J, Ohgi KA, Notani D, et al. (2014). Required enhancer-matrin-3 network interactions for a homeodomain transcription program. Nature 514, 257–261. [PubMed: 25119036]

Slack JM (2007). Metaplasia and transdifferentiation: from pure biology to the clinic. Nat Rev Mol Cell Biol 8, 369–378. [PubMed: 17377526]

Spitz F, and Furlong EE (2012). Transcription factors: from enhancer binding to developmental control. Nat Rev Genet 13, 613–626. [PubMed: 22868264]

Stange DE, Koo BK, Huch M, Sibbel G, Basak O, Lyubimova A, Kujala P, Bartfeld S, Koster J, Geahlen JH, et al. (2013). Differentiated Troy+ chief cells act as reserve stem cells to generate all lineages of the stomach epithelium. Cell 155, 357–368. [PubMed: 24120136]

Stergachis AB, Neph S, Reynolds A, Humbert R, Miller B, Paige SL, Vernot B, Cheng JB, Thurman RE, Sandstrom R, et al. (2013). Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. Cell 154, 888–903. [PubMed: 23953118]

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102, 15545–15550. [PubMed: 16199517]

Sugimoto S, Ohta Y, Fujii M, Matano M, Shimokawa M, Nanki K, Date S, Nishikori S, Nakazato Y, Nakamura T, et al. (2018). Reconstruction of the Human Colon Epithelium In Vivo. Cell Stem Cell 22, 171–176 e175. [PubMed: 29290616]

Sugimoto S, and Sato T (2017). Establishment of 3D Intestinal Organoid Cultures from Intestinal Stem Cells. Methods Mol Biol 1612, 97–105. [PubMed: 28634937]

Szemes M, Gyorgy A, Paweletz C, Dobi A, and Agoston DV (2006). Isolation and characterization of SATB2, a novel AT-rich DNA binding protein expressed in development- and cell-specific manner in the rat brain. Neurochem Res 31, 237–246. [PubMed: 16604441]

Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, and Prins P (2015). Sambamba: fast processing of NGS alignment formats. Bioinformatics 31, 2032–2034. [PubMed: 25697820]

Tata PR, Mou H, Pardo-Saganta A, Zhao R, Prabhu M, Law BM, Vinarsky V, Cho JL, Breton S, Sahay A, et al. (2013). Dedifferentiation of committed epithelial cells into stem cells in vivo. Nature 503, 218–223. [PubMed: 24196716]

Tetteh PW, Farin HF, and Clevers H (2015). Plasticity within stem cell hierarchies in mammalian epithelia. Trends Cell Biol 25, 100–108. [PubMed: 25308311]

Thaiss CA, Levy M, Korem T, Dohnalova L, Shapiro H, Jaitin DA, David E, Winter DR, Gury-BenAri M, Tatirovsky E, et al. (2016). Microbiota Diurnal Rhythmicity Programs Host Transcriptome Oscillations. Cell 167, 1495–1510 e1412. [PubMed: 27912059]

Thompson CA, DeLaForest A, and Battle MA (2018). Patterning the gastrointestinal epithelium to confer regional-specific functions. Dev Biol 435, 97–108. [PubMed: 29339095]

Tian H, Biehs B, Warming S, Leong KG, Rangell L, Klein OD, and de Sauvage FJ (2011). A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable. Nature 478, 255–259. [PubMed: 21927002]

Umesaki Y, Okada Y, Matsumoto S, Imaoka A, and Setoyama H (1995). Segmented filamentous bacteria are indigenous intestinal bacteria that activate intraepithelial lymphocytes and induce MHC class II molecules and fucosyl asialo GM1 glycolipids on the small intestinal epithelial cells in the ex-germ-free mouse. Microbiol Immunol 39, 555–562. [PubMed: 7494493]

van Es JH, Sato T, van de Wetering M, Lyubimova A, Yee Nee AN, Gregorieff A, Sasaki N, Zeinstra L, van den Born M, Korving J, et al. (2012). Dll1+ secretory progenitor cells revert to stem cells upon crypt damage. Nat Cell Biol 14, 1099–1104. [PubMed: 23000963]

VanDussen KL, Sonnek NM, and Stappenbeck TS (2019). L-WRN conditioned medium for gastrointestinal epithelial stem cell culture shows replicable batch-to-batch activity levels across multiple research teams. Stem Cell Res 37, 101430. [PubMed: 30933720]

Verzi MP, Shin H, Ho LL, Liu XS, and Shivdasani RA (2011). Essential and redundant functions of caudal family proteins in activating adult intestinal genes. Mol Cell Biol 31, 2026–2039. [PubMed: 21402776]

Verzi MP, Shin H, San Roman AK, Liu XS, and Shivdasani RA (2013). Intestinal master transcription factor CDX2 controls chromatin access for partner transcription factor binding. Mol Cell Biol 33, 281–292. [PubMed: 23129810]

Wang Y, Chiang IL, Ohara TE, Fujii S, Cheng J, Muegge BD, Ver Heul A, Han ND, Lu Q, Xiong S, et al. (2019). Long-Term Culture Captures Injury-Repair Cycles of Colonic Stem Cells. Cell 179, 1144–1159 e1115. [PubMed: 31708126]

Wells JM, and Spence JR (2014). How to make an intestine. Development 141, 752–760. [PubMed: 24496613]

Wells JM, and Watt FM (2018). Diverse mechanisms for endogenous regeneration and repair in mammalian organs. Nature 557, 322–328. [PubMed: 29769669]

Xie H, Xu J, Hsu JH, Nguyen M, Fujiwara Y, Peng C, and Orkin SH (2014). Polycomb repressive complex 2 regulates normal hematopoietic stem cell function in a developmental-stage-specific manner. Cell Stem Cell 14, 68–80. [PubMed: 24239285]

Yasui D, Miyano M, Cai S, Varga-Weisz P, and Kohwi-Shigematsu T (2002). SATB1 targets chromatin remodelling to regulate genes over long distances. Nature 419, 641–645. [PubMed: 12374985]

Yu G, Wang LG, Han Y, and He QY (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16, 284–287. [PubMed: 22455463]

Yui S, Azzolin L, Maimets M, Pedersen MT, Fordham RP, Hansen SL, Larsen HL, Guiu J, Alves MRP, Rundsten CF, et al. (2018). YAP/TAZ-Dependent Reprogramming of Colonic Epithelium Links ECM Remodeling to Tissue Regeneration. Cell Stem Cell 22, 35–49 e37. [PubMed: 29249464]

Zarate YA, and Fish JL (2017). SATB2-associated syndrome: Mechanisms, phenotype, and practical recommendations. Am J Med Genet A 173, 327–337. [PubMed: 27774744]

Zaret KS, and Mango SE (2016). Pioneer transcription factors, chromatin dynamics, and cell fate control. Curr Opin Genet Dev 37, 76–81. [PubMed: 26826681]

Highlights:

- Colon-to-ileum mucosal conversion in SATB2 mutant mice and human organoids
- SATB2 maintains colonic identity of both stem and differentiated cells
- Ectopic SATB2 converts ileum to colon
- SATB2 regulates enhancer binding of CDX2 and HNF4A

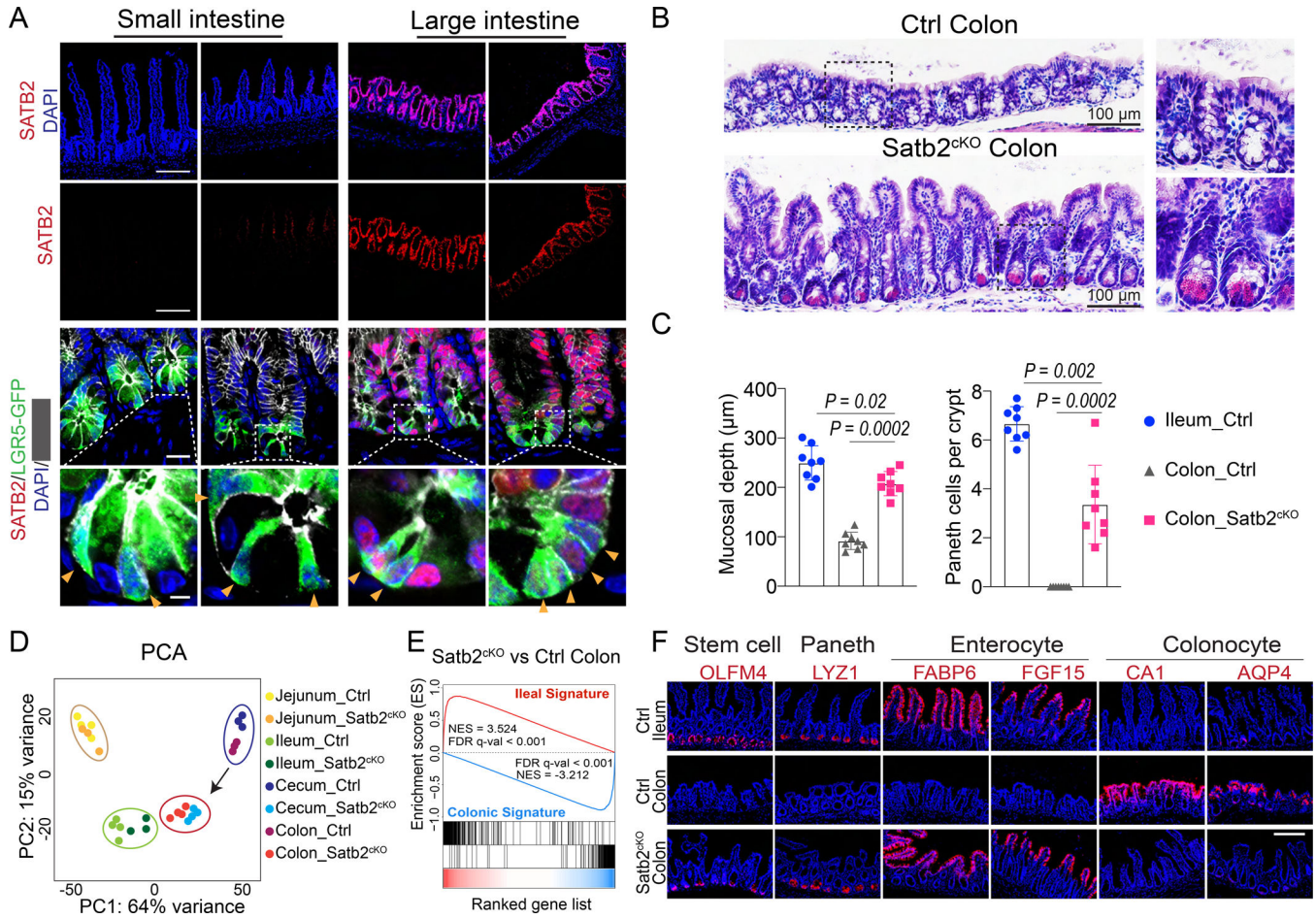


Figure 1. Conversion of large intestine mucosa to one that resembles ileal small intestine in Satb2^{cKO} mice.

A. SATB2 is expressed in adult murine large intestine epithelial cells including LGR5^{+} stem cells and absent in small intestine epithelium. ECAD: E-cadherin.

B, C. 30 days after intestinal deletion of *Satb2* from 2-month old *Vil-Cre^{ER};Satb2$^{f/f}$* (Satb2^{cKO}) mice, H&E staining showed the characteristic flat colonic glands were replaced by elongated villi-like glands, and many cells bearing Paneth morphology (pink-colored cells) appeared at the bottom of the glands. N = 8 mice. Mean ± S.D. P value by Mann Whitney U-test.

D, E. RNA-seq of intestinal epithelia revealed a shift of the large intestine transcriptomes (cecum and colon) in Satb2^{cKO} mice towards small intestine ileal transcriptomes by principal component analysis (PCA) (**D**) and Gene Set Enrichment Analysis (**E,** NES: normalized enrichment score; FDR: false discovery rate).

F. Immunofluorescence showed the appearance of OLFM4^{+} small intestine stem cells, LYZ1^{+} Paneth cells, FABP6^{+} and FGF15^{+} ileal enterocytes in Satb2^{cKO} colon, and concomitant disappearance of CA1^{+} and AQP4^{+} colonocytes.

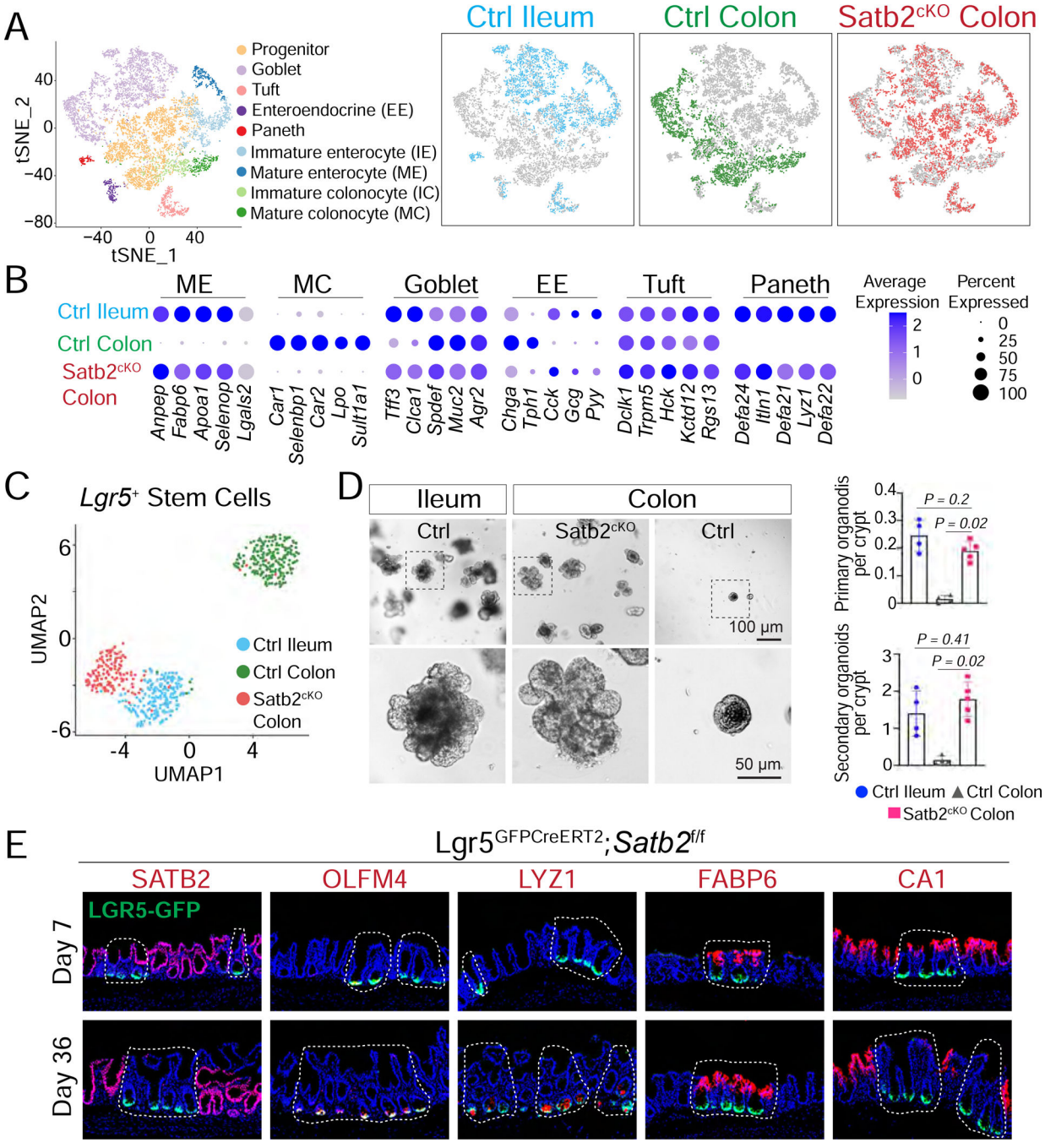


Figure 2. Conversion of LGR5^{+} colonic stem cells to ileal-like stem cells after SATB2 loss.

A. scRNA sequencing and post hoc annotation showed that a majority of cells in Satb2^{cKO} colon clustered with ileum (t-SNE plots, 3,912 cells from ileum, 3,627 cells from control colon, and 4,370 cells from Satb2^{cKO} colon). Satb2^{cKO} colonic sample harvested 30 days post TAM.

B. Dot plots of 30 representative genes of the major intestinal cell lineages.

C. UMAP visualization of 594 LGR5^{+} stem cells at G1/S phase. Satb2^{cKO} colonic stem cells cluster with ileal rather than colonic stem cells.

D. When grown in standard small intestine medium in 3D Matrigel, primary colonic glands from Satb2^{cKO} mice yielded branching organoids at a similar efficiency as control ileal glands, which can be further propagated into secondary organoids. In contrast, control colonic glands generated few small spheroids. N = 4, 5 mice, Mean ± S.D. P value by Mann Whitney U-test.

E. SATB2 deletion from colonic stem cells in *Lgr5*^{CreERGFP};*Satb2*$^{f/f}$ mice led to progressive conversion of colonic epithelium to ileum. The deleted clones were marked by crypt GFP expression. 7 days after tamoxifen treatment, SATB2 disappeared from the lower part of the glands but OLFM4 was activated only in some of the GFP^{+} colonic stem cells, indicating incomplete reprogramming at this stage. By day 36, the conversion of colonic stem cells appeared complete with OLFM4 expression in most of the GFP^{+} cells, presence of LYZ1^{+} Paneth cells, and replacement of CA1^{+} colonocytes by FABP6^{+} enterocytes.

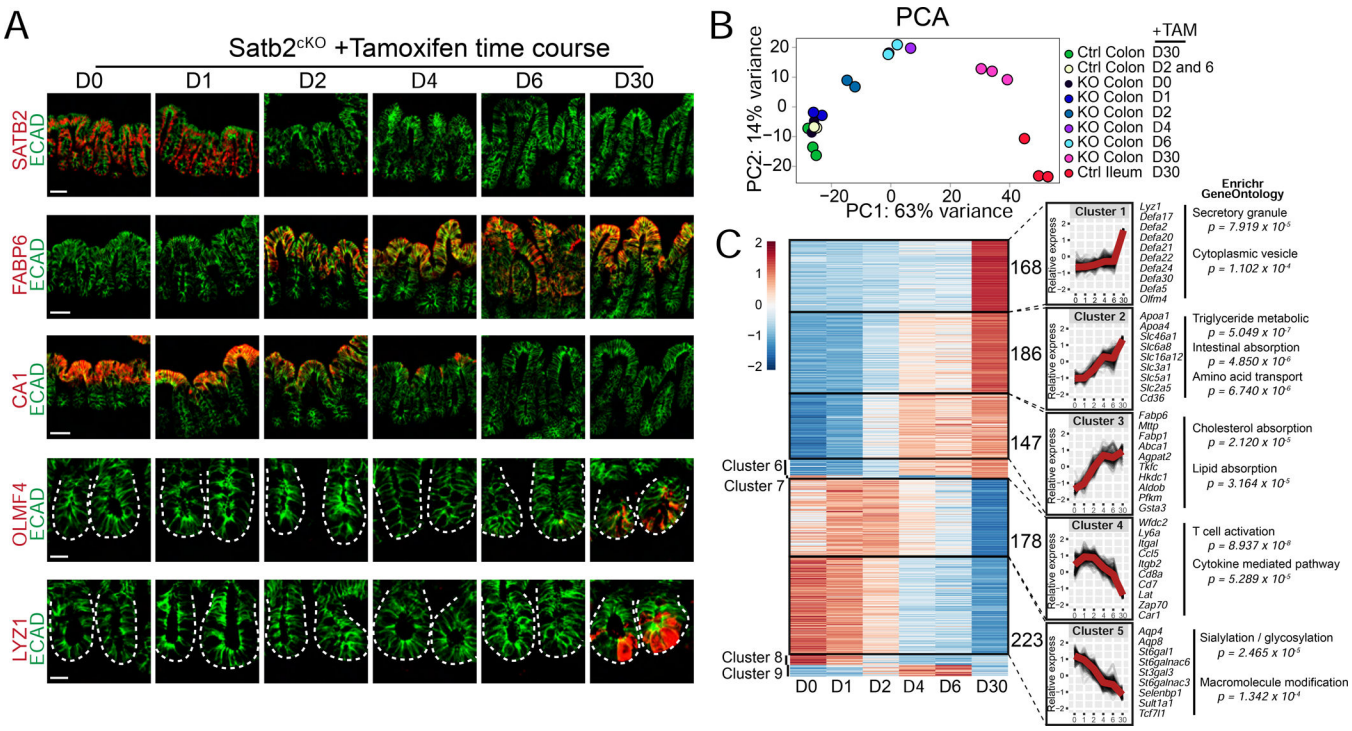


Figure 3. Rapid conversion of colonocytes to enterocytes after SATB2 loss.

A. A time course study of colonic mucosa after a single dose of TAM treatment in Satb2^{cKO} mice showed rapid activation of FABP6 and down-regulation of CA1 at day 2 and complete replacement of CA1^{+} cells by FABP6^{+} cells by day 6. OLFM4 and LYZ1 were not robustly activated until day 30.

B, **C**. Principal component analysis (PCA) and heatmap representation of time course RNA-seq data showed rapid activation of pathways typical of enterocytes and down-regulation of pathways characteristic of colonocytes. Paneth and stem cell genes were only strongly activated at day 30.

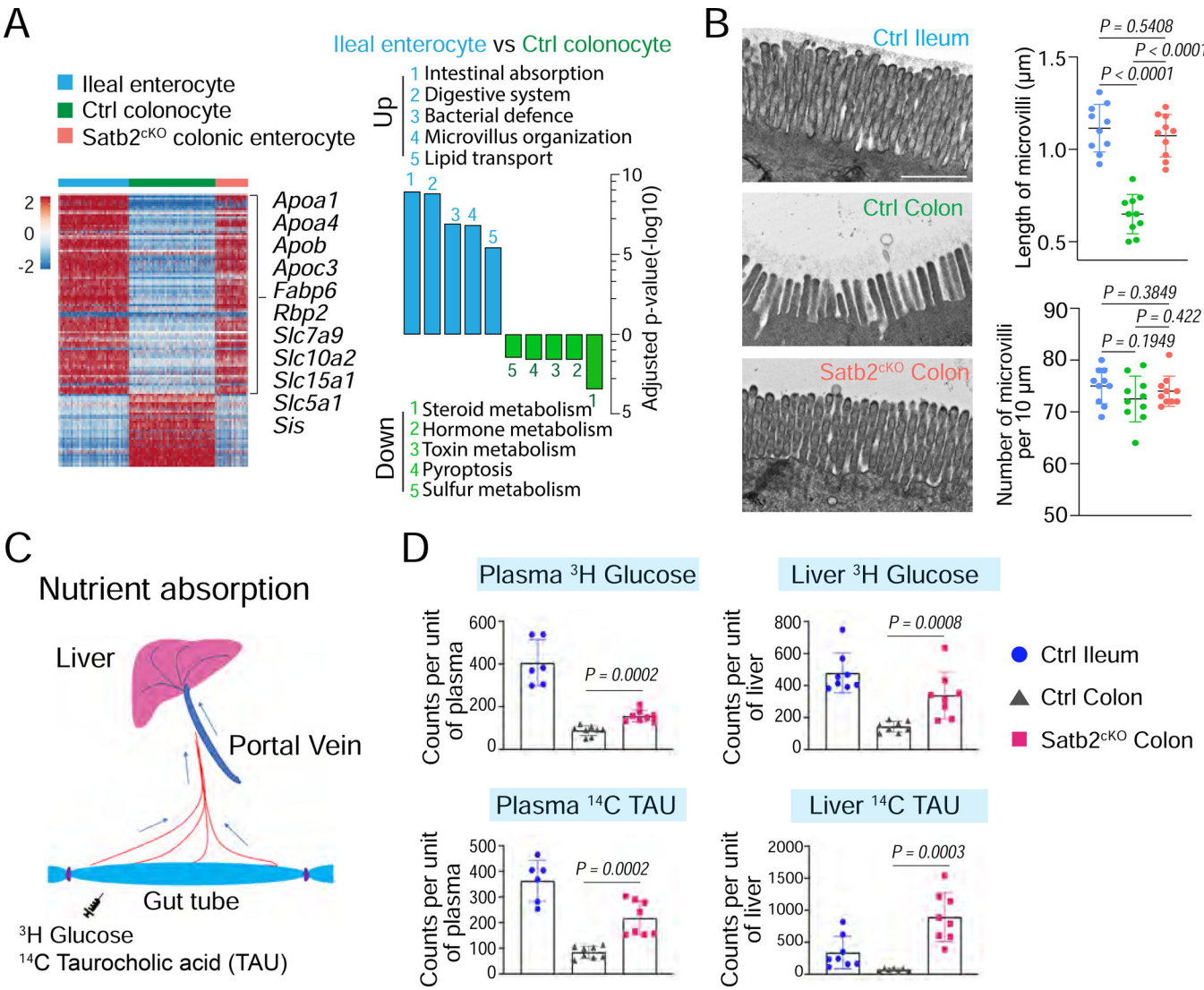


Figure 4. Generation of bona fide nutrient absorbing enterocytes in Satb2^{cKO} colon.

A. The scRNA profiles of Satb2^{cKO} colonic enterocytes closely resemble ileal enterocytes. Heatmap was plotted using the top 100 DEGs between ileal enterocytes and control colonocytes. Bar graph showed the top five differential GeneOntology pathways between ileal enterocytes and control colonocytes. Some of the nutrient transporters were highlighted in heatmap.

B. The microvilli length of Satb2^{cKO} enterocytes was significantly longer than that of control colon and comparable to ileal enterocytes. N = 10 randomly selected cells. Mean ± S.D. P value by Mann Whitney U-test.

C. Schematic diagram of the assay to measure glucose and taurocholic acid absorption and trans-epithelial transport into portal circulation. A segment of the ileum or colon was tied on both ends to create a pouch; radiolabeled chemicals were injected into the pouch to allow absorption and transport.

D. The amount of glucose and taurocholic acid being transported in portal vein plasma or deposited in the liver tissue after infusion into the Satb2^{cKO} colon is significantly higher

compared with infusions into the control colon, indicating enhanced absorptive functions in the colon of SATB2 null mice. N = 6–8 mice. Mean ± S.D. P value by Mann Whitney U-test.

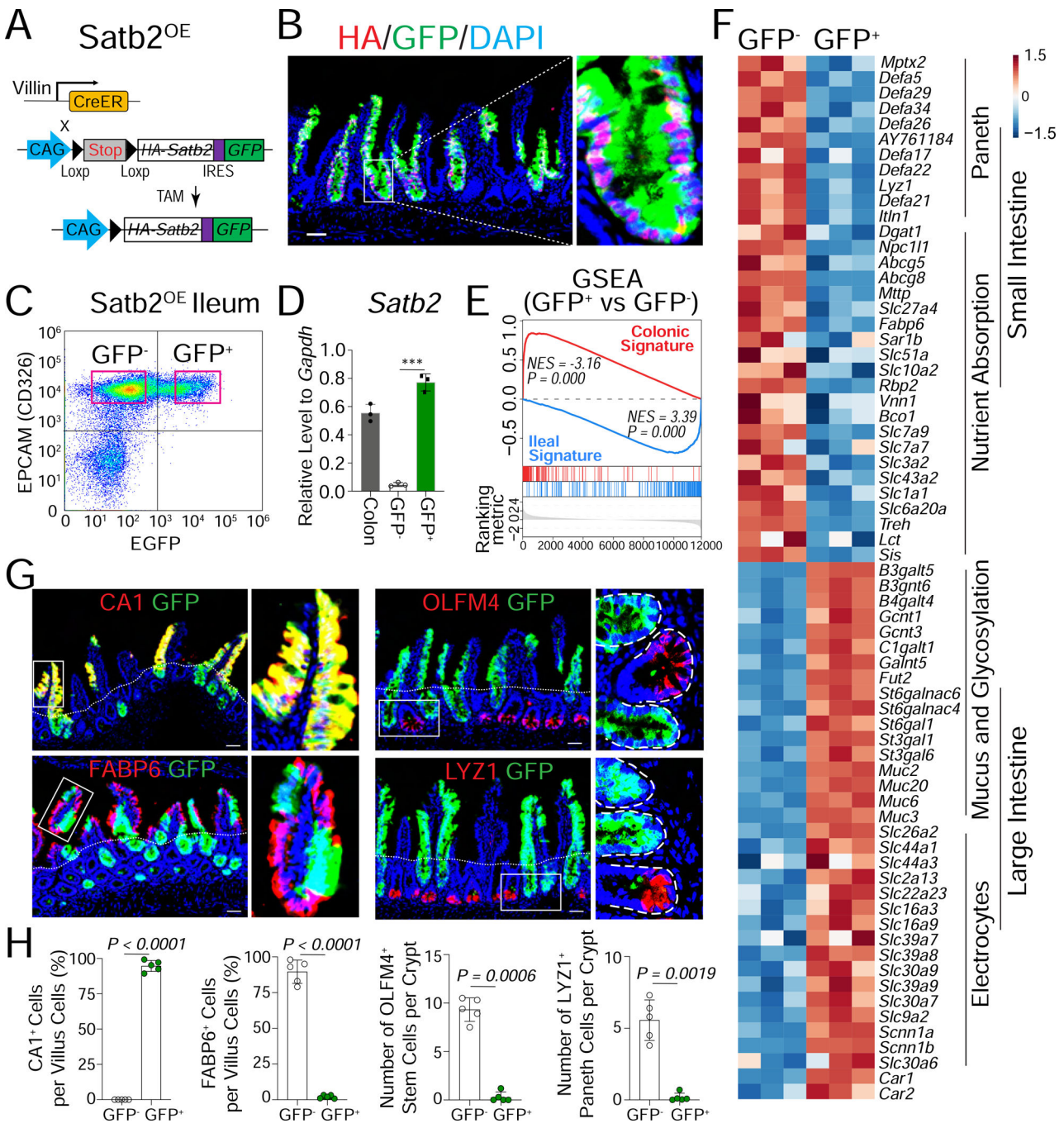


Figure 5. SATB2 confers colonic characteristics to adult ileum.

A, B. Construction of a *Satb2* transgenic mouse line (CAG^{Satb2GFP}). **A**. Diagram of the construct. Co-expression of murine SATB2 (with a HA epitope tag) and GFP is activated in the adult intestine after TAM treatment of *Vil-Cre^{ER};CAG^{Satb2GFP}* (Satb2^{OE}) mice. **B**. Immunostaining of ileal sections (30 days after TAM treatment of 2-month old mice) confirmed co-localization of HA tag and GFP.

C-F. Ectopic expression of *Satb2* activated colonic genes and suppressed ileal genes. **C**. A representative FACS plot of GFP and EPCAM in the purification of GFP^{+} and GFP^{-} ileal

epithelial cells. **D**. qPCR quantification of *Satb2* transcript levels in colon epithelial, GFP^+ and GFP^- ileal epithelial cells. N = 3 mice. *** $P < 0.001$. Mean ± S.D. Unpaired t-test. **E**. Gene Set Enrichment Analysis of transcriptomes from GFP^+ vs GFP^- cells showed an enrichment of colonic and depletion of ileal signature genes (NES: normalized enrichment score; P: Nominal P value). **F**. Heatmap of representative small and large intestine genes illustrating that GFP^+ ileal cells lost expression of enterocyte nutrient transporters and Paneth bacterial defense factors while gaining expression of electrolyte transporters and glycosylation enzymes characteristic of colonic function.

G, H. Ectopic SATB2 activated the colonic marker CA1 and suppressed OLFM4, LYZ1, and FABP6 in the ileum. **G**. Immunofluorescence staining of Ileum 30 days after TAM. White lines delineate villi and crypts. Dashed lines in magnified pictures outline crypts. **H**. Quantitation of $CA1^+$ and $FABP6^+$ cells among GFP^+ or GFP^- cells on the ileal villi. Numbers of $OLFM4^+$ stem cells and $LYZ1^+$ Paneth cells were quantified in GFP^+ and GFP^- crypts. N = 5 mice. Mean ± S.D. P value by paired t-test.

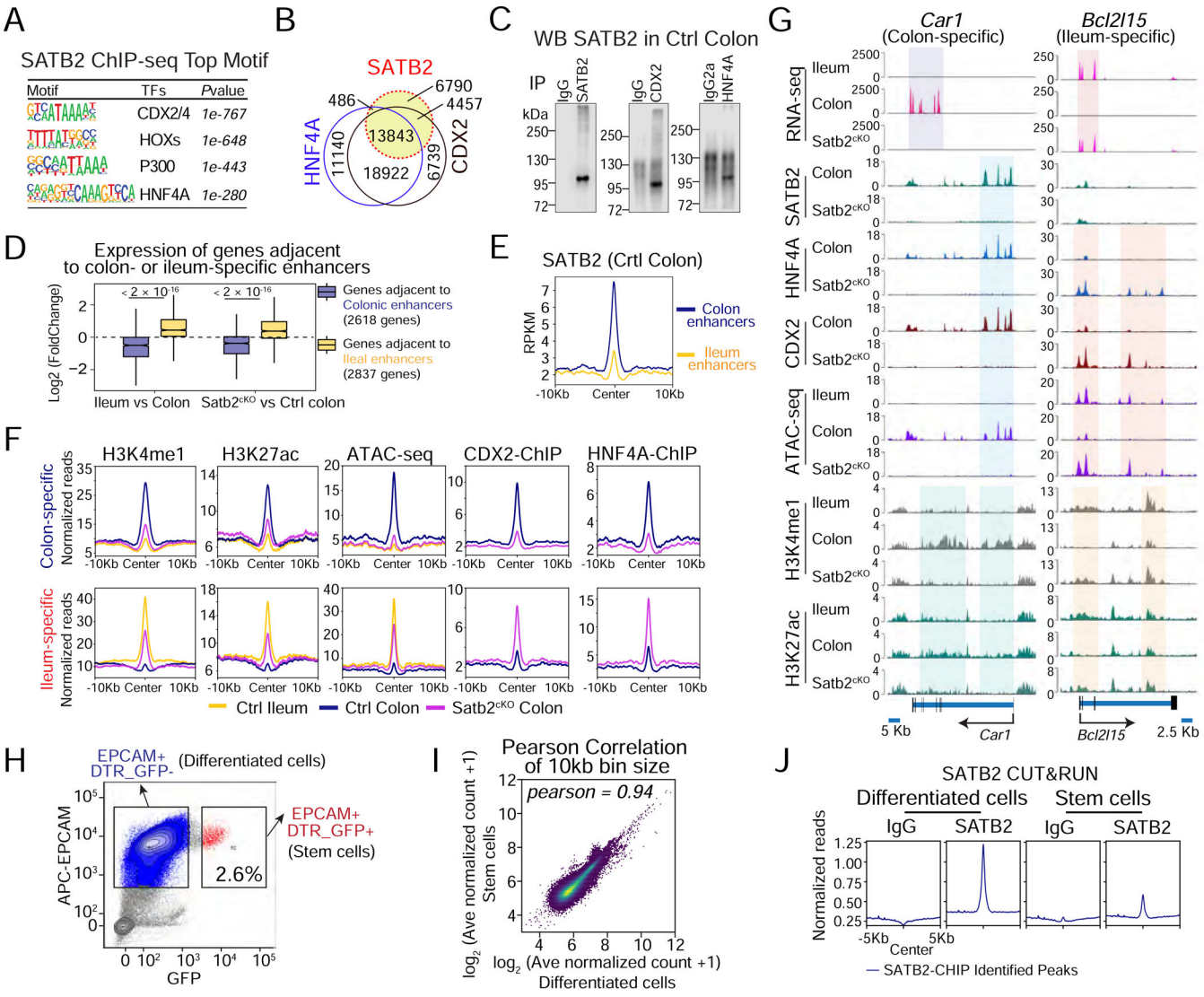


Figure 6. SATB2 regulates enhancer dynamics and binding of intestinal transcription factors CDX2 and HNF4A.

A-C. Extensive genomic co-binding of SATB2 with CDX2 and HNF4A in colonic epithelium. **A**. Top TF binding motifs enriched in SATB2 ChIP-seq sites (MACS $P < 1 \times 10^{-9}$) by HOMER in control colonic epithelium. Motifs were ranked by $-\log_{10}$ (P value). **B**. Venn diagram showed the overlaps among SATB2, CDX2 and HNF4A bound regions. Two biological replicates for each of the factors. **C**. CDX2 and HNF4A antibodies can pull down SATB2 proteins from primary colonic tissues.

D. Box-and-whisker plots representing relative gene expression changes. Genes adjacent to colonic enhancers (MAnorm $P < 0.01$; distance < 50kb; 2618 genes) were expressed at higher levels in control colon ($P < 2 \times 10^{-16}$) whereas genes adjacent to ileal enhancers (MAnorm $P < 0.01$; distance < 50kb; 2837 genes) were expressed at higher levels in ileum and Satb2^{cKO} colon ($P < 2 \times 10^{-16}$). Each box represents the median and interquartile range; whiskers extend to 1.5 times the interquartile range. P value by unpaired, two-sided Wilcoxon rank-sum test; N = 3 mice.

E. SATB2 ChIP-seq signals were enriched at colon-specific enhancers compared to ileum-specific enhancers in control colon ($P < 2 \times 10^{-16}$). Plot shown for a 20 kb window centered on each specific enhancer binding sites.

F. Inactivation of colon-specific enhancers and activation of ileum-specific enhancers in $Satb2^{cKO}$ colon. All plots were shown with a 20 kb window centered at colon-specific or ileum-specific enhancers for H3K4me1 and H3K27ac (CUT&RUN-seq), ATAC-seq, CDX2 and HNF4A ChIP in ileum, colon and $Satb2^{cKO}$ colon (N = 2 biological replicates).

G. Genome Browser tracks of RNA-seq, histone modifications (CUT&RUN-seq), and TF ChIP data at genomic loci of the colonic gene *Car1* and the ileal gene *Bcl2l15*. Regions with significant enhancer and TF binding changes among samples were highlighted.

H-J. SATB2 bound similar genomic loci in colonic $LGR5^{+}$ stem cells as non-stem cells. **H**. FACS plot of cell fractions enriched for stem cells ($EPCAM^{+}GFP^{+}$) or differentiated cells ($EPCAM^{+}GFP^{-}$) isolated from $LGR5^{DTRGFP}$ murine colon. **I**. Pearson correlation of SATB2 binding signals in stem vs differentiated cells by CUT&RUN-seq showed concordant binding patterns. **J**. SATB2 CUT&RUN binding profiles in a 10Kb window centered on SATB2 peaks identified by ChIP-seq.

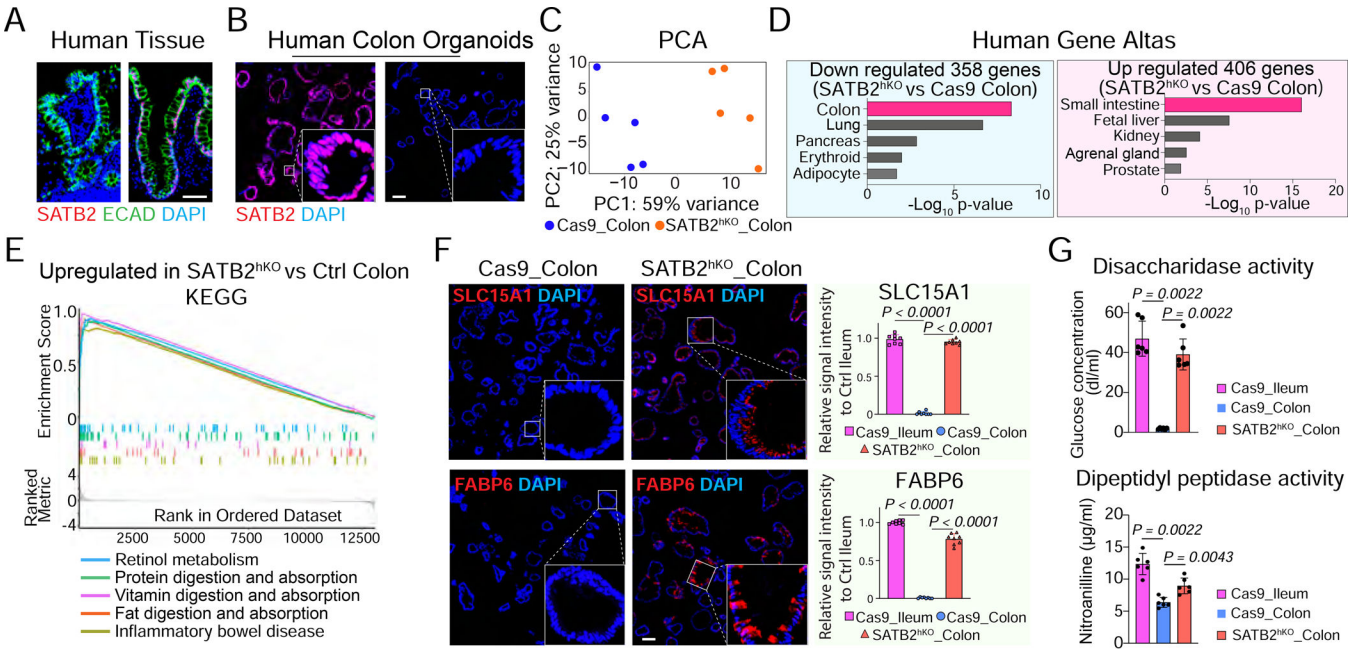


Figure 7. Colonic to ileal plasticity after SATB2 loss in human colonic organoids.

A. SATB2 was expressed in ECAD^{+} (E-CADHERIN) epithelial cells of human colon but not ileum.

B. Representative images of SATB2 expression in one of the primary human organoid lines (#87) and its absence after CRISPR-mediated deletion. sgRNA: single guide RNA.

C-E. Transcriptomes of *SATB2* deleted (SATB2^{hKO}) colonic organoids shifted towards ileum, as seen in the PCA plot (**C**), tissue enrichment (**D**), and up-regulated KEGG pathways (**E**).

F. The ileal makers SLC15A1 and FABP6 were activated in colonic organoids after SATB2 loss and localized to the luminal epithelial side and the cytoplasm, respectively. Relative signal intensity was calculated by comparison with control ileal organoids (Fig. S6G). N = 8 samples (4 biological replicates each in 2 independent experiments). Mean ± S.D. P value by Mann Whitney U-test.

G. Significant activities of the small intestine disaccharidase and dipeptide peptidase were detected in SATB2^{hKO} colonic organoids. N = 6 samples (3 biological replicates each in 2 independent experiments). Mean ± S.D. P value by Mann Whitney U-test.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti SATB2 (For CUT&RUN)	Abcam	Cat# Ab34735; RRID: AB_2301417
Rabbit monoclonal anti SATB2	Abcam	Cat# Ab92446; RRID: AB_10563678
Rabbit monoclonal anti CDX2	Cell Signaling Technology	Cat# 12306; RRID: AB_2797879
Mouse monoclonal anti HNF4A	Abcam	Cat# Ab41898; RRID: AB_732976
Rabbit monoclonal anti H3K27Ac	Millipore Sigma	Cat# MABE647; RRID: AB_2893037
Rabbit monoclonal anti H3K4Me1	Cell Signaling Technology	Cat# 5326; RRID: AB_10695148
Rabbit monoclonal anti CA1	Abcam	Cat# Ab108367; RRID: AB_10863424
Rabbit polyclonal anti FABP6	Abcam	Cat# Ab91184; RRID: AB_10563324
Goat polyclonal anti FABP6	R&D System	Cat# AF3880; RRID: AB_2100347
Rabbit polyclonal anti KI67	Abcam	Cat# Ab15580; RRID: AB_443209
Rabbit polyclonal anti SLC15A1	Abcam	Cat# Ab78020; RRID: AB_1566730
Rabbit polyclonal anti RBP2	Abcam	Cat# Ab180494; RRID:AB_2893036
Rabbit monoclonal anti OLFM4	Cell Signaling Technology	Cat# 39141; RRID: AB_2650511
Rabbit monoclonal anti AQP4	Cell Signaling Technology	Cat# 59678; RRID: AB_2799571
Rabbit monoclonal anti HA-Tag	Cell Signaling Technology	Cat# 3724; RRID: AB_1549585
Rabbit polyclonal anti cleaved CASPASE 3	Cell Signaling Technology	Cat# 9661; RRID: AB_2341188
Rabbit polyclonal anti CEACAM1	Cell Signaling Technology	Cat# 44464; RRID: AB_2799265
Rabbit polyclonal anti LYSOZYME1	Dako	Cat# A0099; RRID: AB_2341230
Rat anti E-CARDHERIN	Life technologies	Cat# 131900; RRID: AB_2533005
Rabbit polyclonal anti MUC2	Invitrogen	Cat# PA5–79702; RRID: AB_2746817
Sheep polyclonal anti FGF15	Invitrogen	Cat# PA5–47870; RRID: AB_2610314
APC Rat anti-mouse CD326 (Ep-CAM)	BD biosciences	Cat# 563478; RRID: AB_2738234
Alexa Fluor® 488 Rat anti-mouse CD31	3D biosciences	Cat# 563607; RRID: AB_2738312
APC-Cy7 Rat Anti-mouse CD45	BD biosciences	Cat# 561037; RRID: AB_10563075
APC Rat IgG2a Control	3D biosciences	Cat# 554690; RRID: AB_10056902
Alexa Fluor® 488 Rat IgG2a Control	BD biosciences	Cat# 557676; RRID: AB_396787
APC-Cy7 Rat IgG2b Control	3D biosciences	Cat# 552773; RRID: AB_394459
Biological samples		
Prep80 (Human Ascending Colon Organoid)	University of Michigan	29Y, Female
Prep83 (Human Ascending Colon Organoid)	University of Michigan	45Y, Female
Prep87 (Human Ascending Colon Organoid)	University of Michigan	21Y, Male
Prep88 (Human Ascending Colon Organoid)	University of Michigan	33Y, Female
Prep89 (Human Ascending Colon Organoid)	University of Michigan	55Y, Male
Z010A (Ileum Biopsy)	Weill Cornell Medicine	27Y, N/A
Z011A (Ileum Biopsy)	Weill Cornell Medicine	27Y, N/A
Z010B (Ascending Colon Biopsy)	Weill Cornell Medicine	27Y, N/A
Z011B (Ascending Colon Biopsy)	Weill Cornell Medicine	27Y, N/A

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
TrypLE Express Enzyme	Thermo Fisher Scientific	Cat# 12604021
VECTASHIELD PLUS Antifade Mounting Medium	Vector Laboratories	Cat# H-1900
Goat anti-Rabbit IgG ImmPRESS Secondary	Vector Laboratories	Cat# MP-7451
BLOXALL Endogenous Peroxidase and Alkaline Phosphatase Blocking Solution	Vector Laboratories	Cat# SP-6000
Hematoxylin QS	Vector Laboratories	Cat# H-3404
ImmPACT DAB EqV Peroxidase (HRP) Substrate	Vector Laboratories	Cat# SK-4103
Corning Matrigel Matrix Phenol Red Free	VWR	Cat# 47743–722
Advanced DMEM/F-12	Thermo Fisher Scientific	Cat# 12634028
Applied Biosystems PowerUp SYBR	Thermo Fisher Scientific	Cat# A25743
Cell Recovery Solution, Corning	VWR	Cat# 47743–696
Recombinant Human EGF Protein, CF	R&D Systems	Cat# 236-EG-01M
CHIR99021, > 98%(hplc)	Sigma-Aldrich	Cat# SML1046
Recombinant Murine Noggin	PeproTech	Cat# 250–38
Recombinant Murine R-Spondin-1	PeproTech	Cat# 315–32
Primocin	Invivogen	Cat# ant-pm-1
A 83–01	Cayman Chemical	Cat# 9001799
Any KD Mini-PROTEAN Gel	Bio-Rad	Cat# 4568126
4x Laemmli Sample Buffer	Bio-Rad	Cat# 1610747
Tamoxifen	Sigma-Aldrich	Cat# T5648
Corn Oil	Sigma-Aldrich	Cat# C8267
Falcon Cell Stainer 70um	Corning	Cat# 352350
Falcon Cell Stainer 40um	Corning	Cat# 352340
Niacinamide (Nicotinamide)	Sigma-Aldrich	Cat# N5535
Gastrin 1 human	Sigma-Aldrich	Cat# G9020
Y-27632, Dihydrochloride Salt	LC Laboratories	Cat# Y-5301
N-2 supplement (100X)	Thermo Fisher Scientific	Cat# 17502048
B-27 Supplement (50X)	Thermo Fisher Scientific	Cat# 17504044
HEPES	Thermo Fisher Scientific	Cat# 15-630-080
GlutaMAX Supplement	Thermo Fisher Scientific	Cat# 35050061
N-Acetyl-L-Cysteine	Sigma-Aldrich	Cat# A9165
Recombinant Human FGF-basic (FGF-2)	PeproTech	Cat# 100–18B
Recombinant Human IGF-1	BioLegend	Cat# 590908
DAPT	Cayman Chemical	Cat# 13197
SB 202190	Tocris	Cat# 1264
L-161,982	Cayman Chemical	Cat# 10011565
Protein A/G Magnetic Beads	Thermo Fisher Scientific	Cat# 88803
Fetal Bovine Serum (FBS)	R&D Systems	Cat# S11150H
Normal Donkey Serum	Jackson ImmunoResearch	Cat# 017-000-001

REAGENT or RESOURCE	SOURCE	IDENTIFIER
AmPure XP Beads	Beckman	Cat# A63880
Lipofectamine 3000	Thermo Fisher Scientific	Cat# L3000015
Alt-R S.p. Cas9 Nuclease V3	IDT	Cat# 1081058
UltraPure EDTA	Thermo Fisher Scientific	Cat# 15-575-020
GIBCO DMEM	Thermo Fisher Scientific	Cat# 11-965-118
Collagenase type IV	Worthington	Cat# LS004188
Deoxyribonuclease I (Dnase 1)	Worthington	Cat# LS002007
Ham's F12K Medium	Thermo Fisher Scientific	Cat# 21-127-022
Cell Strainer 100μm	VWR	Cat# 10199–658
Puromycin	Sigma-Aldrich	Cat# P8833
T7 Endonuclease 1	New England Biolabs	Cat# M0302S
Phusion High-Fidelity DNA polymerase	New England Biolabs	Cat# M0530S
Digitonin	Promega	Cat# G9441
Tween-20,10% non-ionic solution	Sigma-Aldrich	Cat# 11332465001
IGEPAL® CA-630 (Chemically indistinguishable from NP40)	Sigma-Aldrich	Cat# 13021
Magnesium chloride (MgCl2) solution, BioUltra	Sigma-Aldrich	Cat# 68475
Sodium chloride (NaCl) solution, BioReagent	Sigma-Aldrich	Cat# S5150
PBS without calcium magnesium	Cytiva	Cat# SH30028
Disuccinimidyl Glutarate (DSG)	Thermo Fisher Scientific	Cat# 20593
Formaldehyde Solution	Sigma-Aldrich	Cat# F8775
N-Lauroylsarcosine sodium salt solution (Sarkosyl)	Sigma-Aldrich	Cat# L7414
Halt Protease and Phosphatase Inhibitor Cocktails	Thermo Fisher Scientific	Cat# 78443
Sodium Dodecyl Sulfate (SDS) UltraPure	Invitrogen	Cat# 15553–035
Triton X-100	Sigma-Aldrich	Cat# T8787
Tris-HCl	Sigma-Aldrich	Cat# T2194
Proteinase K	Thermo Fisher Scientific	Cat# 26160
14C-Taurocholic acid	American Radiolabeled Chemicals	Cat# ARC3489
3H-Glucose	American Radiolabeled Chemicals	Cat# ART1353
D-(+)-Glucose, ≥ 99.5% (GC) (dextrose)	Sigma-Aldrich	Cat# G8270
Gly-Pro-p-nitroanilide hydrochloride	Sigma-Aldrich	Cat# G0513
Sucrose	Sigma-Aldrich	Cat# S0389
4-Nitroanilide	Sigma-Aldrich	Cat# 185310
DL-Dithiothreitol (DTT)	Sigma-Aldrich	Cat# 43815
Critical commercial assays		
DNA Clean & Concentrator-5	Zymo Research	Cat# D4014
Alcian Blue (pH2.5) Stain Kit	Vector Laboratories	Cat# H-3501
ImmPRESS Duet Double Staining Polymer Kit	Vector Laboratories	Cat# MP-7724
Click-iT Edu Cell Proliferation Kit for Image, Alex Fluor 555 dye	Thermo Fisher Scientific	Cat# C10338

REAGENT or RESOURCE	SOURCE	IDENTIFIER
MinElute PCR Purification Kit	QIAGEN	Cat# 28004
RNeasy Plus Mini Kit	QIAGEN	Cat# 74136
RNeasy Plus Micro Kit	QIAGEN	Cat# 74034
ThruPLEX DNA-seq 12S Kit	Takara Bio	Cat# R400428
SMARTer ThruPLEX DNA-seq 48S Kit	Takara Bio	Cat# R400427
ThruPLEX DNA-seq Kit	Takara Bio	Cat# R400665
E.Z.N.A tissue DNA kit	Omega Bio_Tek	Cat# D3396–02
Kapa Hifi PCR Kits	Fisher Scientific	Cat# NC0580933
Illumina Tagment DNA TDE1 Enzyme and Buffer Kits	Illumina	Cat# 20034197
Nebnext® High-Fidelity 2X PCR Master Mix	New England Biolabs	Cat# M0541S
Dead Cell Removal Kit	Miltenyi Biotec	Cat# 130-090-101
Glucose Colorimetric Assay Kit	Cayman Chemical	Cat# LS-K221
TRIzol Plus RNA Purification Kit	Thermo Fisher Scientific	Cat# 12183555
RNase-Free DNase Set (50)	QIAGEN	Cat# 79254
Deposited data		
ATAC-seq of Intestine Epithelia (Ileum and Colon) from Satb2-deficient and WT mice	This paper	GEO: GSE148690
scRNA-Seq of Intestine Epithelia (Ileum and Colon) from Satb2-deficient and WT mice	This paper	GEO: GSE148693
ChIP-seq of Colonic Intestine Epithelium from Satb2-deficient and WT mice	This paper	GEO: GSE167287
CUT&RUN Analysis of Satb2 Binding in Colonic Stem Cells and Differentiated Cells	This paper	GEO: GSE180029
Enhancer Analysis (CUT&RUN) of Intestine Epithelium from Satb2-deficient and WT mice	This paper	GEO: GSE167289
Bulk RNA-seq of Intestine Epithelia (Jejunum, Ileum, Cecum and Colon) from Satb2-deficient and WT mice	This paper	GEO: GSE148692
Bulk RNA-seq (> 20M reads) of Intestine Epithelium (Ileum and Colon) from Satb2- deficient and WT mice	This paper	GEO: GSE167284
Bulk RNA-seq of Early Time Course Colonic Epithelia from Satb2-deficient and WT mice	This paper	GEO: GSE180023
Bulk RNA-seq of 3D Cultured Intestinal Organoids (Ileum and Colon) from Satb2-deficient and WT mice	This paper	GEO: GSE167281
Bulk RNA-seq of Epitope Expressed SATB2 in Mouse Ileum Epithelium	This paper	GEO: GSE167282
Bulk RNA-seq of Mouse Colonic Organoids: SATB2 and FOXD2KO CRISPR Knockout	This paper	GEO: GSE167283
Bulk RNA-seq of CRISPR SATB2 Knockout Human Colonic Organoids	This paper	GEO: GSE167285
Bulk RNA-seq of Human Duodenal and Colonic Organoids	This paper	GEO: GSE167286
Bulk RNA-seq of Colonic Epithelia from EED and EED/Satb2 Double Knockout Mice	This paper	GEO: GSE180013
Bulk RNA-seq of Mouse Duodenal and	Murata et al., 2020;	GEO: GSE130822; GEO: GSE71713
Colonic Stem Cells	Jadhav et al., 2016	
ATAC-seq of Embryonic Midgut Epithelia from WT mice	Banerjee et al., 2018	GEO: GSE115541

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ATAC-seq of Embryonic Hindgut Epithelia from WT mice	This paper	GEO: GSE180037
Experimental models: Organisms/strains		
Mouse strain: Satb2$^{loxp/loxp}$	Dobreva et al., 2006	gift from Dr. Jeffrey Macklis (Harvard University)
Mouse strain: CAG^{SATB2GFP}	This paper	N/A
Mouse strain: Vil-Cre^{ERT2}	el Marjou et al., 2004	gift from Sylvie Robine (Institut Pasteur)
Mouse strain: Lgr5^{GFPCreER}	Barker et al., 2007	Jackson lab (Stock No: 008875)
Mouse strain: Lgr5^{DTRGFP}	Tian et al., 2011	Genentech (Dr. Frederic de Sauvage)
Mouse strain: Eed$^{loxp/loxp}$	Jadhav et al., 2016; Xie et al., 2014	provided by Dr. Ramesh Shivdasani of Dana-Farber Cancer Institute
Oligonucleotides		
Table S7	This paper	N/A
Software and algorithms		
GraphPad Prism 9	https://www.graphpad.com	v9.1.2 (225); RRID: SCR_002798
ImageJ	https://imagej.nih.gov/ij/	v1.51 (100); RRID: SCR_003070
ImageJ (Fiji)	https://imagej.net/software/fiji/	v2.1.0/1.53c; RRID:SCR_002285
Adobe Photoshop and Illustrator	https://www.adobe.com	RRID: SCR_014199;
RRID: SCR_010279		
QuantStudio Design&Analysis Software	https://www.thermofisher.com/us/en/home.html	v1.4.3
Image Studio Digits	https://www.licor.com/bio/image-studio/	v5.2; RRID: SCR_015795
FlowJo	https://www.flowjo.com/	RRID:SCR_008520
Enrichr	https://maayanlab.doud/Enrichr/	RRID:SCR_001575
TapeStation	https://www.agilent.com/	vA.02.02; RRID:SCR_018435
Easeq	https://easeq.net/	N/A
CaseVeiwer	https://www.3dhistech.com/solutions/caseviewer/	v2.4
ZEN	https://www.zeiss.com/corporate/int/home.html	v2.3; RRID:SCR_018163
R-Studio	https://www.rstudio.com/	v1.4.1717
R	https://www.r-project.org/	v4.0.5
DESeq2	Love et al., 2014	1.28.1; RRID:SCR_015687
STAR	Dobin et al., 2013	v2.7.5b; RRID:SCR_016533
featureCounts	Liao et al., 2014	v2.0.1; RRID:SCR_012919
Clusterprofiler	Yu et al., 2012; (Wu, 2021)	v3.16.1; RRID:SCR_016884
Limma	Ritchie et al., 2015	v3.44.3; RRID:SCR_010943
GSEA	Subramanian et al., 2005	v4.1.0; RRID:SCR_003199

REAGENT or RESOURCE	SOURCE	IDENTIFIER
CellRanger	https://support.10xgenomics.com/	V3.1.0; RRID:SCR_021002
Seurat package	Butler et al., 2018	v3.2.0; RRID:SCR_016341
Deeptools	Ramírez et al., 2014	v3.4.3; RRID:SCR_016366
Bowtie2	Langmead and Salzberg, 2012	v2.4.1; RRID:SCR_016368
Samtools	Li et al., 2009	v1.9; RRID:SCR_002105
Macs2	Feng et al., 2012	v2.2.7; RRID:SCR_013291
Manorm	Shao et al., 2012	v1.1.4; RRID:SCR_010869
HOMER	Heinz et al., 2010	v4.11; RRID:SCR_010881
Recombinant DNA		
pR26CAG/GFP Dest	Chu et al., 2016	Addgene #74281
LentiCRISPRv2	Sanjana et al., 2014	Addgene #52961