Supplemental Methods

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contract, D. Brent Polk (dpolk@ucsd.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Bulk RNA sequencing data shown in Figure S3 are available on the Gene Expression Omnibus (GEO), with accession number GSE168053: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE168053. Single-cell RNA sequencing data shown in Figs. 3,4,6 are available on GEO, with accession number GSE168033: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE168033. Analysis scripts for single-cell sequencing are available on GitHub: https://github.com/stalepig/2021-SNEC-single-cell. Imaging processing routines for images obtained via the Deep Mucosal Imaging (DMI) pipeline are available on GitHub: github.com/stalepig/deep-mucosal-imaging. All other datasets generated and analyzed during the current study are available from the lead contact on reasonable request.

❖

Experimental model and subject details

Mice

All animals were maintained ethically and humanely at Children's Hospital Los Angeles (CHLA). The Institutional Animal Care and Use Committee of CHLA approved the study under internal protocol #288. Sox2::CreER (stock 017593), Rosa26::Confetti (stock 013731), Rosa26::LSL-DTA (stock 009669), Rosa26::mTmG (stock 007676), Rosa26::tdTomato (stock 007914), Lgr5::EGFP-IRES-CreER (stock 008875), and C57Bl6/J (stock 000664) mice were purchased from Jackson Laboratory. Vil1::Cre mice[1]

were a kind gift from Sylvie Robine and Daniel Louvard (Institut Curie). Krt7::CreER mice[2] were generously provided by Jianwen Que (Columbia). Krt20::CreER mice[3] were donated by Andrew McMahon (University of Southern California). Lgr5::DTR-EGFP mice[4] were kindly donated by Frederic de Sauvage (Genentech).

Mice were housed in the Animal Care Facility at CHLA. The mice were provided with standard bedding and airflow. Irradiated food pellets and drinking water supplied from a valve were provided for consumption ad libitum. The housing room was kept on a strict 12 h light/dark cycle and maintained at 20 ºC. Littermate mice were separated by sex at 3 wks-old and co-housed to a maximum of 5 mice per cage. Lineage tracing and immunolabeling experiments were performed on both male and female transgenic mice. For sequencing studies, cohorts of co-housed 6-wk-old male wildtype mice (C57Bl/6J) were ordered from Jackson Laboratory and allowed to habituate to the animal room for 3 wks prior to experimentation.

Retrospective analyses of Il10-/- mice were performed from archived slides of a previous study[5].

❖

Method details

Dissection and fixation

For experimental studies, mice were anesthetized with isoflurane and euthanized via cervical dislocation. The colon and adjacent anal ring were removed, opened longitudinally, and washed of fecal material. Tissue was immediately flattened against blotting paper and fixed overnight in methanol-free 4% cold paraformaldehyde/phosphate-buffered saline (PBS). The next day, the mucosal layer was finely freed under a stereomicroscope from serosal/muscular layers and additionally processed.

Deep Mucosal Imaging (DMI)

The pipeline for mounting the specimen, imaging, and in silico reconstruction of images was as previously published[6, 7]. However, the clearing reagents used are different. Briefly, fixed tissues were washed with PBS, incubated with 20% sucrose/PBS and frozen. Tissues were thawed and incubated with CUBIC-1[8] clearing agent supplemented with 1:1,000 dilution of a 4% w/v methyl green stock[9] that was purified of

contaminating crystal violet through multiple phase separations in chloroform. Tissues were cleared for 1-7 d prior to imaging. For imaging, tissues were embedded between two #1 glass coverslips and mounted with the used clearing solution. Tiled images were captured on a Zeiss LSM 700 confocal microscope with either 10X/0.45NA or 20X/0.8 air objectives. The distal 1-3 cm of mouse colonic mucosa and adjacent anal tissue were captured in overnight scans.

Whole-mount antibody staining

For procedures not requiring the preservation of fluorescent proteins, the iDISCO procedure[10] was modified for mouse colon. Fixed tissue was washed with PBS and permeabilized in an ascending (20-100%) methanol/PBS series, and then incubated overnight in 5% hydrogen peroxide/methanol at 4ºC. Tissue was rehydrated in a descending methanol/PBS series, washed successively with PBS and 0.3% Triton X-100 in PBS (PBTx), and incubated for 18 h at 37ºC in 20% DMSO/0.3M glycine in PBTx and for 24 h at 37ºC in 10% DMSO/5% normal horse serum prepared in PBTx. Primary antibody was diluted in PBTx supplemented with 10 µg/ml heparin, 3% normal horse serum, and 5% DMSO. Samples were stained with primary antibody at 37ºC for 4 d. Washes were performed with PBTx with 10 µg/ml heparin. Secondary antibodies conjugated with Alexa dyes (Thermo Fisher Scientific) were diluted in PBTx with 10 µg/ml heparin, 3% normal horse serum, and a 1:10,000 dilution of 4% w/v methyl green stock. Secondary staining was performed for 3 d at 37ºC. Stained samples were washed and chemically cleared in CUBIC-2 for 1-7 d and mounted, imaged, and reconstructed using the DMI pipeline.

Antibodies used in the iDISCO protocol were rabbit anti-cleaved caspase-3 (Cell Signaling Technology 9661, 1:100), rabbit anti-CHGA (Immunostar 20085, 1:200), and rat anti-KI-67 (eBioscience 41-5698, clone SolA15, 1:100).

For procedures requiring the preservation of fluorescent proteins, we first cleared the tissue using CUBIC-1 for 3 d at room temperature. The tissue was washed with PBS and blocked at room temperature overnight in 0.3% PBTx supplemented to 5% with horse serum. Primary antibodies were diluted in blocking solution and incubated with the tissue for 3 d at room temperature. Tissue was washed with PBS. Secondary antibodies were diluted in 0.3% PBTx and incubated with tissue for 3 d at room temperature. After a final series of washes with PBS, tissue was cleared with CUBIC-2 for 1-4 d, mounted, and directly imaged in the clearing solution. Primary antibodies were directed against KRT7 (Abcam ab181598, 1:200), KRT14 (Abcam ab206100, 1:500), and

KRT10 (Proteintech 18343-1-AP, 1:100).

In situ hybridization

Chromogenic in situ hybridization was performed using the RNAScope 2.5 HD Assay – Brown (ACDBio/Biotechne) according to the manufacturer's instructions. Paraffin-embedded sections were hybridized with probes targeted to murine *Krt17* or *Ass1*. Signal was revealed with DAB (Vector Labs), and the tissue was counterstained with hematoxylin and blued with ammonium hydroxide.

Fluorescent in situ hybridization was performed on paraffin-embedded sections with hybridization chain reaction (HCR) technology (Molecular Instruments) according to the manufacturer's recommendations. Probes were directed against *Krt7*. After staining, samples were mounted in Fluoro-Gel (Electron Microscopy Sciences) prior to confocal imaging.

Histochemistry

Immunohistochemical experiments on paraformaldehyde-fixed, paraffin-embedded or frozen sections was performed according to standard procedures. For paraffin-embedded sections, a citric acid/heat antigen retrieval step was performed prior to antibody incubation. Endogenous peroxidases were quenched with a 30 min incubation in 3% hydrogen peroxide in PBS. Primary antibodies used were rat anti-CDH1 (R&D Systems FAB7481F, paraffin, 1:300), rabbit anti-P63 (Proteintech 12143-1-AP, paraffin, 1:200), rabbit anti-SOX2 (Abcam ab97959, paraffin, 1:1500), rabbit anti-CLDN1 (Thermo Fisher Scientific 71-7800, paraffin, 1:1000), rabbit anti-LOR (Biolegend 905101, 1:5000), mouse anti-MUC4 (Thermo Fisher Scientific 35-4900, 1:2000), rabbit anti-KRT17 (Sigma-Aldrich AV41733, 1:3000), rabbit anti-COL17A1 (Abcam ab184996, 1:250), rabbit anti-GSTO1 (Novus NBP1-33763, 1:100), rabbit anti-KRT8 (Abcam ab53280, 1:1000), rabbit anti-IVL (Biolegend 924401, 1:10000), and rabbit anti-KRT1 (Biolegend 905602, 1:1000), with HRP-linked secondaries. Chromogen development was performed using a VIP kit (Vector Labs), and slides were counterstained with methyl green.

Mouse endoscopy

Mice were sedated with ketamine and xylazine. Distal colonoscopy was

performed with a KARL STORZ Veterinary Endoscope (Karl Storz, Tuttlingen, Germany).

DSS colitis

Dextran sulfate sodium (DSS)-induced colitis was administered as described previously[6]. Mice were trained to drink solely from a water bottle provided to the cage for 3-7 d prior to administration of DSS (38-50 kDa). On exp d 0, the contents of the water bottle were replaced with dextran sulfate sodium supplemented to de-ionized drinking water to a final concentration of 3% w/v. After 6 d, the water bottle was removed, and the cage was returned to the housing room's water valve supply.

Lineage tracing experiments

Induction of CreER proteins for lineage tracing was induced by a single injection of 2 mg tamoxifen. Sox2::CreER;Rosa26::Confetti lineage tracing for clonal analysis was initiated with a single intraperitoneal injection of 0.5 mg tamoxifen. Low-dose lineage tracing in Krt7::CreER;Rosa26::mTmG mice was initiated with 0.02 mg tamoxifen.

Target cell ablation experiments

Systemic activation of DTA expression was induced with 2 mg tamoxifen administered as in lineage tracing. Local or topical activation of DTA in the distal colon/anus was induced via enema. 4-hydroxytamoxifen (4-OHT) was dissolved in a heated mixture of 65%/35% Witepsol H-15/corn oil at 0.25 mg/ml and cooled to form a suppository. A per-dose volume of 0.1 ml was intrarectally distilled through an 18-gauge blunt needle to sedated mice. Mice received 3 doses of 4-OHT; the first dose was an "induction" dose through which mice were deeply sedated by ketamine and xylazine and inverted after enema administration to maximize rectal retention. In the other "maintenance" doses, mice were sedated by isoflurane, given the enema, and allowed to recover quickly.

EdU/BrdU double labeling

Mice were intraperitoneally injected with 2.5 mg EdU (Cayman 20518) 24 h

prior to dissection and 2.5 mg BrdU (Abcam ab142567) 2 h prior to dissection. Nucleotides were prepared as 10 mg/ml stocks in PBS.

For dual revelation of EdU and BrdU incorporation in paraffin-embedded sections, slides with tissue were de-waxed and boiled for 20 min in a pressure cooker in 10 mM sodium citrate buffer (pH 6.0) supplemented with 0.5% Tween. The Click reaction was performed for 30 min at room temperature with 2 mM $CuSO_4$, 10 mM THPTA, 2 µM Cy3 azide (Click Chemistry Tools AZ119-1), and 100 mM ascorbic acid freshly prepared and diluted in Gomori buffer (pH 7.4). After the Click reaction, immunofluorescent staining with a rat anti-BrdU antibody (1:100, clone BU1/75, Abcam ab6326) was performed according to standard protocols.

For dual revelation of EdU and BrdU in full-thickness specimens, the iDISCO protocol was followed, except the EdU solution (lacking THPTA) was applied after initial permeabilization of tissue with methanol and DMSO/Triton-X. The Click reaction was performed overnight at room temperature. After washing, tissue was acid-treated with 2N HCl for 90 min at room temperature. Tissue was then stained with the anti-BrdU antibody (1:100) according to the iDISCO timings, and cleared in CUBIC-2 prior to imaging.

Single-cell RNA-Seq

Primary tissue was dissected from the anorectal junction with a 2-mm margin. The mucosal/surface tissue was separated from the muscle layer and digested in warmed 0.25% Trypsin-EDTA for 30 min with agitation at 160 rpm. Tissue was triturated and passed through a 100-µm mesh filter, treated with RBC lysis solution (Biolegend), washed and spun twice at 300 rcf with divalent-free HEPES-buffered saline[11] supplemented with 1% BSA, and purified for live cells using the EasySep Live/Dead magnetic selection kit (Stem Cell Technologies). Suspensions were passed through a 40-µm mesh filter to obtain single cells.

Single-cell suspensions of the combined distal colon, SNEC, transition zone, and anus were submitted representing 3 timepoints (exp d 0, 7, and 42) in mice treated with 3% DSS for 6 d (exp d 0-6). Five wildtype Bl/6 mice were pooled for each timepoint in order to obtain sufficient numbers of cells for analysis. Barcoding and library preparation was performed, allocating a separate microfluidic well per timepoint, on the Chromium (10X Genomics) platform. Targeted recovery was 10,000 cells per timepoint. Libraries were sequenced on a HiSeq 4000 machine (Illumina).

Bulk RNA-Seq experiments

C57Bl6/J mice were obtained from Jackson Labs and exposed to 3% DSS for 6 d. Mice were euthanized at 0, 3, 6, 9, 14, and 20 d from the beginning of DSS exposure. The distal colon and anus mucosal layers were isolated and manually separated from the underlying muscle under the guidance of a stereomicroscope. The mucosal “peel” from each mouse was divided into 4 pieces: 1) anus, 2) metaplasia (applicable from d 9 onward), 3) distal colon, or the distal-most 500-µm colonic length of colonic crypts, and 4) less-distal colon, or the piece of colon representing 1000-500-µm linear distance from the anal margin, which is overall similar to distal colon but more resistant to DSS-induced ulceration. Tissue was agitated and rendered with a manual homogenizer (Squisher, Zymo Research) in Trizol, and RNA isolation was performed according to standard procedures. Sequencing libraries were prepared from polyA-selected RNA using Illumina stranded library preparation kits and were sequenced on 3 NextSeq high-output lanes.

❖

Quantification and statistical analysis

Deep Mucosal Imaging (DMI)

Software routines for post-processing of images are available at github.com/stalepig/deep-mucosal-imaging Images were stitched from metadata and visualized with a custom-written interface atop of ImageJ. Refractive signal loss through the depth of the image stack was estimated from the mean intensity of cell nuclei, fit to an exponential function, and corrected in other channels as necessary. Confetti fluorophores were excited with 405 nm (mCFP), 488 nm (nGFP and YFP), and 543 nm (RFP), while methyl green was excited with a 633 nm laser.

Single-cell RNA-Seq

Quantification of transcript abundance per cell was obtained from cellranger 3.0 (10X Genomics). Matrix files were imported into Monocle 3[12]. Cells with fewer than 100 UMI reads were discarded. Count matrices were merged across samples for downstream analysis. Matrices were normalized by dividing by a size factor (the sum total of counts within a cell) and then log-transformed. Dimensionality reduction was performed via principal component analysis, taking the top 50 principal components. These principal components were used for further dimensional reduction to two-dimensional space via uniform manifold approximation and projection (UMAP). This was executed using the

default options in Monocle 3 for the reduce_dimension() function. Unsupervised clustering was performed from the UMAP values using the leiden algorithm, except in Figure 3B, which was clustered using the louvain algorithm. Dimensional reduction and clustering were repeated after subsetting of data.

Markers of cell clusters were identified using the top_markers() function in Monocle 3. All cells were used for marker analysis. The specificity of each marker for each cell group was computed using the Jensen-Shannon method. The significance of each marker was assessed using general linear model with a binomial family function and a likelihood ratio test between a full and simplified model of expression. The false discovery rate is reported as the adjusted p value (q value). The top 150 markers by specificity and meeting a q value < 0.001 were then uploaded into the pathway analysis aggregator enrichr[13] (https://maayanlab.cloud/Enrichr/) to identify targets of organ similarity (e.g., Figure 3G) using the BioGPS database[14]. These markers were also plotted as heatmaps. The top 5 markers were used for manual annotation of clusters and celltype assignment. Clusters were not merged except to facilitate display of epithelial and TZ cells as a single unit in Figure 3A and Figure S6B,D. To compare epithelial cells of different tissue types across time in aggregate (e.g., Figure 4H, S2B), clusters were merged for Krt8+ (colonic), Krt14+ Krt7+ Sox2+ (TZ-derived), or Krt14+ Krt7- Sox2+ (anal epidermal) cells. Explicit differential gene expression analysis between samples was performed with the fit_models() function in Monocle 3; genes with an adjusted p value (q value) < 0.2 were selected for enrichr analysis as shown in Figure 6B.

Heatmaps were generated for select groups of genes by computing the average normalized counts per gene per cluster. The expression of each gene was scaled across clusters prior to display. For GSEA[15], the average normalized counts for all genes were exported as text files. Gene names were translated to human symbols using the Homologene database. GSEA 4.0.3 (Broad Institute) with the default options was used to determine enrichment of KEGG, Reactome, Gene Ontology, or HALLMARK pathway databases in the expression data. To assess SNEC similarity to marker genes of fetal intestinal epithelium, a dataset of gene expression from fetal mouse intestinal organoids[16] was used as a reference.

To perform partition-based graph abstraction (PAGA) analysis[17], cellranger-derived count matrices were imported into scanpy[18]. The count matrix was normalized for size and log-transformed. High variable genes were selected with a minimum expression > 0.0125, a maximum expression < 3, and a minimum dispersion > 0.5. Twenty principal components were used to compute neighbors for UMAP. The leiden algorithm was used to perform clustering. PAGA analysis was performed using the default options in the scanpy.pl.paga function.

Bulk RNA-Seq experiments

Resultant 1x75 bp reads were demultiplexed and pseudoaligned with kallisto 0.42.5 software (https://pachterlab.github.io/kallisto/) to obtain tpm (transcripts per million) values[19]. Pathway analyses were performed using GSEA software (Broad Institute) with the HALLMARK [20] reference. Normalized enrichment scores for individual pathways represent a comparison of the experimental timepoint for a given tissue fraction with the 0-d timepoint of the same tissue fraction (e.g., colon at exp d 9 vs. colon at exp d 0). SNEC pathway enrichment was compared against 0-d timepoints of both the distal colonic and anal tissue fractions. Global markers of SNEC were obtained by examining individual genes with expression ratio >4 compared to distal colonic and anal tissue fractions, merged across all times. Distinguishing pathways of marker genes were classified using functional annotation clustering on DAVID (https://david.ncifcrf.gov/)[21].

Clonal analysis

The volume of Confetti+ clones was measured from DMI-reconstructed z-stacks of SNEC. A region of interest corresponding to the xy-extent of the fluorescently marked cytosolic YFP+ or RFP+ cells was picked for each clone at each image plane in ImageJ. The volume of the clone was summed from each z plane and normalized to the total volume of the rete peg identified by its outlined structure from a reference nuclear stain. The allocation of clonal volume across tip, casing, or core compartments was made according to the following definitions: 1) tip - the single-cell thick periphery of the rete peg beginning at the abluminal vertex and extending until the peg has reached maximum cylindrical thickness, 2) casing - the single-cell thick peripheral layer of the cylindrical subportion of the rete peg, and 3) core - cells harboring a partially differentiated “spread” morphology in the interior of the rete peg. Sixty clones from combined 3-4 mice were measured per lineage tracing timepoint.

Statistical methods

Measurements were taken from distinct samples, as full-volume imaging allows the whole sample to be processed. Data with an exponential distribution were log-transformed to obtain normal distributions prior to statistical testing. Statistical comparisons between experimental and control conditions were performed using

pairwise two-sided t-tests. Unless otherwise indicated, the Benjamini-Hochberg method of false discovery adjustment for multiple comparisons was applied to the p values. The adjusted p values (q values) are reported. Statistical tests were not performed on lineage tracing data when the effect size multiple was >3 or the number of samples was <3; in these cases the individual data points are plotted along with the mean.

Modeling

From the clonal data, we found that, over time, individual rete pegs in SNEC can become monoclonally labeled. In principle, such behavior could indicate that rete pegs are maintained by a single stem cell that renews through asymmetric division. Alternatively, such behavior could derive from an equipotent proliferative population that, like the normal intestinal colonic epithelium, undergoes neutral drift dynamics involving the chance replacement of neighboring tip cells lost through differentiation. To assess whether the tip proliferative cells of the rete pegs in SNEC represent an equipotent progenitor cell population, we turned to a statistical modeling-based approach.

Motivated by previous studies of crypt clone dynamics in the intestine[22] we considered a "voter model" to quantify cell dynamics within an equipotent progenitor population. In this formulation, a tip proliferative cell *S* carrying a hereditary label 'a' can replace its neighbor with label 'b' through symmetric cell duplication at a rate λ, described by the replacement "reaction:"

$$S^a S^b \xrightarrow{\lambda} S^a S^a$$

For an individual clone, the resulting pattern of lineage expansion and contraction follows a random walk or diffusion-like pattern. The mathematical formalism of this model has been detailed previously [22] and we only describe the essential findings here. When the renewing cells are arranged as a one-dimensional annulus or "ring," such as that formed by intestinal stem cells at the crypt base, the smoothed distribution of clone sizes at early times, prior to significant clone fixation, takes the statistical scaling form:

$$f(x) = \frac{\pi x}{2} e^{-\pi x^2/4}$$

where $x = n/\overline{n(t)}$ denotes the clone size n normalized by the average $\overline{n(t)}$. Therefore, a key test of whether the cellular system exhibits scaling is to plot the relationship $P_n(t)\overline{n(t)}$ vs. $n(t)/\overline{n(t)}$, where *P* describes the probability of finding a clone of size *n* at time *t*. In a scaling system, clonal data from all timepoints should collapse onto the same time-

independent scaling curve, $f(x)$, provided the empirical tracing data are analyzed sufficiently early as to be within the scaling regime. Moreover, in the one-dimensional geometry, the surviving clone probability is given by:

$$R(t) = \frac{1}{\sqrt{\pi\lambda t}}$$

while the average size *n* varies as:

$$n(t) = \sqrt{\pi\lambda t}$$

However, in the present case, the abluminal tip of the rete peg forms an ostensibly two-dimensional grid-like organization. In this case, the neutral drift dynamics are predicted to result in a scaling form in which the probability distribution converges over time to a simple exponential:

$$f(x) = e^{-x}$$

However, in stochastic simulations we found that in relatively small 2d compartments (i.e., lattices of 6x6 cells or smaller, within the size ranges of rete peg tip compartments), in which diagonal neighbor replacements are disallowed, the resulting clonal distributions could be reasonably modeled as 1d-type dynamics. Technically, an improved fit could be obtained with a linear combination of 1d (major contribution) and 2d (minor contribution) scaling functions (Figure S13A). However, for simplicity we primarily used the 1d voter model to study the tip clonal fate data. Importantly, when overall applied to the clonal data, we found that the 1d model provided a good fit to the data (Figure 7E).

To estimate the loss/replacement rate of the tip progenitors from the empirical long-term tracing of tip cells, we turned to the random walk equations that define exact solutions that account for clonal extinction and fixation within the scaling regime and during the later stages of clonal fixation. For the 1d voter model, the clone size probability distributions take the form:

$$P_n(t) = \frac{2}{N}\sum_{k=1}^{N-1} \sin\left[\frac{\pi k}{N}\right] \sin\left[\frac{\pi k n}{N}\right] e^{-4\sin^2\left(\frac{\pi k}{2N}\right)\lambda t} \text{ for } 1 \leq n \leq N-1$$

$$P_0(t) = \frac{2}{N}\sum_{k=1}^{N-1} \cos^2\left[\frac{\pi k}{2N}\right]\left(1 - e^{-4\sin^2\left(\frac{\pi k}{2N}\right)\lambda t}\right)$$

$$P_N(t) = \frac{2}{N}\sum_{k=1}^{N-1}(-1)^{k+1}\cos^2\left[\frac{\pi k}{2N}\right]\left(1 - e^{-4\,sin^2\left(\frac{\pi k}{2N}\right)\lambda t}\right)$$

where *N* denotes the total size of the stem cell pool. To apply the model to the data, it is tempting to assume that the total number of renewing cells equates to the total number of tip proliferative cells. However, studies[23] in the intestine have shown that the effective stem cell number *N* may be smaller than the number of cells that have renewal potential. Therefore, to address the experimental data, it makes sense to score clone size as a fraction of the total number of cells in the rete peg, defining the effective clone size as the fraction of the total area of the tip domain that are labeled, which translates to *n/N*. In this case, by fitting the distribution of persisting clone sizes $P_n^{(pers)}(t) = {P_n(t)}/{(1 - P_0(t))}$ to the model prediction, we can estimate the effective rate:

$$\nu = {\lambda}/{N^2}$$

Moreover, to account for a practical delay between lineage tracing initiation (tamoxifen injection) and observation of clonal patches, the experimental data were shifted by $\tau = -1$ week.

Since it was difficult to consistently resolve individual nuclei at the depth of the rete peg tip, we made use of clone morphologies in 3d images to quantify the fraction of labelled cells in rete pegs. The relationship of these fractions to cell numbers is simple if one assumes that tip compartment size is uniform. However, we found that the empirical fractional clone size converged to ~0.6 instead of the theoretical value of 1 (peg fixation), which challenged the fit at later tracing timepoints (> 12 wks, Figure S13B). We therefore devised a simple correction, based on our observation of heterogeneity in rete peg sizes. We found that rete peg diameters could range between 5-11 cells (data not shown). If one assumes that the tip cell compartment is of uniform absolute size, irrespective of the diameter of the rete peg, essentially implying that we have over-estimated the size of the tip compartment, then we can normalize the empirical clone size fraction by the maximally observed size. Such a normalization results in an improved fit (Figure 7F), with $\nu = 0.02$ /wk, which is similar to the scaling kinetics found in small intestine and colon[22].

Finally, we note that the scaling behavior was robust to heterogeneity in tip compartment sizes. In particular, if *N* is not fixed, but is sampled from a statistical distribution, we could also generate good fits to the data. A key question is what is an

appropriate distribution to use for values of *N*. A sample distribution can be obtained experimentally by examining the distribution of clone fractions at timepoints >12 wks, when the mean clone size does not change appreciably (Figure S13C). Thus, except for a small fraction of rete pegs that have become fully fixed as a result of a slow or infrequent transfer of cells from the tip to the casing compartments, this distribution provides insight into the spread of tip compartment sizes; that is, the tip compartment has been fixed. This tip "fixation distribution" can be sampled from a probability density function (Figure S13D) that has the empirical form:

$$\Gamma(N) = \alpha e^{-\alpha(N_{max}-N)}$$

where N_{max} describes the maximum tip compartment size (we used $N_{max} = 36$ as it represents a 6x6 compartment). The result of this modification is a scaling system whose behavior represents the weighted sum of scaling subsystems evolving with different kinetics (λ/N^2, assuming a fixed λ value). This model provided an excellent fit to the surviving clone size distribution, the growth in the mean clone size, and the shape of the clone size distribution throughout the tracing experiment (Figure S13E,F). However, it precludes reporting a single value of λ/N^2. We note that previous theory has demonstrated that scaling behavior is preserved even with heterogeneity in values in λ/N^2, provided that these values are sampled from a distribution with finite variance. Thus, any tip compartment size distribution meeting this criterion would result in scaling behavior.

Methods References

1. el Marjou F, Janssen KP, Chang BH, et al. Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. Genesis 2004;39:186-93.
2. Jiang M, Li H, Zhang Y, et al. Transitional basal cells at the squamous-columnar junction generate Barrett's oesophagus. Nature 2017;550:529-533.
3. McMahon AP, Aronow BJ, Davidson DR, et al. GUDMAP: the genitourinary developmental molecular anatomy project. J Am Soc Nephrol 2008;19:667-71.
4. Tian H, Biehs B, Warming S, et al. A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable. Nature 2011;478:255-9.
5. Dube PE, Yan F, Punit S, et al. Epidermal growth factor receptor inhibits colitis-associated cancer in mice. J Clin Invest 2012;122:2780-92.
6. Liu CY, Dube PE, Girish N, et al. Optical reconstruction of murine colorectal mucosa at cellular resolution. Am J Physiol Gastrointest Liver Physiol 2015;308:G721-35.
7. Liu CY, Polk DB. Cellular maps of gastrointestinal organs: getting the most from tissue clearing. Am J Physiol Gastrointest Liver Physiol 2020;319:G1-G10.
8. Susaki EA, Tainaka K, Perrin D, et al. Whole-brain imaging with single-cell resolution using chemical cocktails and computational analysis. Cell 2014;157:726-39.
9. Prieto D, Aparicio G, Machado M, et al. Application of the DNA-specific stain methyl green in the fluorescent labeling of embryos. J Vis Exp 2015:e52769.
10. Renier N, Wu Z, Simon DJ, et al. iDISCO: a simple, rapid method to immunolabel large tissue samples for volume imaging. Cell 2014;159:896-910.
11. Liu CY, Tam SS, Huang Y, et al. TNF Receptor 1 Promotes Early-Life Immunity and Protects against Colitis in

Mice. Cell Rep 2020;33:108275.
12. Qiu X, Mao Q, Tang Y, et al. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods 2017;14:979-982.
13. Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 2016;44:W90-7.
14. Wu C, Jin X, Tsueng G, et al. BioGPS: building your own mash-up of gene annotations and expression profiles. Nucleic Acids Res 2016;44:D313-6.
15. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545-50.
16. Mustata RC, Vasile G, Fernandez-Vallone V, et al. Identification of Lgr5-independent spheroid-generating progenitors of the mouse fetal intestinal epithelium. Cell Rep 2013;5:421-32.
17. Wolf FA, Hamey FK, Plass M, et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol 2019;20:59.
18. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol 2018;19:15.
19. Bray NL, Pimentel H, Melsted P, et al. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 2016;34:525-7.
20. Liberzon A, Birger C, Thorvaldsdottir H, et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst 2015;1:417-425.
21. Dennis G, Jr., Sherman BT, Hosack DA, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol 2003;4:P3.
22. Lopez-Garcia C, Klein AM, Simons BD, et al. Intestinal stem cell replacement follows a pattern of neutral drift. Science 2010;330:822-5.
23. Kozar S, Morrissey E, Nicholson AM, et al. Continuous clonal labeling reveals small numbers of functional stem cells in intestinal crypts and adenomas. Cell Stem Cell 2013;13:626-33.

Supplemental Table

Table S1: Key resources used in the study.

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit anti-cleaved caspase-3	Cell Signaling Technology	Cat# 9661
Rat anti-Ki-67, clone SolA15	eBioscience	Cat# 41-5698
Rabbit anti-KRT7	Abcam	Cat# ab181598
Rabbit anti-KRT14	Abcam	Cat# ab206100
Rat anti-CDH1	R&D Systems	Cat# FAB7481F
Rabbit anti-CLDN1	Thermo Fisher Scientific	Cat# 71-7800
Rabbit anti-LOR	Biolegend	Cat# 905101
Rabbit anti-KRT17	Millipore Sigma	Cat# AV41733
Rabbit anti-collagen XVII	Abcam	Cat# ab184996
Rabbit anti-GSTO1	Novus	Cat# NBP1-33763
Rabbit anti-CHGA	Immunostar	Cat# 20085
Rabbit anti-KRT8	Abcam	Cat# ab53280
Rabbit anti-IVL	Biolegend	Cat# 905602
Rabbit anti-KRT1	Biolegend	Cat# 905602
Rabbit anti-KRT10	Proteintech	Cat# 18343-1-AP
Rat anti-BrdU, clone BU1/75	Abcam	Cat# ab6326
Goat anti-rabbit IgG, Alexa 568	Thermo Fisher Scientific	Cat# A-11011
Goat anti-rat IgG, Alexa 488	Thermo Fisher Scientific	Cat# A-11006
Donkey anti-rabbit IgG, Alexa 647	Thermo Fisher Scientific	Cat# A-31573
Biological samples		
Archived colons from Il10-/- mice	[1]	N/A
Chemicals, peptides, and recombinant proteins		
Dextran sulfate sodium 36-50kDa	MP Bio	Cat# 0216011091
EdU	Cayman Chemical	Cat# 20518
BrdU	Abcam	Cat# ab142567
Cy3 azide	Click Chemistry Tools	Cat# AZ119-1
Tamoxifen	Millipore Sigma	Cat# T5648-1G
4-Hydroxytamoxifen	Millipore Sigma	Cat # H6278-50MG
Trypsin-EDTA	Thermo Fisher Scientific	Cat# 25200056
DNase I	Millipore Sigma	D5025
RBC lysis solution	Biolegend	Cat# 420301
Witepsol H-15	Medisca	Cat# 1699
CUBIC-1	[2]	N/A
CUBIC-2	[2]	N/A
Critical commercial assays		
RNAScope 2.5 HD Assay – Brown	Biotechne / ACDBio	Cat# 322300
ImmPACT DAB Substrate	Vector Labs	Cat# SK-4105

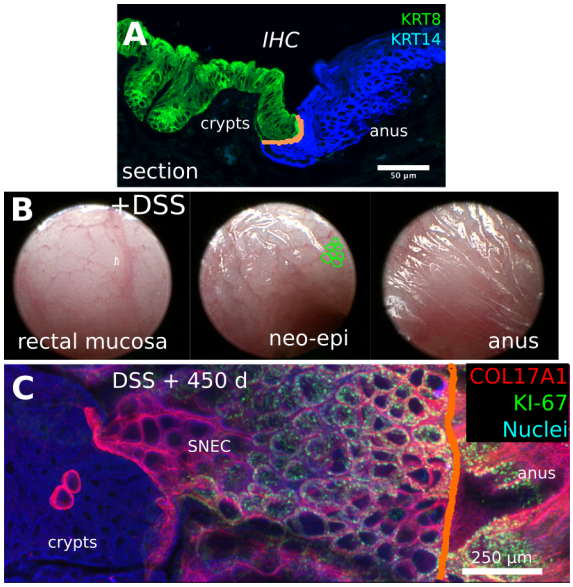
ImmPACT VIP Substrate	Vector Labs	Cat# SK-4605
ImmPRESS HRP anti-Rabbit IgG Polymer Kit	Vector Labs	Cat# MP-7801-15
ImmPRESS HRP anti-Mouse IgG Polymer Kit	Vector Labs	Cat# MP-7802-15
Chromium	10X Genomics	N/A
EasySep Dead Cell Removal Kit	Stemcell Technologies	Cat #17899
Deposited data		
Bulk RNA-Seq	This paper	GEO: GSE168053
Single-cell RNA-Seq	This paper	GEO: GSE168033
Experimental models: Genetically modified mice		
Rosa26::-tdtTomato/-EGFP (mTmG)	Jackson Laboratory	Cat# 007676
Rosa26::Brainbow2.1 (Confetti)	Jackson Laboratory	Cat# 013731
Rosa26::LSL-DTA	Jackson Laboratory	Cat# 009669
Rosa26::LSL-tdTomato	Jackson Laboratory	Cat# 007914
Lgr5::EGFP-IRES-CreER	Jackson Laboratory	Cat# 008875
Vil1::Cre	[3]	N/A
Sox2::CreER	Jackson Laboratory	Cat #017593
Krt7::CreER	[4]	N/A
Krt20::CreER	Jackson Laboratory	Cat #030600
Lgr5::DTR-EGFP	[5]	N/A
Bl/6 (WT)	Jackson Laboratory	Cat #000664
Software and algorithms		
Cellranger V3.0.2	10X Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/overview/welcome
Monocle3 V1.0.0	[6]	https://cole-trapnell-lab.github.io/monocle3/
Enrichr	[7]	https://maayanlab.cloud/Enrichr/
GSEA V4.0.0	[8]	https://www.gsea-msigdb.org/gsea/index.jsp
ScanPy / PAGA	[9, 10]	https://scanpy-tutorials.readthedocs.io/en/latest/index.html
Kallisto V0.42.5	[11]	https://pachterlab.github.io/kallisto/
DAVID	[12]	https://david.ncifcrf.gov/
Deep Mucosal Imaging (DMI)	This paper	https://github.com/stalepig/deep-mucosal-imaging
Single-cell RNA-Seq analysis scripts	This paper	https://github.com/stalepig/2021-SNEC-single-cell
Other		
RNAScope probe – Mm-Krt17	Biotechne / ACDBio	Cat# 479911
RNAScope probe – Mm-Ass1	Biotechne / ACDBio	Cat# 447991

RNAScope probe – Mm-Ly6a	Biotechne / ACDBio	Cat# 427571
Molecular Instruments probe – Mm-Krt7	Molecular Instruments	Accession# NM_033073.3

Table References

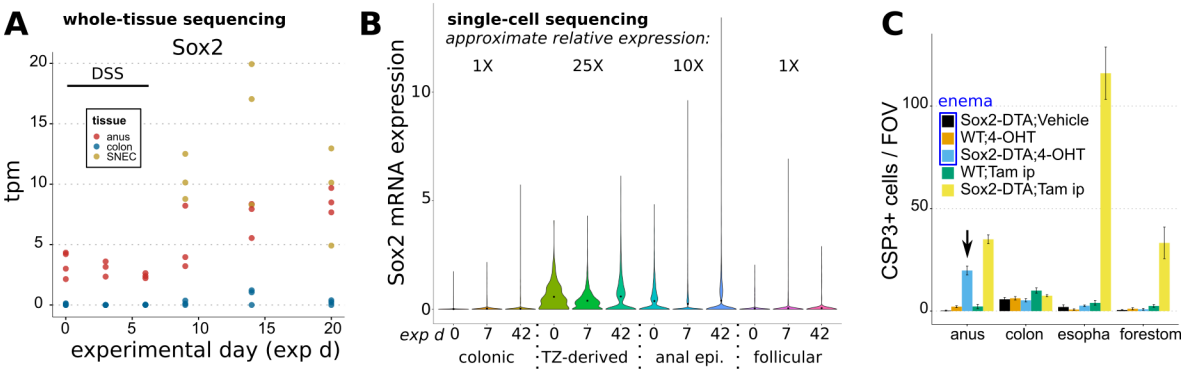
1. Dube PE, Yan F, Punit S, et al. Epidermal growth factor receptor inhibits colitis-associated cancer in mice. J Clin Invest 2012;122:2780-92.
2. Susaki EA, Tainaka K, Perrin D, et al. Whole-brain imaging with single-cell resolution using chemical cocktails and computational analysis. Cell 2014;157:726-39.
3. el Marjou F, Janssen KP, Chang BH, et al. Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. Genesis 2004;39:186-93.
4. Jiang M, Li H, Zhang Y, et al. Transitional basal cells at the squamous-columnar junction generate Barrett's oesophagus. Nature 2017;550:529-533.
5. Tian H, Biehs B, Warming S, et al. A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable. Nature 2011;478:255-9.
6. Qiu X, Mao Q, Tang Y, et al. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods 2017;14:979-982.
7. Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 2016;44:W90-7.
8. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545-50.
9. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol 2018;19:15.
10. Wolf FA, Hamey FK, Plass M, et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol 2019;20:59.
11. Bray NL, Pimentel H, Melsted P, et al. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 2016;34:525-7.
12. Dennis G, Jr., Sherman BT, Hosack DA, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol 2003;4:P3.

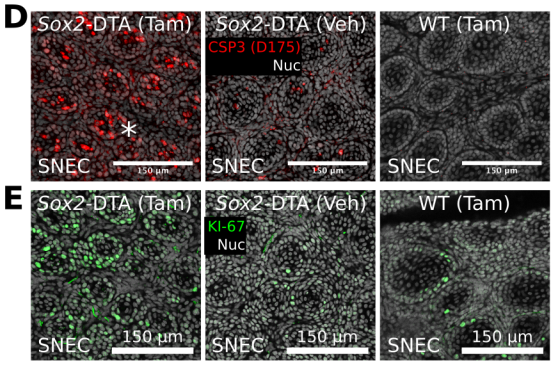
Figure S1

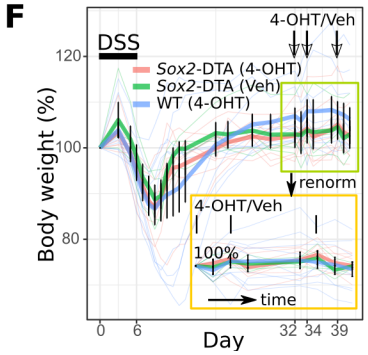


Squamous neo-epithelium of colon (SNEC) contributes to long-term colonic wound healing. Related to Figure 1. A) Thin-section photomicrograph of immunostained colonic (KRT8+) and anal squamous (KRT14+) epithelium reveals a sharp boundary between columnar and squamous epithelium at the dentate line (orange line), in the absence of injury. B) Mouse colonoscopy shows distinct structure of SNEC after DSS-induced injury. Example rete pegs are traced in green. C) In this *en face* projection of an interior section of a whole-mount staining, basal cells in SNEC express the squamous marker COL17A1 and the proliferation marker KI-67. Neighboring anal epithelium also expresses COL17A1 but colonic crypts do not. The dentate line is highlighted in orange. Error bars: SE.

Figure S2

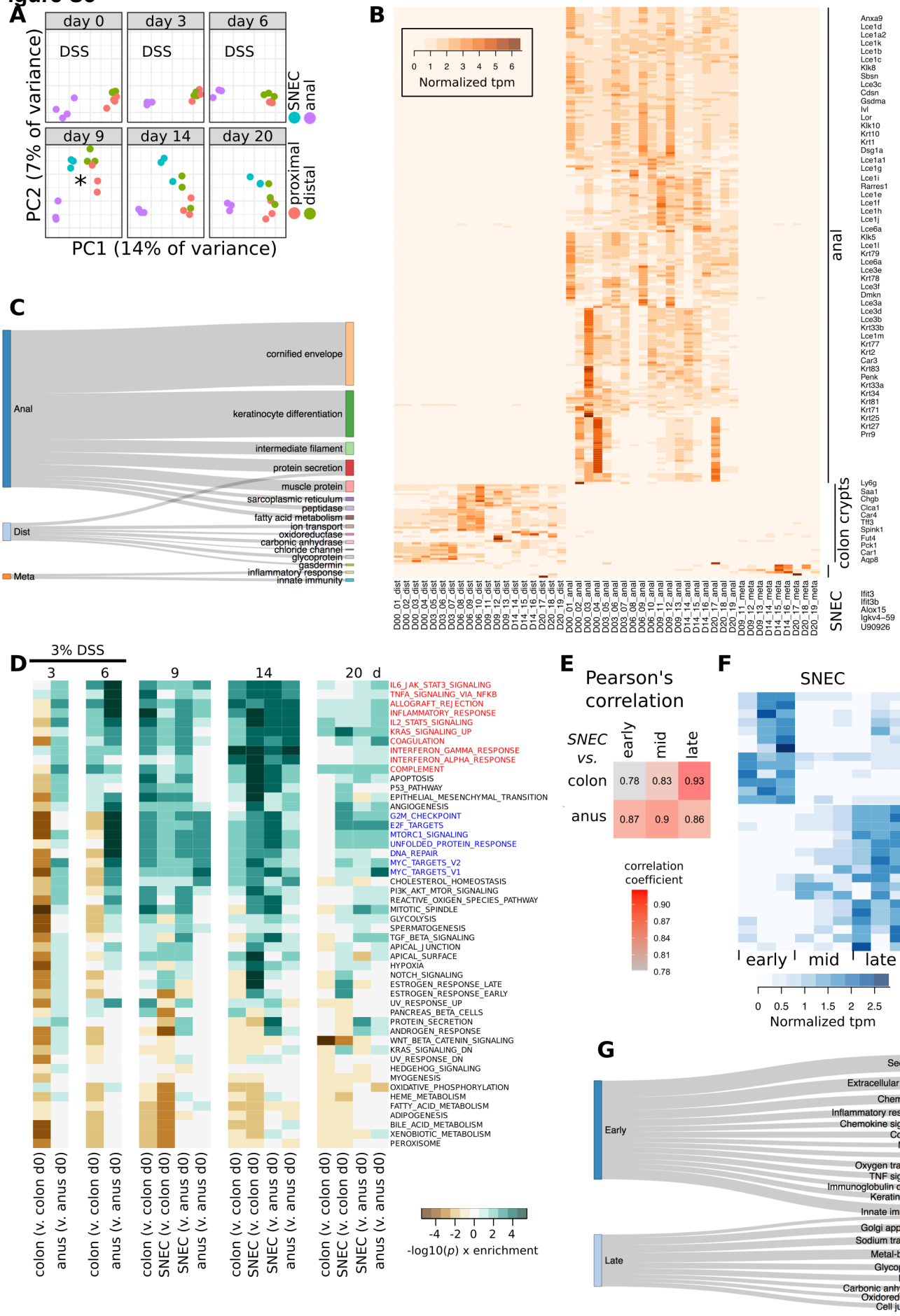






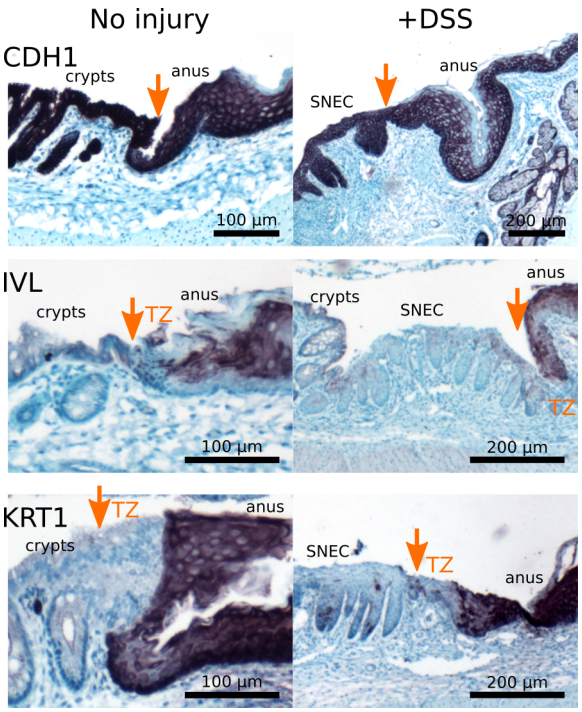
Induction of apoptosis in anorectal squamous cells with a conditional genetic construct. Related to Figure 1. A,B) Expression of Sox2 mRNA in whole-tissue sequencing (A) or single-cell sequencing (B) across different tissue or epithelial types shows consistent >10-fold enrichment for Sox2 expression in squamous epithelium vs. colonic epithelium. These sequencing data were obtained from experiments described in detail in Figures 4-6 and S3. C) We tested the specificity of targeting apoptotic signals to anal squamous tissue using Sox2-DTA mice (Sox2::CreER;Rosa26::LSL-DTA mice). Counting of caspase-3+ (CSP3, apoptotic) cells in response to enema (topical) or injection (systemic) of 4-OHT/tamoxifen shows specific apoptosis of anal epithelium after enema administration in Sox2-DTA mice (n=4 mice/condition). D) Whole-mount images demonstrate localization of cleaved caspase-3 (CSP3, asterisk) to SNEC cells after treatment of Sox2-DTA mice with DSS (to induce injury and SNEC formation) and 2 mg tamoxifen on exp d 32. E) Proliferative signaling, as revealed through KI-67 immunostaining, persisted in SNEC and may have been upregulated in surviving cells after apoptotic targeting to SNEC tissue. F) In contrast to treatment during acute injury and repair, enema administration of 4-OHT after mucosal healing does not affect mouse body weight (n=6 mice/condition). Error bars: SE.

Figure S3


A
day 0
day 3
day 6
day 9
day 14
day 20
DSS
PC2 (7% of variance)
PC1 (14% of variance)
SNEC
anal
proximal
distal
B
Normalized tpm
0 1 2 3 4 5 6
anal
colon crypts
SNEC
C
Anal
Dist
Meta
cornified envelope
keratinocyte differentiation
intermediate filament
protein secretion
muscle protein
sarcoplasmic reticulum
peptidase
fatty acid metabolism
ion transport
oxidoreductase
carbonic anhydrase
chloride channel
glycoprotein
gasdermin
inflammatory response
innate immunity
D
3% DSS
3
6
9
14
20 d
colon (v. colon d0)
anus (v. anus d0)
SNEC (v. colon d0)
SNEC (v. anus d0)
-log10(p) x enrichment
E
Pearson's correlation
SNEC vs.
early
mid
late
colon
anus
0.78
0.83
0.93
0.87
0.9
0.86
correlation coefficient
F
SNEC
early
mid
late
Normalized tpm
G
Early
Late
Secretion
Extracellular matrix
Chemotaxis
Inflammatory response
Chemokine signaling
Collagen
Mitosis
Lectin
Oxygen transport
TNF signaling
Immunoglobulin domain
Keratinization
Innate immunity
Golgi apparatus
Sodium transport
Metal-binding
Glycoprotein
Kinase
Carbonic anhydrase
Oxidoreductase
Cell junction

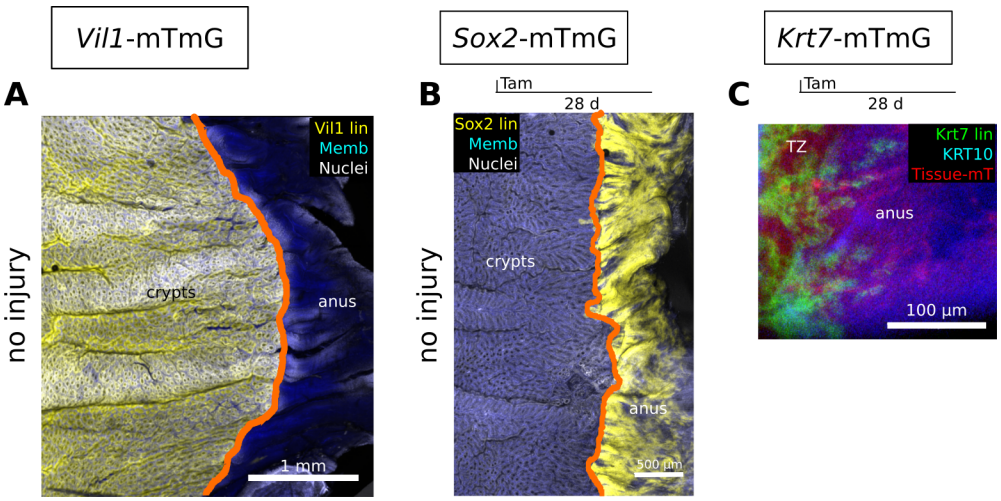
Whole-tissue profiling of colonic, anal, and SNEC samples. Related to Figure 1. A) Principal component plot of RNA-Seq data obtained from proximal-to-anal segments of colons from DSS-treated mice. Each dot represents a tissue sample. The initial emergence of SNEC is indicated by an asterisk (*). B-C) Tissue marker analysis (B) with gene ontology analysis (C) of anus, colonic crypts, or SNEC identify known squamous differentiation markers in anus, absorptive and secretory cell markers in colonic crypts, and interferon-inducible and inflammatory markers in SNEC (“meta”). D) GSEA of pathway activation/suppression at different timepoints versus the 0-day (uninjured) timepoint in colon or anus. Inflammatory pathways are highlighted in red. Proliferative pathways are highlighted in blue. “Enrichment” is defined as in the GSEA software. E-G) Changes in the overall expression characteristics of SNEC samples over time, shown as they correlate (median Pearson’s correlation coefficient) to mouse-matched anal or colonic samples (E), at the single-gene level gene (F), and in ontology analysis (G). Early SNEC (d9) exhibits higher expression of inflammatory and squamous differentiation genes, while late SNEC (d20) shows elevated expression of colon-like genes. However, we note that the absolute expression levels of these genes were ~10-fold less than in colon.

Figure S4



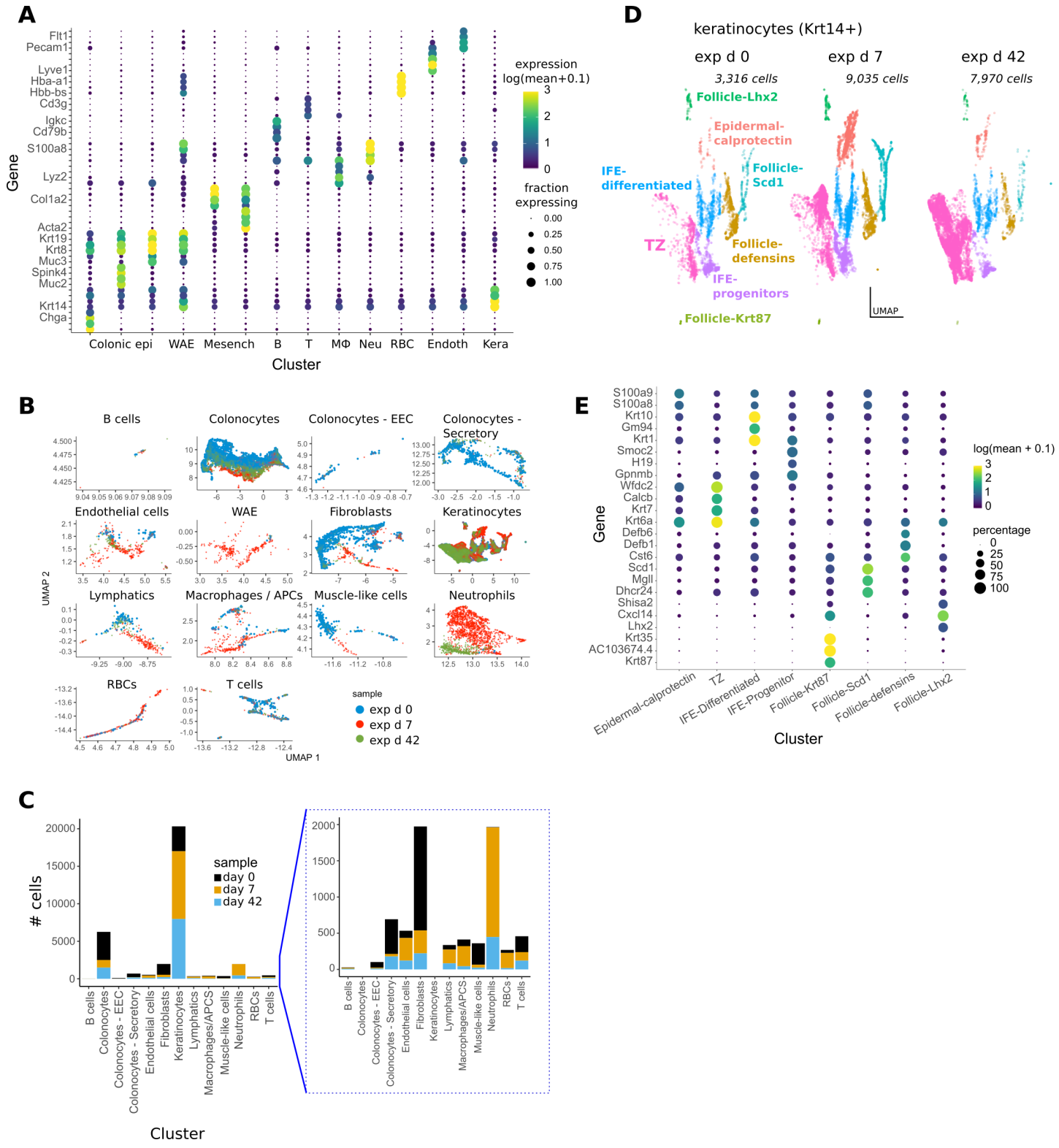
Related to Figure 2. Anal TZ represents a zone lacking terminal epidermal differentiation. Immunostaining of key markers of squamous differentiation at the anorectal junction, before (exp d 0) and after DSS (exp d 36) treatment. The orange arrow shows the dentate line, with the anal transition zone (TZ) marked immediately distal to it. Like SNEC, the TZ does not express IVL or KRT1 differentiation markers. Both colonic and anal epithelium express E-cadherin (CDH1) (n=3 mice/stain).

Figure S5



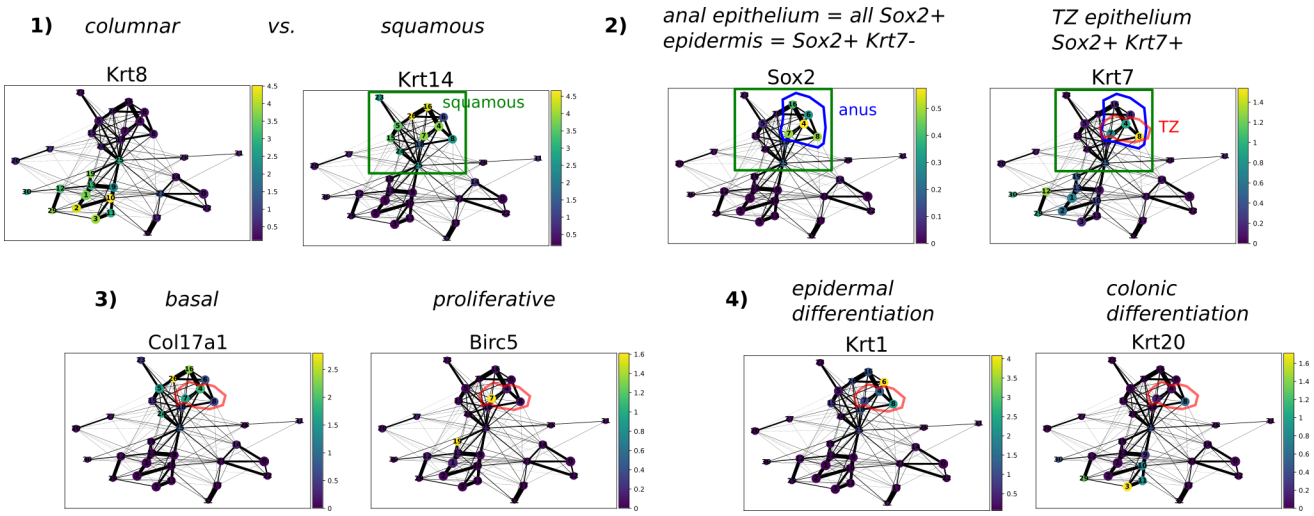
Assessing lineage contributions to the anorectal region in homeostasis. Related to Figure 2. Mice used: Vil1-mTmG (Vil1::Cre;Rosa26-mTmG, no tamoxifen injection), Sox2-mTmG (Sox2::CreER;Rosa26-mTmG, 2 mg i.p. tamoxifen), Krt7-mTmG (Krt7::CreER;Rosa26-mTmG, 0.2 mg i.p. tamoxifen). A) Whole-mount surface reconstruction of the anorectal region in Vil1-mTmG mice demonstrates specific labeling of crypts (n=3 mice). The dentate line is marked in orange. B) Results from Sox2-mTmG mice showing that tracing results in specific labeling of most anal cells (n=2 mice). C) Low doses (0.02 mg) of tamoxifen administered to Krt7-mTmG mice induce specific labeling of anal transition zone (TZ)-derived cells. After 28 d of tracing, labeled cells were found to extend distally into the differentiated anal epidermal zone marked by KRT10 immunostaining (n=4 mice).

Figure S6



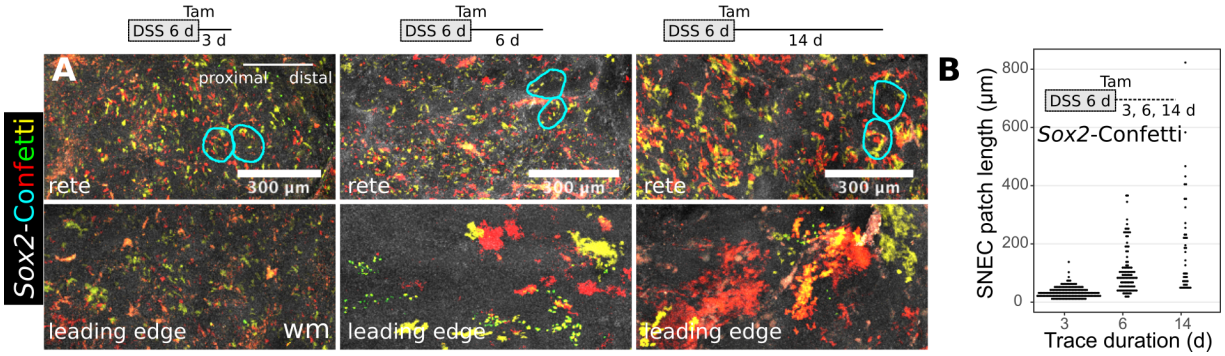
Transcriptomic characterization of cell type-specific responses during DSS colitis. Related to Figure 3. A) The dotted plot shows major gene markers driving the top-level clustering of the merged scRNA-Seq data obtained from the anorectal region at exp d 0, 7, and 42. Abbreviations: epi = epithelium, WAE = wound associated epithelium, Mesench = mesenchyme, B = B cells, T = T cells, Mϕ = macrophages, Neu = neutrophils, RBC = red blood cells, Endoth = endothelium, Kera = keratinocytes. B,C) UMAP visualization of changes in individual cell clusters (B) and their abundance (C) associated with DSS-induced injury; clusters where the dot colors do not overlap represent cell types undergoing broad shifts in transcriptional profile. D) The time-resolved UMAP visualization shows major subtypes of keratinocytes. Note the increased abundance of TZ-related cells at exp d 42, which corresponds to the presence of SNEC. IFE = interfollicular epithelium. E) The dotted plot reveals markers used to classify keratinocytes shown in D.

Figure S7



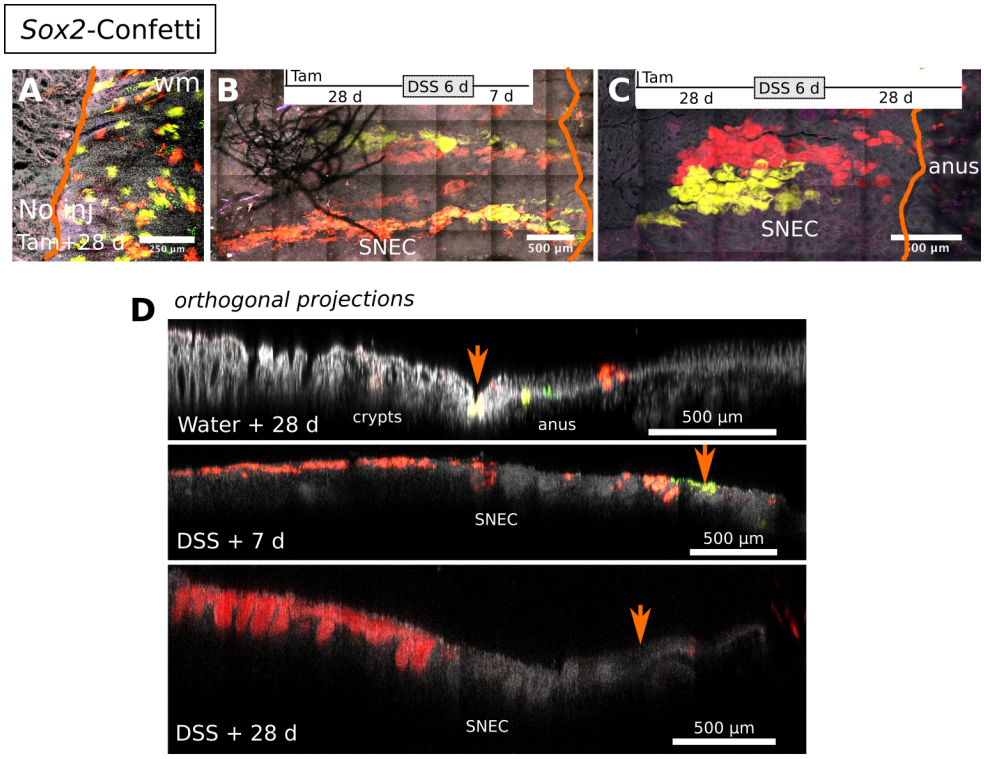
Graph analysis of single-cell mucosal transcriptomes obtained from exp d 0. Related to Figure 3. The count matrix was imported into scanpy, size-normalized and log-normalized, trimmed to highly variable genes, dimensionally reduced using 20 principal components, and projected using UMAP. Clusters were identified using the leiden algorithm. Similarity relationships between clusters were computed using partition-based graph abstraction (PAGA). Each point on the plot represents a cluster of cells. The color shading indicates the expression level of the gene per cluster. A “gating” strategy was applied such that squamous cells (Krt14+) were distinguished from colonocytes (Krt8+). Keratinocytes could be divided into anal squamous epithelium (Sox2+) or follicular-type epithelium (Sox2-negative). Within Sox2+ epithelium, three clusters of Krt7^{hi} cells were identified; these clusters putatively represent the anal TZ. Similar to the analytical pipeline performed in monocle3, two of the TZ clusters represented basal cells (Col17a1+), with one cluster harboring highly proliferative cells (Birc5+). The suprabasal TZ cluster was positive for the colonocyte marker Krt20 but lacked high expression of the squamous cell differentiation marker Krt1. Note that the TZ clusters are embedded within the squamous cell graph. This suggests that TZ cells are fundamentally squamous in nature, but have elevated expression of colonic markers. This finding is in agreement with the results independently obtained in Figure 3.

Figure S8



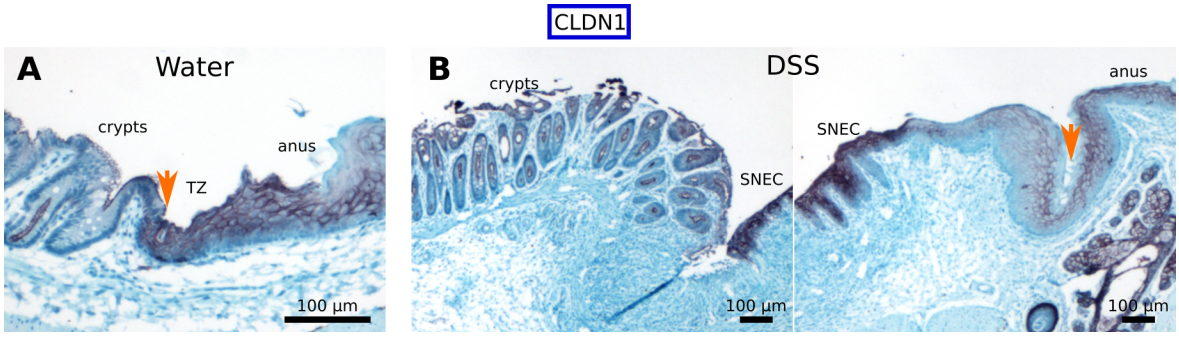
Sox2+ cells expand during acute colitis. Related to Figure 5. A) Short-term analysis in Sox2-Confetti mice during epithelial repair shows, in surface maximum intensity projections, the appearance of color patches after 3 d (exp d 9) and their expansion in the leading region (bottom row) after 6 (exp d 12) and 14 (exp d 20) d of tracing. Example rete pegs are outlined in light blue. B) Color patch size increases over time after DSS-induced injury. This shows that Sox2+ cells within SNEC drive the progressive re-epithelialization process.

Figure S9



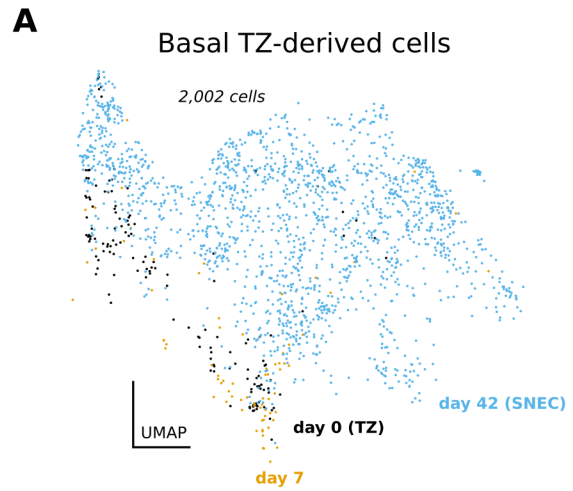
Emergence and remodeling of anal-derived cellular streams during SNEC formation. Related to Figure 5. A-C) Results from Sox2-Confetti mice show that clonal streams of cells forming SNEC arise from clonal collections of cells in the anus. The orange line marks the dentate line in whole-mount images. Surface reconstructions are displayed from uninjured colon (A) and from DSS-injured colon at 7 (B) and 28 d (C) after DSS withdrawal. D) Orthogonal projections from the images shown in A-C show the remodeling of the clonal streams from flat epithelium to rete peg epithelium. This demonstrates that SNEC is initially formed as a flat epithelium and that rete pegs emerge later. Orange arrows mark the dentate line.

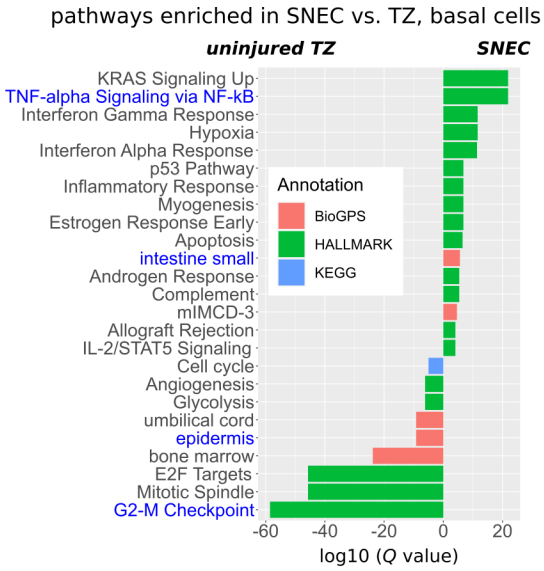
Figure S10

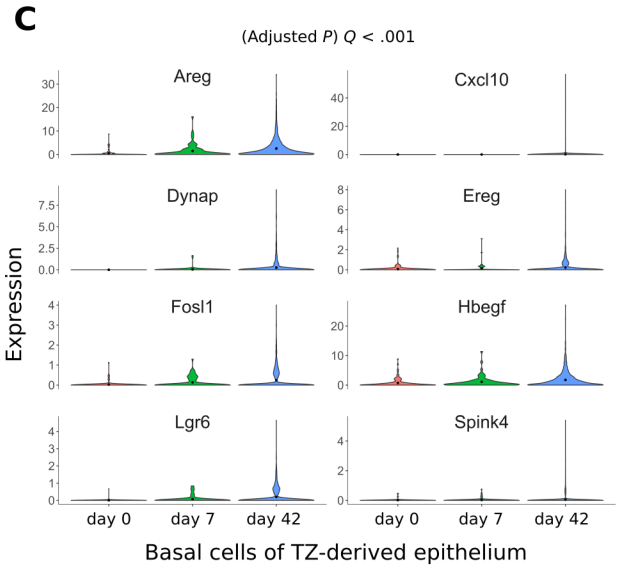


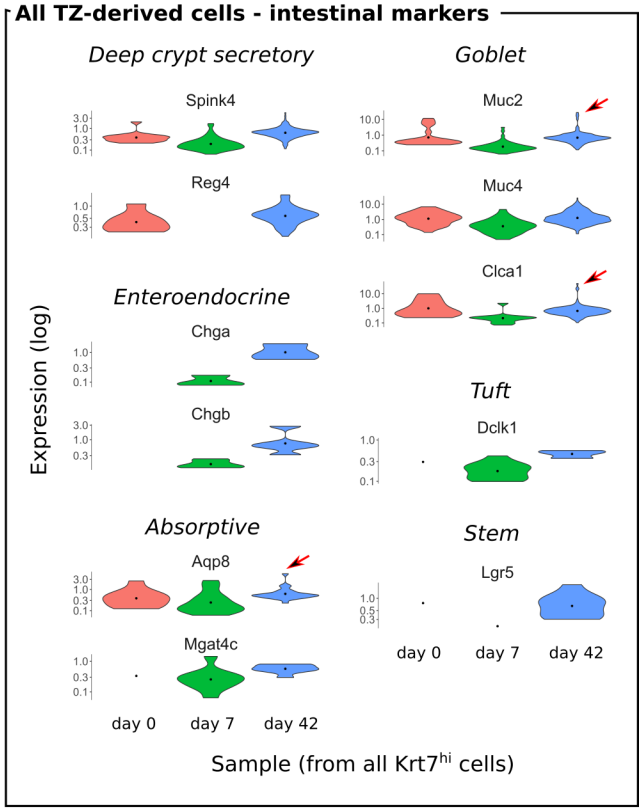
SNEC cells exhibit elevated expression of claudin-1 (CLDN1). Related to Figure 6. A) Immunohistochemical labeling of CLDN1 at the uninjured anorectal junction shows staining in rectal crypts and the TZ. B) Staining after DSS-induced injury shows elevated expression of CLDN1 protein in SNEC relative to anal epidermis and the original TZ. Orange arrows mark the dentate line.

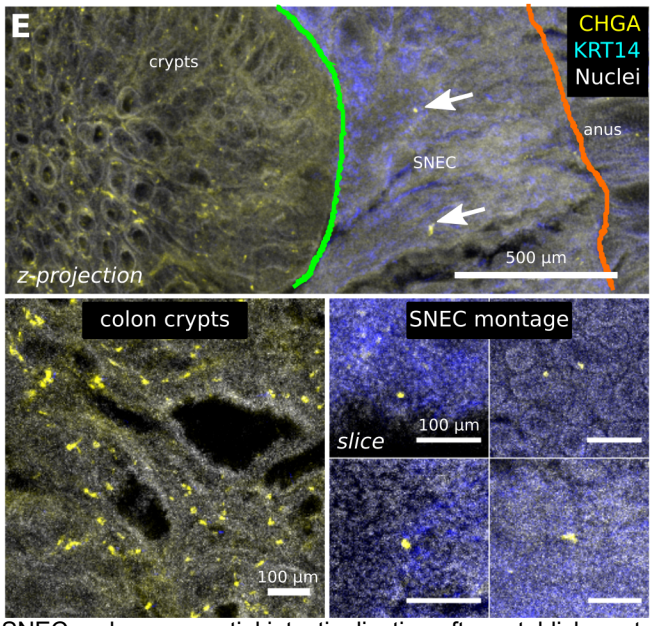
Figure S11









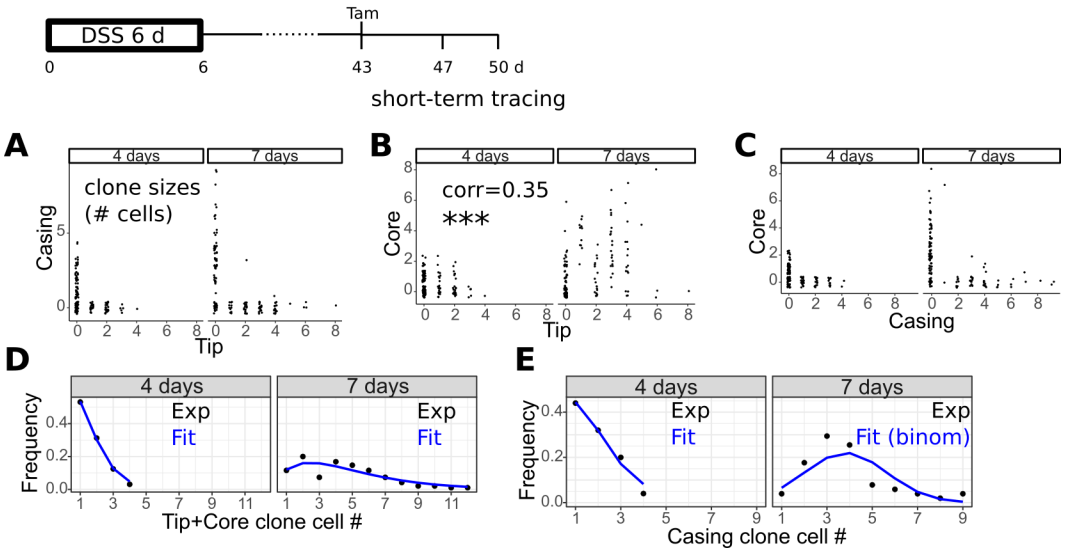


SNEC undergoes partial intestinalization after establishment. Related to Figure 6. A) UMAP of the basal cell cluster of Krt7^{hi} cells shows displacement and transcriptional remodeling of basal cells in SNEC (day 42) compared to the TZ (day 0). B) Pathway analysis of differentially expressed genes in basal cells at exp d 42. Upregulated are small-intestinal genes and inflammatory signals such as TNF;

downregulated are epidermal genes and mitotic signals. C) Violin plot of specific transcripts upregulated in SNEC basal cells. These included EGFR ligands (Areg, Ereg, Hbegf), inflammatory signals (Cxcl10), regulators of proliferation and stemness (Dynap, Lgr6, Fosl1), and markers of deep crypt secretory cells (Spink4). D) Violin plots of expression of “classic” colonic epithelial differentiation markers in all Krt7^{hi} cells stratified by sample. Red arrows point to subsets of SNEC cells with log-fold increase in expression compared to the TZ. The evidence overall supports enrichment of colonic differentiation transcripts, although cells with paramount expression were still quite rare. E) Whole-mount immunostaining of the anorectal region at exp d 82 with antibodies targeted against CHGA (yellow) and KRT14 (blue). In colonic crypts, we found labeling of relatively abundant CHGA+ enteroendocrine cells; in SNEC, we found <5 CHGA+ cells per animal (n=4 mice). White arrows in the maximum intensity projection (upper panel) demonstrate labeled CHGA+ SNEC cells. Green and orange lines denote the new squamocolumnar junction and the original dentate line, respectively. Each panel in the SNEC montage shows a labeled CHGA+ SNEC cell obtained from a different mouse.

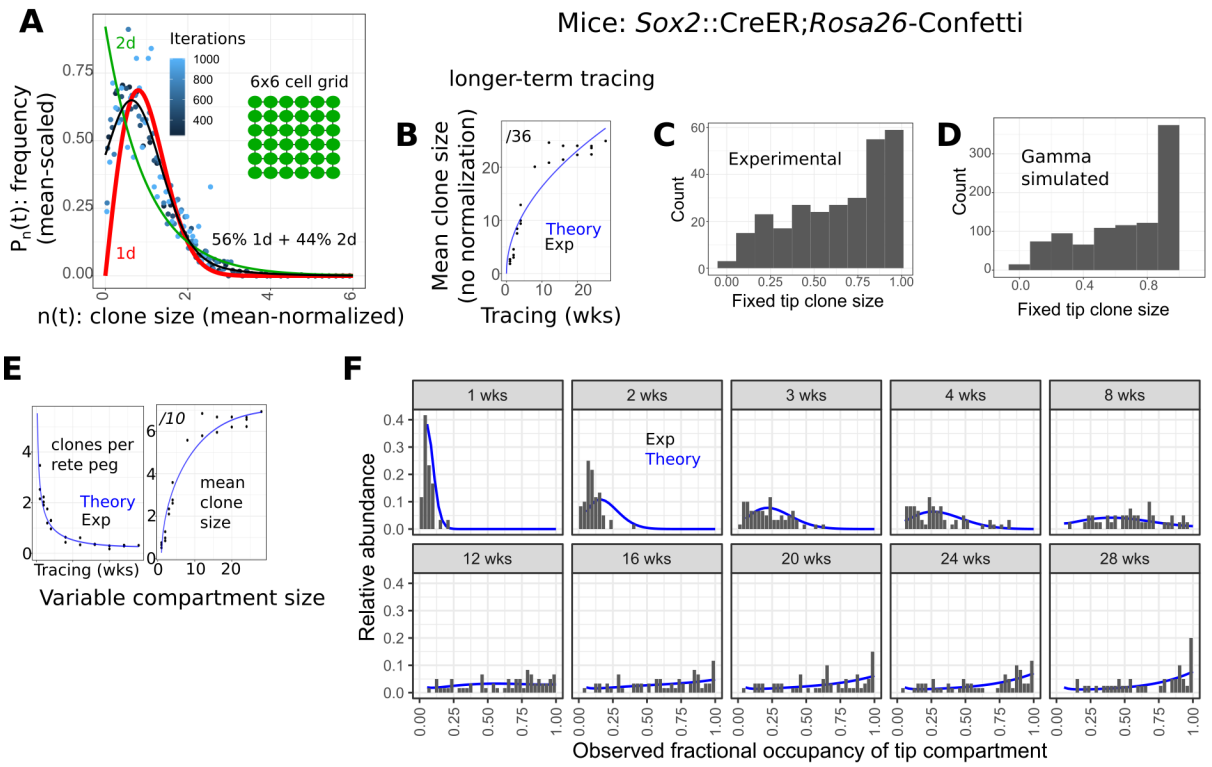
Figure S12

Mice: *Sox2*::CreER;*Rosa26*-Confetti



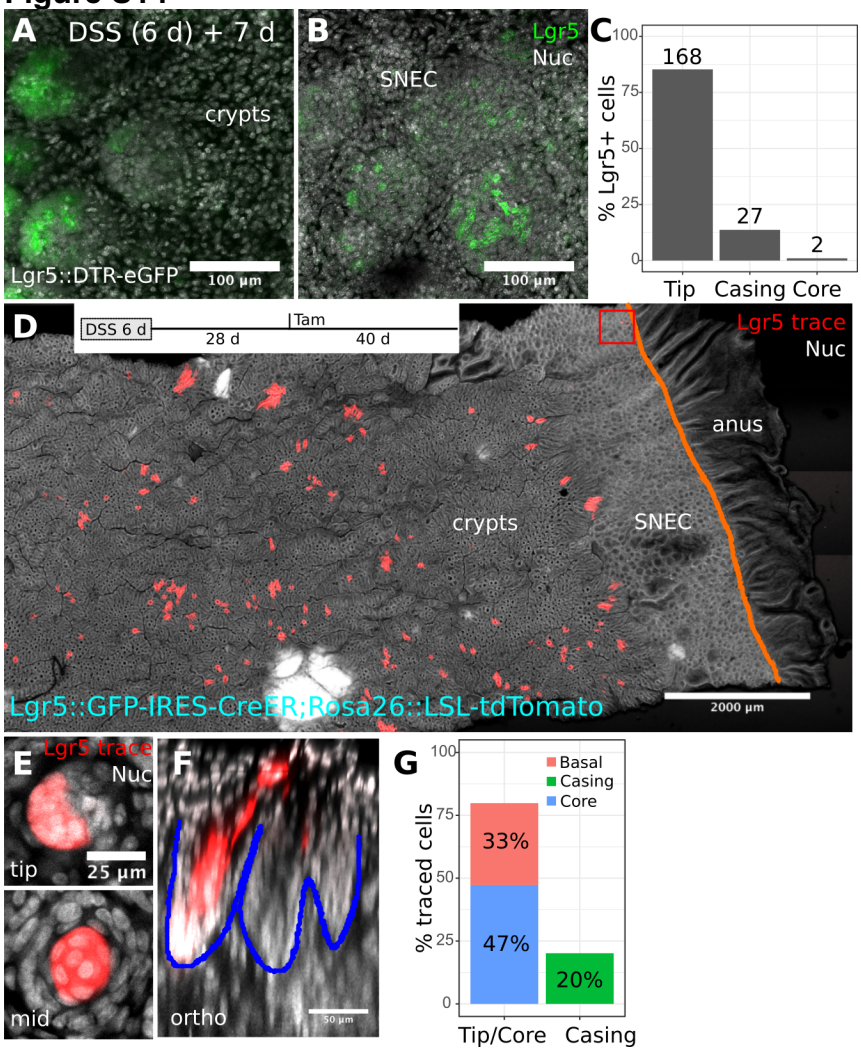
Short-term lineage tracing of progenitor cell populations in SNEC. Related to Figure 7. A-E) Short-term lineage traces (4 and 7 days, 0.5 mg tamoxifen per mouse, n=3 mice/timepoint) in SNEC begun 43 d after DSS treatment in Sox2-Confetti mice. The relationships between the number of cells in each color patch in the tip, casing, and core compartments are plotted (A-C). Only tip and core cell occupancy were correlated (corr., Pearson's r), supporting the regeneration of core cells by tip cells and separate function of the casing domain. The size distributions of tip+core (D) and casing (E) cells were plotted and show smooth distributions that are fit by exponential or binomial functions.

Figure S13



Modeling long-term regeneration in SNEC with variable tip compartment sizing. Related to Figure 7. A) Simulation of clone sizes emerging at different times (iterations) from random replacement within an equipotent 6x6 square grid of cells; the time-invariant scaling form is shown (see “Methods”). Diagonal replacement is not allowed in the simulation. A 1d scaling function provided a good fit to the simulated clone sizes. B) Fitting of the theoretical evolution (λ/N^2=0.008 /wk) of the un-normalized mean clone size over longer-term (1-28 wks) tracing. The fitting at later timepoints (>12 wks) is suboptimal because the mean clone fractional size converges to ~0.6 (<1). C) Shown is the distribution of empirical tip clone sizes expressed as a fraction of the total tip domain identified from imaging. The distribution is aggregated from data after 12 wks of tracing. This “fixation distribution” explains why the mean clone size asymptotically approaches 0.6 within the timeframe of the tracing experiment. D) The fixation distribution can be modeled with a simulated gamma function. E,F) Adjusting the neutral drift model to account for heterogeneity in tip compartment sizes provides a good fit to the empirical elimination and growth rate of clones (E), as well as the empirical shape of the tip clonal size distribution (F).

Figure S14



Low expression of Lgr5 by some tip cells in SNEC. Related to Figure 7. A-B) GFP-labeled Lgr5+ cells from Lgr5::DTR-EGFP mice were found in colonic crypts (A) and in certain patches of SNEC rete pegs (B). C) Most Lgr5+ cells were found in the tip cell compartment (n=3 mice). Numbers on the graph denote the absolute number of cells counted from representative photos captured from 3 mice. D) Lineage tracing from Lgr5+ stem cells after DSS injury (in Lgr5::EGFP-IRES-CreER;Rosa26::LSL-tdTomato [Ai14] mice) reveals tracing within SNEC (red box outline). E-G) Consistent with tip cells regenerating core cells, tracing from Lgr5+ cells was mostly restricted to the tip and core domain, as demonstrated in z-plane reconstructions (E), orthogonal projections (F), and quantification of cell position (G) (n=20 color patches).