

Tissue and cellular spatiotemporal dynamics in colon aging

Received: 16 April 2024

Accepted: 22 August 2025

Published online: 22 October 2025

 Check for updates

Aidan C. Daly^{1,2,13}, Francesco Cambuli^{1,11,13}, Tarmo Äijö², Britta Lötstedt^{1,3,4}, Nemanja Despot Marjanovic^{3,5}, Sara Fernandez¹, Olena Kuksenko^{3,6}, Matthew Smith-Erb¹, Daniel Domovic¹, Nicholas Van Wittenberghe³, Eugene Drokhyansky³, Gabriel K. Griffin^{3,7}, Hemali Phatnani^{1,6}, Richard Bonneau^{2,8,12}✉, Aviv Regev^{3,5,12}✉ & Sanja Vickovic^{1,3,9,10}✉

Tissue structure and molecular circuitry in the colon can be profoundly impacted by systemic age-related effects but many of the underlying molecular cues remain unclear. Here, we build a cellular and spatial atlas of the colon across three anatomical regions and 11 age groups, encompassing ~1,500 mouse gut tissues profiled by spatial transcriptomics and ~400,000 single nucleus RNA-sequencing profiles. We develop a computational framework, cSpotch, which learns a hierarchical Bayesian model of spatially resolved cellular expression associated with age, tissue region and sex by leveraging histological features to share information across tissue samples and data modalities. Using this model, we identify cellular and molecular gradients along the adult colonic tract and across the main crypt axis and multicellular programs associated with aging in the large intestine. Our multimodal framework for the investigation of cell and tissue organization can aid in the understanding of cellular roles in tissue-level pathology.

A typical colon extends >12 cm in mice and >1.5 m in humans^{1,2}, with considerable variance in length, thickness and folding, impacted by multiple variables, including age, sex, weight and diet. The inner lumen of the large intestine is punctuated by millions of invaginations, each harboring a colonic crypt, the key anatomic unit responsible for its continuous regeneration and differentiation³. Underlying the mucosal epithelium, the submucosa hosts lymphoid clusters, nerve fibers and the lymphovascularity, while the outer muscular wall enables peristaltic motility. From the cecum to the rectum, the colon carries out spatially confined regional functions, which emerge postnatally and are required for the digestion of solid food^{4,5}. During aging, the decline of colonic function is accompanied by dysbiosis and excessive epithelial

permeability, allowing the gut microbiota to infiltrate the lumen⁶ and causing a generalized and protracted inflammatory state⁷, as well as the emergence of common pathologies, including constipation, diverticulitis, malnutrition and colorectal cancer (CRC)⁸.

Despite the crucial importance of colon function, the cellular and molecular features associated with functional diversity across colonic regions, the crypt axis and major lifespan stages have not yet been comprehensively characterized. In recent years, single-cell and single-nucleus profiling of the mouse and human intestines^{9–20} has discovered and classified cell types and functions in the gut during development^{15,20}, in adults^{16,17,19} and in aging^{18,21} but with limited spatial context. Spatial *in situ* profiling methods^{22–33} are poised to address

¹New York Genome Center, New York, NY, USA. ²Center for Computational Biology, Flatiron Institute, New York, NY, USA. ³Klarman Cell Observatory Broad Institute of MIT and Harvard, Cambridge, MA, USA. ⁴Science for Life Laboratory, Department of Gene Technology, KTH Royal Institute of Technology, Stockholm, Sweden. ⁵Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. ⁶Department of Neurology, Columbia University Irving Medical Center, New York, NY, USA. ⁷Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA. ⁸Center for Data Science, New York University, New York, NY, USA. ⁹Department of Biomedical Engineering and Herbert Irving Institute for Cancer Dynamics, Columbia University, New York, NY, USA. ¹⁰Science for Life Laboratory, Department of Immunology, Genetics and Pathology, Beijer Laboratory for Gene and Neuro Research, Uppsala University, Uppsala, Sweden. ¹¹Present address: Human Technopole, Milan, Italy. ¹²Present address: Genentech, San Francisco, CA, USA. ¹³These authors contributed equally: Aidan C. Daly, Francesco Cambuli. ✉e-mail: bonneau.richard@gene.com; aviv.regev.sc@gmail.com; svickovic@nygenome.org

this gap, through either targeted or genome-wide profiling. However, robust computational frameworks for spatial analysis of large tissue cohorts are still lacking. Both the size and the heterogeneity of the organ pose notable analytical challenges, including how to infer molecular information at the single-cell level, how to integrate imaging and molecular data and how to perform biologically meaningful comparisons between functionally relevant tissue domains across space and time. For example, many spatial analysis methods reduce noise by smoothing gene expression data across neighboring spots or cellular neighborhoods at the risk of true signal loss^{34–36}. Most methods for testing spatial differential expression (DE) through clustering^{37,38} or explicit modeling^{39,40} are applied only to single tissue sections and those that integrate data from multiple tissue sections^{41–44} are usually limited to the alignment of serial sections from the same tissue. Other methods, focused on deconvolution of multicellular spatially resolved measurements to the single-cell level (through Bayesian modeling^{45–47}, non-negative matrix factorization (NMF)⁴⁸ or deep learning⁴⁹), typically use little or no information about tissue anatomy or histology, which may yield biologically unrealistic results and limit their efficacy^{50,51}. Furthermore, while a common coordinate system⁵² can facilitate integration, it is more challenging in large tissues such as the colon that lack a strict stereotypical architecture. Thus, to characterize the molecular and cellular variation underlying functional variation in the colon, both spatial profiling of large tissue cohorts and computational means to integrate these data across space and age are needed.

Here, we created a comprehensive experimental and computational framework to construct an integrated cell and tissue atlas of the mouse colon across temporal, anatomical and morphological variation, by combining spatial transcriptomics (ST)^{31,53} and single-nucleus RNA sequencing (snRNA-seq)^{17,54}. We define the relative abundance of cell types using multimodal estimation, address missing data imputation and technical noise correction through information sharing across tissue sections and use explicit, hierarchical modeling of spatial and covariate-specific effects on cellular gene expression in Bayesian hypothesis testing across cell types, tissue regions, ages or other covariate groups informed by both snRNA-seq and ST data. Our work provides insights into tissue-level and cell-level function and organization and serves as an important reference for understanding the biology of aging.

Results

A spatiotemporal atlas of the colon

To build a comprehensive atlas, we collected colonic specimens from proximal to distal anatomical regions through three major phases of the mammalian lifespan: juvenile (<4 weeks of age in the mouse), adulthood (6–12 weeks) and aging (6 months–2 years) and profiled them by ST⁵³ and snRNA-seq⁵⁴ (Fig. 1a). The spatial branch comprised ~1,500 sections ($n = 65$) and ~66,500 spatially barcoded spots, each quantifying the expression of 12,976 genes and sampling, on average, 24 tissue sections, 977 spots and 35,730 annotated cell segments from each mouse colon (Fig. 1b). These spanned 66 conditions across combinations of age (11 time points), sex (male and female), colonic region (proximal, middle and distal) and morphological regions of interest (MROIs; 14 histological annotation categories) (Fig. 1c and Supplementary Table 1). The cellular branch of the atlas encompassed a total of ~400,000 snRNA-seq profiles ($n = 21$), integrating newly generated and publicly available data, all collected in our lab (Extended Data Fig. 1a–c and Methods), which we partitioned and annotated into 17 major subsets of epithelial cells (intestinal stem cells (ISCs), transamplifying cells (TAs), cycling TAs, colonocytes, goblet cells, neuroendocrine cells and tuft cells), immune cells (B cells, T cells and macrophages), stromal cells (vascular cells, lymphatic cells, fibroblasts and trophocytes), enteric neurons, glia and smooth muscle cells (SMCs)^{17,21,55} (Fig. 1d, Extended Data Fig. 1d,e, Supplementary Table 2 and Methods).

cSpotch infers cell type compositions and gene expression rates from ST, histology and snRNA-seq

To accurately detect spatial gene expression changes across multiple tissue samples, individuals and conditions in our atlas, we developed cSpotch⁵⁶, a novel hierarchical Bayesian probabilistic model that uses both histological images and snRNA-seq to infer location-dependent and covariate-dependent cell-type-specific gene expression profiles from multicellular ST data (Fig. 1e). cSpotch consists of two major steps. First, it infers the cellular composition in each spatial spot from a multitissue dataset using a two-tier histological annotation schema, where MROIs, such as crypt apex and crypt base, are further labeled with cell-level morphological data, such as epithelial and muscle cells. These labels are then used for molecular integration with snRNA-seq profiles in a cell type deconvolution task. Second, cSpotch uses the deconvolved cell type compositions to infer MROI-specific and covariate-specific expression rates for each gene in each cell type. These expression rates can be used to test for DE across location (for example, crypt base versus apex), condition (for example, proximal versus distal colon or juvenile versus old mice), or cell type (for example, goblet versus T cell) in the entire atlas.

Specifically, to infer cell composition, we first annotated each spatial spot with a single (mutually exclusive) MROI label (Methods), segmented nuclei from the histology image and annotated each nuclear segment with one of five morphological cell superclass labels (Fig. 2a), using a conditioned semantic segmentation workflow with structural, anatomical and neighborhood features (Extended Data Fig. 2, Supplementary Table 3 and Methods). Across held-out test sets for young (≤ 4 weeks old) and adult/aged (≥ 6 weeks old) mice and all anatomical regions, we correctly assigned 85% of all nuclear areas compared to pathologist superclass annotations as the ground truth (Extended Data Fig. 2c,d and Methods). Next, for each spot, cSpotch uses NMF^{52,57} to infer a combination of snRNA-seq profiles whose aggregate marker gene expression profile fits the observed expression measurement but in a manner constrained by the morphological cell composition inferred from histology in the previous step (Methods and Supplementary Table 4). Without introducing any morphological cell composition constraints ($\alpha = 0$), while gene expression profiles were well reconstructed (Extended Data Fig. 3a, blue line), the morphological cell type proportions were not, resulting in large numbers of deconvolved transcriptional cell types assigned to each spot (Extended Data Fig. 3b,c, blue lines). Once constrained by the inferred morphological cell type proportions ($\alpha > 0$; Extended Data Fig. 3c, orange, green and red lines), the cell composition reconstruction improved, without loss of expression reconstruction accuracy. cSpotch performed well in both decomposing the expression profiles and recovering cell compositions consistent with pathologist annotations and known features of the tissue. For example, cell composition estimates for 88 spots randomly selected from 31 tissues (8 weeks) spanning all colonic regions (11 proximal, 12 middle and 8 distal) and MROIs agreed well with ground-truth pathologist annotations and well-known patterns of tissue organization in terms of cell type proportions (Fig. 2a,b). These included high SMC content in the muscularis regions, high goblet cell content in epithelial layers and high B cell content in Peyer's patch (PP) (Fig. 2a).

To infer gene expression rates, cSpotch next uses these inferred cell compositions, MROI annotations and sample covariates to fit a generalized linear model (GLM) of spatial gene expression across the atlas (Extended Data Fig. 4 and Methods). The model performs Bayesian inference to apportion the aggregate expression of each gene in each spot to the contributing cell type(s) in the spot (characteristic expression rate (β); hierarchically formulated to account for covariate-driven variation), along with the effects of neighboring spots (through spatial autocorrelation (ψ)) and spot-specific random effects (ϵ). The three hierarchical levels of the model—age, colon region and sex—are ordered by decreasing levels of hypothesized importance for gene expression

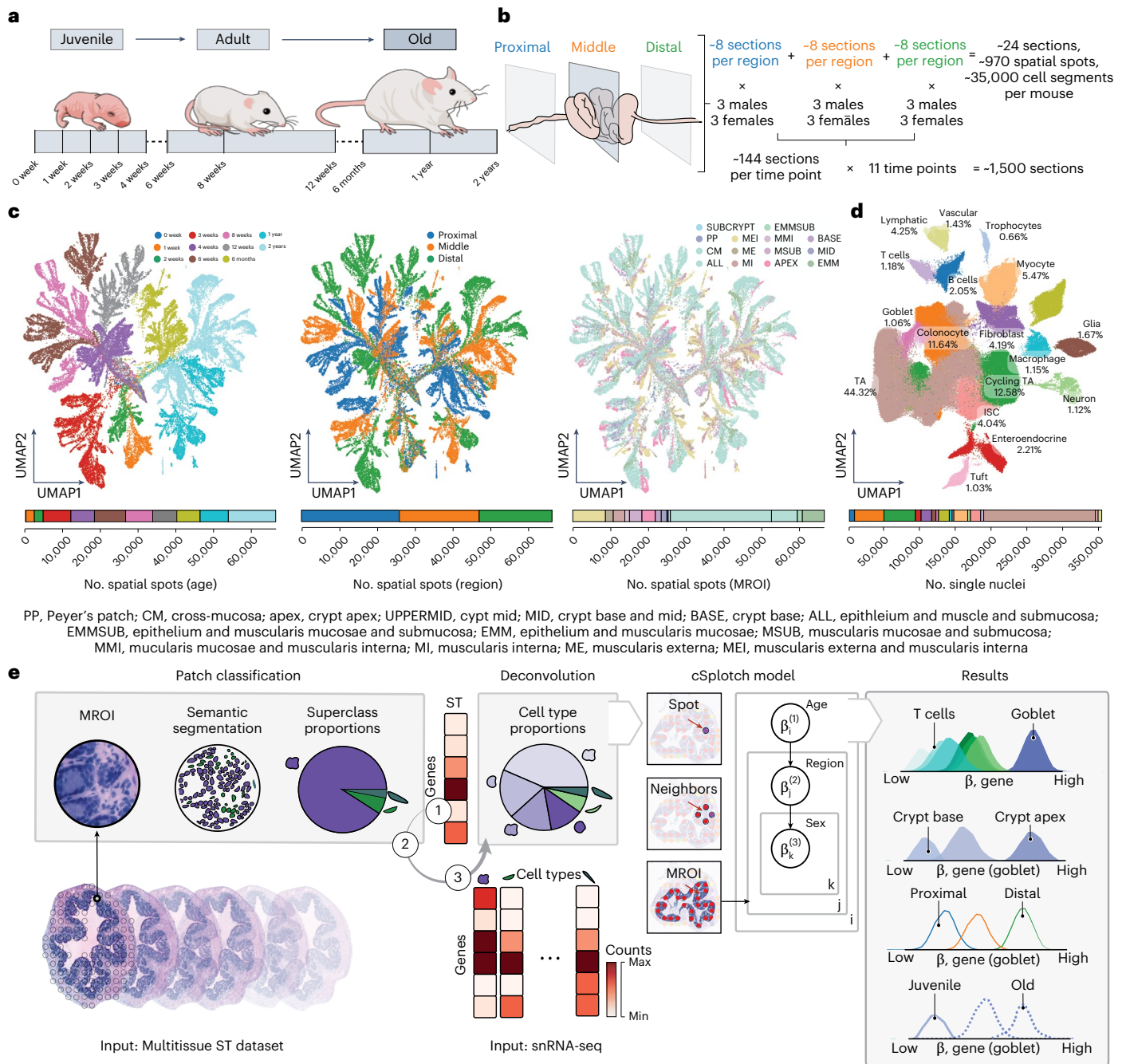


Fig. 1 A cellular and spatial atlas of the mouse colon across regions and ages.

Study design overview. **a**, Sampling time points: birth/juvenile (0–4 weeks), adulthood (6–12 weeks) and aging (6 months–2 years). **b**, Sampling regions: proximal, middle and distal colon. **c**, Spatial atlas. The *t*-distributed stochastic neighbor embedding (*t*-SNE) embedding of 66,481 ST spot profiles colored by age group (left), colon section (center) and MROI (right), along with relative abundances (stacked bars; bottom). **d**, Single-nucleus atlas. The *t*-SNE of

352,195 snRNA-seq profiles colored by cell type cluster (top) and their relative abundances (stacked bars; bottom). **e**, Analysis. Multitissue dataset (bottom left) served as input to cSpotch, which uses histological (MROI) and semantic segmentation and expression (ST and snRNA-seq priors) data to estimate the abundance of each snRNA-seq cell type in each spatial spot and then uses hierarchical Bayesian modeling (cSpotch model) to infer cell-type-specific gene expression conditioned on age, region, sex and MROI annotation (results).

variation. Additionally, to integrate the spatial samples in the context of a common coordinate framework, we used the 14 MROI categories that were manually assigned by the histology at each spatial spot (Methods).

We validated cSpotch’s inferred expression rates in our colon aging atlas data. Firstly, for highly expressed genes (~10% of the genes in the study, detected in >45% of all spots; Fig. 2c, red; for example, *Tff3*, *Ceacam1* and *Acta2*), the expression estimates of cSpotch were highly correlated to those obtained by transcripts per million (TPM) normalization and the spatial autocorrelation and spot-level variation

components of cSpotch captured a substantial portion of the variance (Fig. 2d–f). Secondly, cSpotch recovered the correct spatial DE patterns in each MROI compared to immunofluorescence (IF) staining for four selected marker genes: the highly expressed *Epcam* and *Tff3* (enriched in crypt apex), *Acta2* (*Sma*; enriched in the muscularis interna (MI)), as well as *Cd48*, expressed in only 2.7% of spots (enriched in the PP) (Fig. 2g–i). Thirdly, the characteristic expression levels inferred for each cell type in a given MROI were most correlated with external snRNA-seq profiles from the expected cell type, for types present at

both high (for example, vascular) and low (for example, goblet) frequencies (Fig. 2j). Fourthly, cSpotch correctly assigned genes with known cell-type-specific expression, such as *Prdx6* (goblet cells⁵⁸) and *Pdgfra* (fibroblasts⁵⁹), in the cross-mucosa (CM) and the muscularis externa (ME) and MI (MEI) region, confirming the accuracy of cSpotch in describing gene expression in a complex tissue region comprised of multiple cell types of varying proportions (Fig. 2k–m). Fifthly, cSpotch's accurately deconvoluted simulated ST data, generated by constructing spots using mixtures of snRNA-Seq profiles (Supplementary Table 5 and Methods). cSpotch successfully reconstructed mean profiles for snRNA-seq (Extended Data Fig. 5a), morphological cell type clusters (Extended Data Fig. 5b and Methods) and DE patterns detected from snRNA-seq (Extended Data Fig. 5c and Methods; 94–100% agreement on DE genes), even when the cellular compositions used to deconvolve the simulated data were corrupted by Gaussian noise to reflect imperfect cell type annotations (Extended Data Fig. 5a,b). Sixthly, cSpotch performed better than SpatialDE2 and Spark-X, two methods^{40,60} that analyze ST data at the level of individual tissue sections, on the task of identifying spatially variable genes. Specifically, SpatialDE2 and Spark-X spatial DE statistics for four MROI-specific genes (*Tff3*, *Epcam*, *Acta2* and *Cd48*) vary greatly from section to section (Extended Data Fig. 6), with positive spatial DE determined only for a relatively low fraction of sections ranging from 1.3% (*Cd48*) to 15% (*Tff3*) in SpatialDE2 and 0.5% (*Cd48*) to 38% (*Epcam*) in Spark-X. These low detection rates may be because of the low signal-to-noise ratio and insufficient spatial sampling of individual tissue sections (particularly in PPs), limitations that cSpotch overcomes through hierarchical and multislice integration. Lastly, we estimated the impact of the number of individuals and tissues profiled on statistical power, by incrementally downsampling combinations of individuals ($n = 1, \dots, 6$ of 6) and tissue sections ($n = 2, \dots, 8$ of 52). Even when using one animal and two tissue sections, cSpotch captured meaningful expression signals, as reflected by the low Kullback–Leibler (KL) divergence from the posterior distributions derived using full data for that animal's covariate group (age and region), and accuracy improved significantly with at least four mice (Extended Data Fig. 7a and Methods). Increasing the number of mice improved the estimation independent of expression level, whereas increasing the number of tissue sections per mouse improved the estimation of lowly expressed genes (Extended Data Fig. 7b and Methods). In our full dataset, we sample at least five mice per time point and a minimum of nine tissue sections per mouse, exceeding these thresholds. We confirmed the lesser importance of sex relative to age and colon region on expression variation, finding that at most 140 genes (0.01%) show DE across sex

in any age or MROI (Supplementary Table 6) and that gene expression exhibits similar levels of variation within each sex group across all genes and conditions (coefficient of variation (5th, 95th) percentiles: (0.056, 5.46) for male and (0.057, 5.47) for female). Notably, cSpotch is compatible with higher-resolution ST technologies (for example, Visium and Visium HD; Methods) and recapitulates key spatial results (Extended Data Fig. 8 and Methods).

Cell composition and cell-type-specific expression across the proximal–distal colonic axis correlates with functional variation

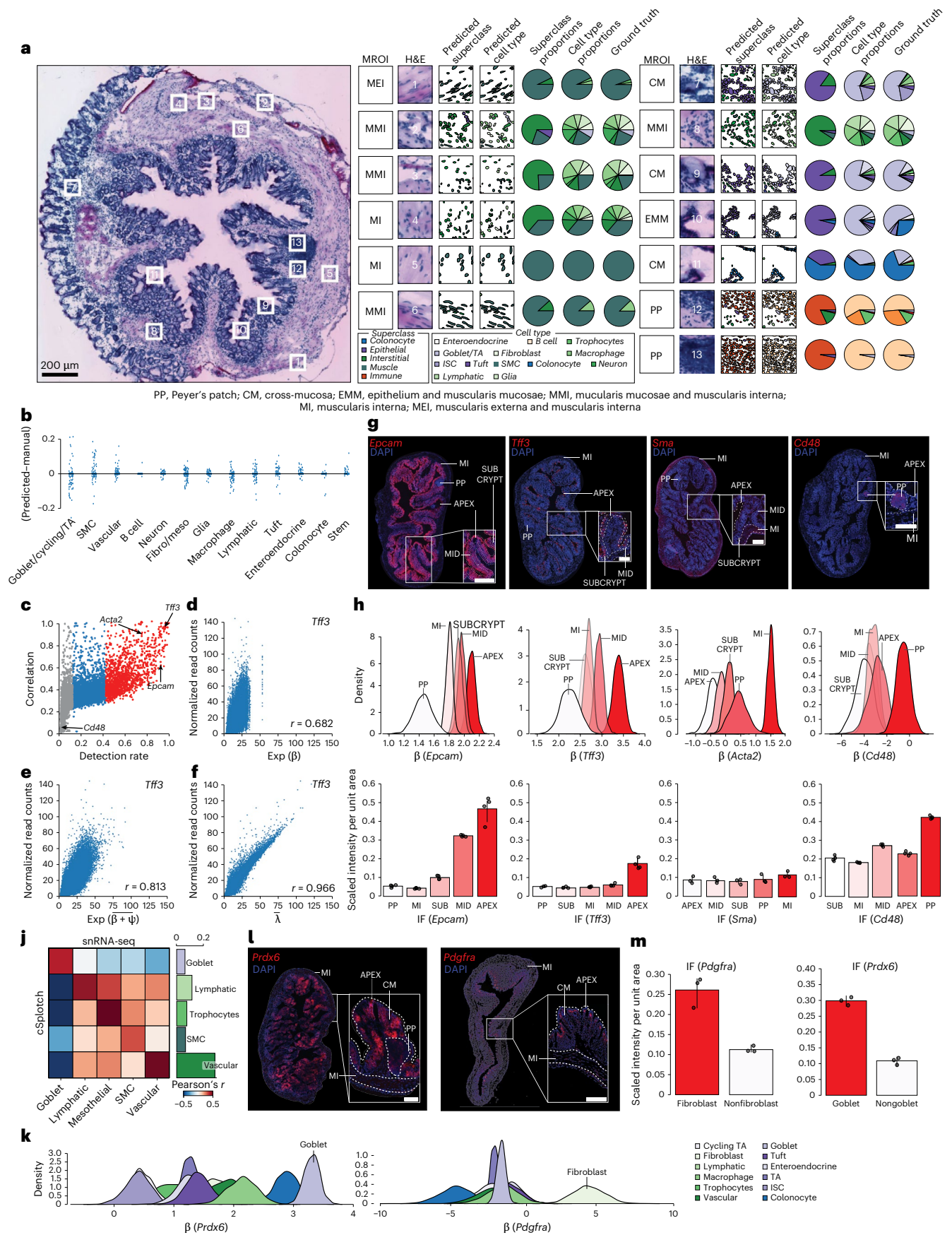
To better describe the variation in tissue structure and function across the colonic tract, we first applied cSpotch to identify tissue-scale changes in cellular composition and spatial cellular gene expression along the proximal–distal axis in the adult (12 weeks) mouse colon. We analyzed 134 sections (43 proximal, 43 middle and 48 distal; 6,399 spots, ~55 nuclei per spot) across six age-matched and gender-balanced mice (three males and three females) (Fig. 3a). Relying on the classification in 14 MROIs, we estimated cell abundance and cell-type-specific expression of 17 distinct cell types.

A comparison of cell abundance in each MROI class across the proximal, middle and distal colon (Fig. 3b) showed widespread variation along the regions, especially in the transition from the proximal to the middle colon (Supplementary Table 7). Among the MROIs with the highest number of statistically significant differences in cell type abundance were the CM (significant regional differences in seven cell types) and the MEI (five cell types) (Benjamini–Hochberg-corrected Welch's *t*-test, $P < 0.05$; Fig. 3c,d). In the CM, goblet cell proportions increased and TA frequencies decreased from the proximal to the distal colonic segment, whereas ISC proportions increased distally. In the MEI, SMC abundance declined distally, whereas fibroblasts, lymphatic cells, macrophages and vascular cells grew in frequency. These observations shed light on the cellular transitions associated with the change in colonic function along its longitudinal axis, with most digestive and absorptive functions occurring in the proximal colon, which handles final steps through peristalsis and bacterial enzymes, breaking down and absorbing residual starch and lipids, compared to water reabsorption, fecal compaction and storage in the distal colon. We compared our results to a previous estimate of cell type proportions along the human intestine, albeit by an earlier study that was not powered to statistically test longitudinal variation across subregions of the large tract¹⁹. Nevertheless, in both studies, goblet cells in the mucosa rose distally, whereas, in the MEI, SMCs increased and endothelial cells decreased in humans but not in mice. Thus, if adequately powered,

Fig. 2 | Deconvolutional modeling using histological and expression features.

a,b, Agreement of morphology-informed deconvolution with manual cell type annotation. **a**, A total of 13 H&E patches from an example (8 weeks, distal) ST array (left; numbered boxes; scale bar, 200 μ m) with their (from left) MROI annotation, H&E image patch, semantic segmentation of morphological superclasses (color code, bottom left) and snRNA-seq cell types (color code, bottom right) and relative proportions of predicted superclasses (pie charts), predicted cell types and manually labeled cell types ('ground truth'). **b**, Error between predicted and manually calculated cell type fraction (y axis) for 13 cell types (x axis) evaluated across $n = 88$ manually annotated spots selected from 31 tissues (8 weeks) spanning all three colonic regions (11 proximal, 12 middle and 8 distal). **c–f**, cSpotch model validation. **c**, Pearson correlation coefficients (y axis) between the expression rate estimated by cSpotch (λ) and TPM-normalized values in ST for the same spots scattered against the detection rate in ST (percentage of spots with nonzero measurements, x axis) for each of 12,796 genes (dots). Red and gray colors denote the top 10% and bottom 50% in detection rate, respectively. **d–f**, Variance in *Tff3* gene expression (y axis, normalized read counts) across spots ($n = 66,481$ dots) explained by region alone (**d**; mean ($\exp(\beta)$), x axis), region and location (**e**; mean ($\exp(\beta + \psi)$), x axis) or all components of the cSpotch model (**f**; mean $\bar{\lambda}$, x axis). Bottom right, Pearson's r . **g–i**, IF validation of spatial gene expression trends predicted by cSpotch. **g**, IF of four gene

products (*Epcam*, *Tff3*, *Sma* and *Cd48*, also highlighted in **c**) in proximal colon sections. Insets, zoomed-in views of a single colon crypt. Dotted lines denote the boundaries between MROIs: crypt apex, crypt mid, subcrypt, PP and MI (*Epcam*, *Tff3* and *Sma* scale bars, 200 μ m; *Cd48* scale bar, 100 μ m). **h**, Probability density (y axis) for posteriors over regional rate terms (β ; x axis) inferred by the cSpotch model for the genes in **g** in each MROI (color code). **i**, Scaled IF intensities per unit area for each gene presented in **g** evaluated in at least three tissue sections (*Epcam*, *Tff3* and *Cd48*, $n = 4$; *Sma*, $n = 3$). The color code is shared with **h**. Error bars, 95% confidence interval (CI); center, mean. **j–l**, Validation of the cellular expression profiles predicted by cSpotch. **j**, Cross-correlation between mean cell type expression profiles derived from two technologies: snRNA-seq (columns) and ST (cSpotch) (rows). The Pearson correlation (color scale) between mean external snRNA-seq profiles (columns) and mean cSpotch β s (rows) was calculated over the union of the top 50 cell type marker genes (Methods) for each of the top five cell types (by mean cell fraction) in the subcrypt for adult mice. **k**, Posteriors over cellular rate terms (β) inferred by cSpotch in CM in each abundant cell type (color code) for a goblet cell (*Prdx6*, top) and fibroblast (*Pdgfra*, bottom) marker in the CM region. **l**, IF images of proximal colon sections for the genes (*Prdx6* and *Pdgfra*) in **k**. Insets are the same as in **g**. Scale bar, 200 μ m. **m**, Scaled IF intensities per unit area for each cell type presented in **l** evaluated in $n = 3$ tissue sections. Error bars, 95% CI; center, mean.



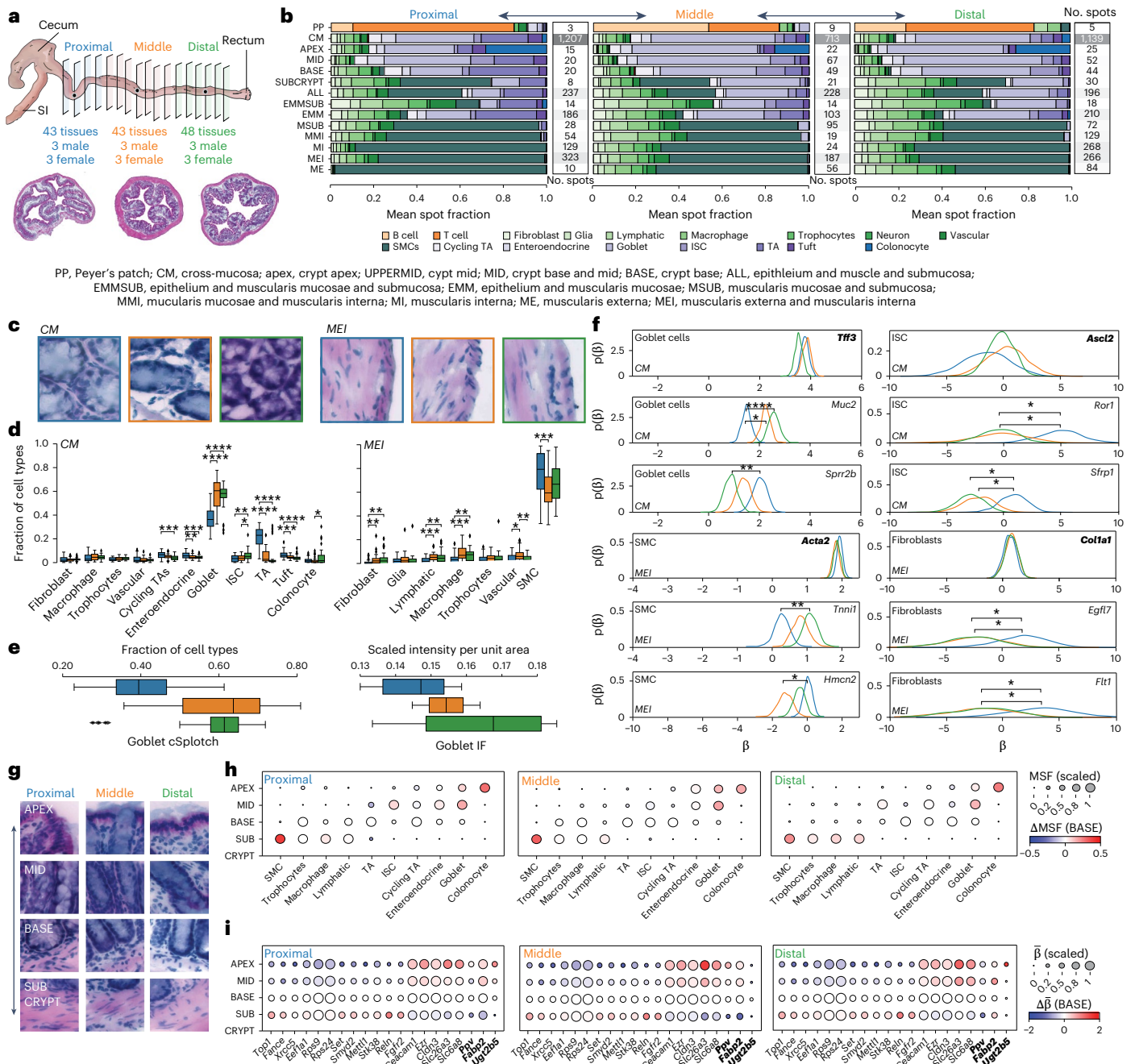


Fig. 3 | Regional differences in cell composition and function in the adult colon. **a**, The atlas spans structural variation in the mouse colon along the proximal–distal axis. Illustrative H&E images of sections from proximal (left; blue), middle (middle; orange) and distal (right; green) regions. **b**, Variation in cellular composition of MROIs across the proximal–distal axis. Proportion of cells of each type (x axis, stacked bars; color code) in each MROI (y axis) in the proximal (left), middle (middle) and distal (right) regions of adult (12 weeks) colon. The number of spots per MROI (color scale) in each colon region is indicated. **c, d**, Variation in cell type composition of the same MROI type along the proximal–distal axis. **c**, Representative H&E image patches from selected ST spots in two MROIs (left to right) from proximal (left; blue), middle (middle; orange) and distal (right; green) regions. In **d**, the fraction of cells (y axis) of each cell type (x axis) in the proximal (blue), middle (orange) and distal (green) regions is shown for each of the MROI types in **c**. Center black line, median; color-coded box, interquartile range; error bars, 1.5× the interquartile range; *0.01 < FDR ≤ 0.05, **10⁻³ < FDR ≤ 0.01, ***10⁻⁴ < FDR ≤ 10⁻³ and ****FDR ≤ 10⁻⁴ (Welch’s *t*-test). Only cell types observed at a rate of 2% or greater across all spots

in an MROI are shown. **e**, Regional goblet cell abundances achieved with cSpotch as presented in **d** and those measured with IF (proximal and distal, *n* = 6; middle, *n* = 2). Center black line, median; color-coded box, interquartile range; error bars, 1.5× the interquartile range. The color code is shared with **d, f**. Cell-specific expression patterns vary across colon regions. Estimated posterior distributions of expression rate (β) for different genes in each colon region (color) in goblet cells and ISCs in CM or SMCs and fibroblasts in MEI. Bold, canonical markers; brackets, significant DE (*BF > 2, **BF > 10, ***BF > 30 and ****BF > 100). **g–i**, Variation in structure, cell composition and gene expression along locations. **g**, Selected H&E image patches from ST across four MROIs (rows) from three colon regions (columns). **h**, Scaled (dot size) and absolute (dot color) change in mean spot fraction (MSF; dot color) of each cell type (columns) in each MROI (rows) relative to crypt base from three colon regions. **i**, Scaled (dot size) and absolute (dot color) change in mean log expression (β) of each gene (columns) in each MROI (rows) relative to crypt base from three colon regions. Bolded genes, regional expression gradients.

a comparison between humans and mice may reveal both similarities and differences, which appear to follow known patterns of evolutionary variance at the anatomical and histological scale. In particular, while the mucosa cytoarchitecture is largely conserved, the external muscular layers are divergent across the two species⁶¹.

At the level of cellular gene expression, cSpotch analysis showed that canonical cell-type-specific markers and processes are similarly expressed across the three colonic regions, according to the posterior distributions of cellular expression rates (Fig. 3f, bold, Extended Data Figs. 9 and 10 and Methods), with stronger cell-type-specific signals than snRNA-seq alone (Supplementary Fig. 11) (*Tff3*, *Clca1* and *Lypd8* for goblet cells^{62,63}; *Ascl2*, *Lgr5* and *Ephb2* for ISCs^{64–66}; *Acta2*, *Cnn1* and *Myh11* for SMCs⁶⁷; *Col1a1*, *Thy1* and *Postn* for fibroblasts⁶⁸), although we acknowledge that this effect may be influenced by snRNA-seq dataset size and/or quality. Other genes had significant cell-type-specific regional differences (Bayes factor (BF) > 2) across the proximal–distal axis (Fig. 3f, nonbold, and Extended Data Fig. 9a,b). For example, in the CM, goblet cells upregulated the antimicrobial genes *Sprr2b* and *Sprr2a3* (ref. 69) proximally and the key gel-forming mucin *Muc2* distally. Such goblet-cell-specific expression patterns are consistent with the role of the proximal colon in controlling microbial fermentation and the requirement for a thick mucus layer as a protective barrier distally⁷⁰. Within the same MROI, ISCs expressed proximally higher levels of canonical Wnt inhibitors (for example, *Sfrp1*) and mediators of the noncanonical pathway (for example, *Ror1* and *Wnt4*). As the strength of canonical Wnt signaling is tightly regulated locally and associated with ISC self-renewal⁷¹, the proximal colon may restrain this pathway and experience a slower cellular turnover. Conversely, the distal colon may require less restrained canonical Wnt signaling to replenish an expanded goblet population, which has the shortest lifespan across mucosal cell types⁷². Such observations are in line with embryological studies on gastrointestinal patterning demonstrating a key instructive role for an anterior-to-posterior Wnt signaling gradient^{73–75}. In MEI, SMCs upregulated *Hmcn2* proximally and *Tnn1* distally. As *Hmcn2* is associated with Hirschsprung's disease⁷⁶ and *Tnn1* helps enforce contractility⁷⁷, their regional-specific expression can support autonomous peristalsis proximally and voluntary contraction distally. In MEI, genes encoding for repressors of vasculogenesis and lymphogenesis, such as *Egfl7*, *Flt1* and *Mdfr*^{78,79}, were upregulated by the fibroblasts proximally, indicating the loss of such mediators as potential mechanism for the expansion of the vascular system distally, where it is required for water reabsorption. Consistent with this, gene set enrichment analysis (GSEA; Methods) of genes differentially expressed between the proximal and distal regions in the CM and MEI showed an enrichment for pathways associated with lipid and sugar metabolism proximally and for ionic exchange distally, while those associated with voluntary contractility, as opposed to autonomous peristalsis, were enriched at the terminal colon (Extended Data Fig. 10c,d). Comparison to the Human Protein Atlas⁵⁸ revealed that 128 of the regionally regulated genes, such as *Tff3*, *Clca1*, *Lypd8*, *Muc2*, *Ascl2*, *Ephb2*, *Myh11* and *Cnn1*, are among the 949 protein markers overrepresented in the human colon from a cohort of healthy individuals, suggesting a conservation of regional expression and likely function across species.

Collectively, we found that the variation in both cell abundance and cell-type-specific gene expression along the colon longitudinal axis correlates with distinct digestive functions along the colonic tract.

Cell type and gene expression gradients along the crypt axis associated with colonic regeneration and functional differentiation

The intestinal crypt is responsible for the continuous regeneration of the highly specialized colonic mucosa³. As ISCs divide at the bottom of the crypt, they migrate toward the apex, differentiating into the main specialized lineages (goblet cells and colonocytes), as well as multiple rarer cell types, including enteroendocrine and chemosensory cells.

A morphogen gradient modulates the balance between epithelial proliferation and differentiation but its composition at different biological scales is not fully characterized⁸⁰. To date, targeted imaging approaches have characterized only a limited number of positionally restricted signaling mediators regulating crypt structure and function⁸¹.

Both cell morphologies and identities displayed distinct zonation patterns across four MROIs of the crypt axis (subcrypt, crypt base, crypt mid and crypt apex) (Fig. 3g,h). The proportions of ISCs, TAs and cycling TAs decrease gradually from the crypt base to apex, goblet cell proportions follow an opposite gradient, colonocyte abundance peaks sharply in the apex and enteroendocrine cells are evenly scattered across the crypt axis. SMCs are almost entirely confined to the subs crypt, whereas lymphatic cells, trophocytes and macrophages gradually decline in a base-to-apex direction. These cell-type-specific distributions are consistent with distinct positional dependencies. Colonocytes and SMCs are localized in close proximity to physical barrier domains (for example, the colonic lumen and the lamina propria), whereas the distribution of epithelial cells (ISCs, TAs and goblet cells), immune cells (macrophages) and some mesenchymal cells (trophocytes and lymphatic cells) is consistent with a reliance on overlapping signaling cues from the opposite crypt poles. Lastly, the homogeneous scattering of enteroendocrine cells may reflect the distribution of enteric nerve fibers.

At the molecular level, 321 genes had significant, monotonic variation across the crypt regardless of the colonic region (Fig. 3i, Supplementary Table 8 and Methods). Some of these genes were significantly associated (BF > 2, log₂ fold change > 0.5; Methods) with specific cell types in a given spatial niche. For example, many of the genes upregulated in BASE were both expressed in ISCs and implicated in key biosynthetic pathways, including DNA replication and repair (*Top1*, *Fance* and *Xrcc5*), protein synthesis (*Eef1a1*, *Rps9* and *Rps24*) and transcriptional (*Set* and *Smyd2*), translational (*Mettl1*) and post-translational (*Stk38*) regulation (Supplementary Table 9). Elevated biosynthetic activity restricted to basally located progenitors has also been observed across the human intestine of healthy individuals¹⁹ and in CRC⁸². Genes upregulated at the crypt apex were associated with the establishment of an apical polarity domain at the plasma membrane of goblet cells, including cytoskeletal (*Ezr*), cell–cell adhesion (*Ceacam1* and *Cldn3*) and transmembrane transport (*Slc26a3* and *Slc6a8*) genes (Supplementary Table 9). Additionally, receptors (*Fgfr2*) and ligands (*Reln*) mediating the crosstalk between ISCs^{83–85} and the underlying stromal compartment (for example, lymphatic cells and trophocytes) were upregulated in the subs crypt and crypt base. Consistent with this, GSEA showed an enrichment of genes from pathways associated with ribosomes and protein synthesis in the subs crypt, whereas genes associated with tight junctions, nutrient absorption and energy metabolism genes were enriched in the crypt apex (Supplementary Fig. 12a, Supplementary Table 9 and Methods).

Another 118 genes had crypt-oriented gradient expression preferentially in one colonic region (23 proximal, 53 middle and 42 distal). These included genes related to key metabolic functions, including the detoxifying enzyme *Ugt2b5*, which is downregulated in experimental models of colitis⁸⁶, the fatty-acid-binding and importer protein *Fabp2* (ref. 87) and the gut hormone *Ppy*, known to regulate food intake⁸⁸ (Fig. 3i). Consistent with this, genes differentially expressed specifically in the proximal, middle or distal crypt were enriched (by GSEA) for biosynthetic programs in the subs crypt. For IGF signaling, which is implicated in ISC biology⁸⁹, DE genes were enriched in proximal crypts (Supplementary Fig. 12b–d). For programs related to vitamin C, DE genes were enriched in the proximal crypt apex. For vitamin A, DE genes were enriched in the middle colon crypt apex. Thus, each of the three colonic regions may rely on partially distinct molecular gradients associated with different digestive functions. Several such genes, especially those expressed in the highly specialized cells in the crypt apex, such as *Ceacam1*, *Cldn3*, *Slc26a3* and *Fabp2*, also display higher expression in the human colonic lumen^{19,58}. A small subset of

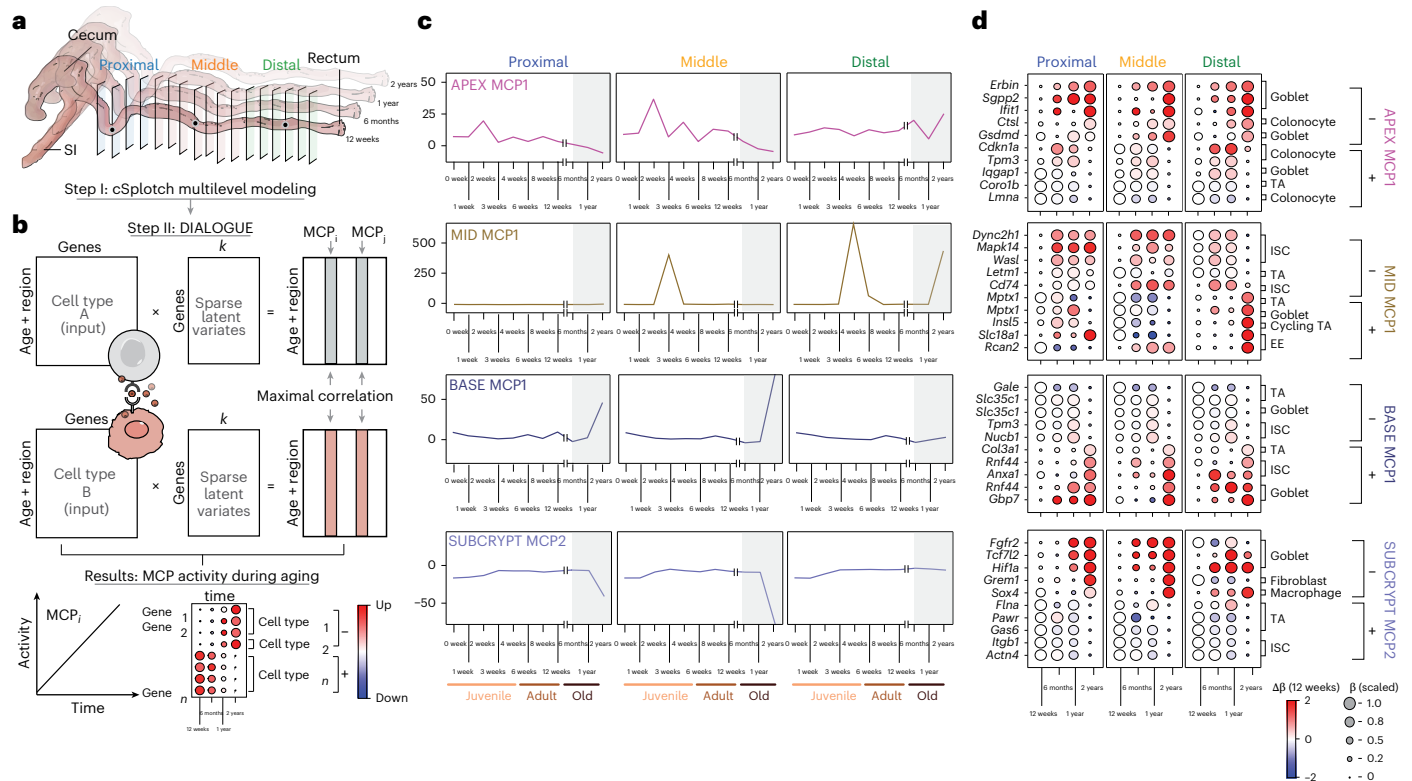


Fig. 4 | Spatiotemporal changes in cell composition and function during colon aging. **a, b**, Analysis approach. cSpotch was used to characterize variability in cell-type-specific gene expression across age and region (a), followed by MCP analysis with DIALOGUE¹⁰⁸ (b) on inferred profiles for abundant cell types (for example, cell types A and B) in each MROI across ages and regions. MCPs are returned as sets of genes with high correlation between two or more cell types across conditions with positively (+, up) and negatively (-, down) contributing

genes from each cell type (Methods). **c, d**, Aging-associated MCPs. **c**, Activity score (y axis) of selected MCPs from different MROIs (top left) for each time point (x axis) and region (top). Gray area: aging window. **d**, Scaled (dot size) and absolute (dot color) change in log expression (β) relative to 12 weeks of each gene (rows) for each time point (columns) for selected upregulated (+) and downregulated (-) genes from the MCPs in c, sorted by their associated cell type (right).

genes presented sex-biased expression across the crypt, including *Guca2a*, encoding the intestinal hormone guanylin⁹⁰, and *Apoa1* and *Cyp3a44*, involved in cholesterol and steroid metabolism^{91,92}, respectively (Supplementary Table 6).

In summary, our analysis revealed cell-type-specific distributions, likely reflecting distinct positional dependencies, and identified a subset of pan-colon or region-specific gradient genes to prioritize mechanistic experiments on the crypt's structure and function.

Spatiotemporal variability of cell type abundances during colon aging

In the elderly human population, the large intestine is affected by common pathological conditions including constipation⁹³, diverticulitis⁹⁴ and malnutrition⁹⁵, a markedly increased risk of CRC⁹⁶ and other features of intestinal senescence^{8,18,97-103}. Yet, aging is a loosely defined condition, which is asynchronously experienced across populations with vastly variable phenotypic impact. Previous studies carried out on small cohorts with few time points, using destructive methodologies, have led to potentially conflicting findings. For example, both loss and gain of secretory cells have been reported to occur in the aging mouse intestine^{18,102,104}.

To characterize intestinal cells during mouse aging, we tracked changes in the proportion of the most abundant cell types (ISCs, TAs, goblet cells and colonocytes) in the four crypt MROIs in each of the colonic regions over time (Supplementary Fig. 13 and Supplementary Table 10). During the juvenile (up to 4 weeks) and adult (4 to 12 weeks) phases, there was limited variation in the ISC fraction over time in the subcrypt, crypt base and crypt mid, consistent with the highly regenerative nature of the colon, both during development and at steady state. In the same compartments, TAs displayed a transient increase during the first 4 weeks

after birth, followed by a decline between 6 and 12 weeks, as they increasingly transitioned toward the secretory lineage. Throughout juvenile and adult life, goblet cells steadily increased across all MROIs, reaching their peak proportion between 8 and 12 weeks. Although the structure and function of colonocytes is impacted by weaning from lactation to solid food starting 3 weeks after birth¹⁰⁵⁻¹⁰⁷, their relative proportion progressively decreased in the crypt apex up to 12 weeks, concomitantly with the increase in goblet cell fraction.

Importantly, there were multiple significant changes in cell composition of the same region between full reproductive maturity (12 weeks) and end of life (2 years) across the crypt axis in the distal and middle colon, as well as, to a lesser extent, in the proximal region (Supplementary Fig. 13, shaded areas, Supplementary Table 10 and Methods). Goblet cell frequency declined from young adult (12 weeks) to aged (2 years) animals in all MROIs of the crypt axis in the mid colon, in the subcrypt, crypt base and crypt mid regions in the distal colon and in the crypt base of the proximal colon. Conversely, the frequency of progenitor cells (ISCs and TAs) increased in old mice (2 years) in both the middle and the distal colon, with TAs increasing with age across the entire crypt axis in the middle colon and ISCs increasing with age in the subcrypt, crypt base and crypt mid MROIs in the distal colon. In the crypt apex of these segments (middle and distal), there was also a marked increase in colonocytes during that period. Hence, cell frequencies are substantially remodeled as animals age, with colonocytes progressively displacing a dwindling goblet population in the crypt apex. This may be especially impactful in the middle and distal mucosa, which rely more heavily on secretory cells at steady state. Goblet cells decline despite the concurrent increase in the proportion of ISCs and TAs in the subcrypt, crypt base and crypt mid. These temporal

dynamics suggest an imbalance between progenitor proliferation and secretory differentiation, providing a cellular basis for the defective regenerative function observed in the elderly.

Variation in cell type abundances associated with activity of multicellular programs (MCPs) during colon aging

To further identify expression changes potentially associated with distinct biological properties across covariate groups, we identified the genes with significant expression differences in the interval of 12 weeks to 2 years for every abundant cell type within each of the four crypt MROI (Supplementary Table 11) and tested the union of all DEGs (across all cell types) within each MROI for functional enrichments by GSEA. Enriched categories only included very broad terms, such as protein kinase signaling, regulation of transcription and vesicular transport, yielding limited insights into specific aging mechanisms (Supplementary Table 11). Because such gene-pooling approaches discard potentially useful information on the cellular source of expression and on cell–cell associations, we next characterized MCPs of genes with coordinated change in expression across multiple mucosal cell types with age. To that end, we applied DIALOGUE¹⁰⁸, which uses multiple canonical correlation analysis (MCCA) to infer MCPs of genes with correlated expression patterns across multiple distinct cell types (Fig. 4a,b and Methods). Each MCP consists of subsets of genes whose mean expression in one cell type is positively or negatively correlated with that of genes in one or more other cell types across time (Supplementary Table 12). Several of the MCPs with the highest degree of correlation between member genes had near-monotonic changes during the aging window (for example, between 6 months and 2 years of age) in at least one colonic region (Fig. 4c, Supplementary Fig. 14, Supplementary Tables 12 and 13 and Methods).

Subcrypt MCP2, whose activity markedly declines at 1 year of age in the proximal and middle colon (Fig. 4d), may help explain some of the substantial changes in cell composition with age in the subs crypt region, where goblet cells decline and progenitors increase at 2 years of age (Supplementary Fig. 13). Specifically, subs crypt MCP2 consists of genes from nine cell types: epithelial (ISCs, TAs, cycling TAs and goblets), macrophages, trophocytes, lymphatic cells, vascular cells and SMCs. These include age-increasing proinflammatory genes in macrophages (*Sox4*)¹⁰⁹ and signaling mediators in goblet cells (*Fgfr2*, *Tcf7l2* and *Hif1a*)^{89,110,111} and trophocytes (*Grem1*)¹¹², many of which are involved in epithelial dedifferentiation^{89,110,113}, inflammation¹¹¹ and malignant transformation^{112,114,115}, especially those associated with colitis^{114,115} (Supplementary Table 13). Age-declining genes included intestinal mechanotransduction (*Itgb1*, *Flna* and *Actn4*)^{116–118}, cell death regulation (*Gas6* and *Pawr*)^{119,120} and tumor suppressor genes (TSGs; *Itgb1*, *Flna*, *Gas6* and *Pawr*)^{116,119,121,122} expressed in ISCs and TAs (Fig. 4d). As the adult distal subs crypt had enhanced Wnt signaling (Fig. 3f), the loss of expression of such TSGs could explain a shift in the balance between self-renewal and differentiation, leading to an expanded stem cell compartment and defective lineage priming. In contrast, the adult proximal subs crypt was enriched for canonical Wnt inhibitors. Here, TSG loss of expression alone may be insufficient to drive a wider progenitor compartment in the absence of a permissive microenvironment. Instead, we observed an upregulation of regenerative signaling mediators (*Fgfr2*, *Tcf7l2* and *Hif1a*), which may eventually promote dedifferentiation of the secretory lineage into progenitor cells. In line with such a hypothesis, the middle colon displayed an intermediate scenario, where the rewiring of goblet signaling was associated with an expansion of the TA fraction.

The activity of the crypt base MCP1 program, which consists of gene expression in ISCs, TAs and goblet cells, increases in the proximal and middle colon following 1 year of age. At 2 years of age, this MROI displayed loss of goblet cells throughout the colon and a gain in progenitors in the middle and distal colon. The crypt base MCP1 genes increasing with age included the key fetal stem cell marker *Anxa1*

(ref. 123), upregulated in ISCs and associated with damage-induced regeneration¹²⁴, *Rnf44* (ref. 125) and *Gbp7* (ref. 126), known to be induced by inflammation and expressed in ISCs and goblets, and the protumorigenic extracellular matrix protein *Col3a1* (refs. 127,128) found in TAs (Fig. 4d). Age-declining genes include metabolic regulators of glycosylated surface proteins and lipids, such as *Slc35a1* (ref. 129) and *Gale*¹³⁰, which are important for immune recognition and pathogen infection, expressed in TAs and goblets, and *Tpm3*, a gene recurrently lost in CRC¹³¹, expressed in ISCs. The activation of damage-induced regeneration genes (*Anxa1*, *Rnf44* and *Gbp7*) suggests that injury-related responses are enhanced with aging predisposing the colon to transformation.

The crypt mid MCP1 program, with genes expressed in ISCs, TAs, goblet cells, enteroendocrine cells and tuft cells, had a different aging-related activity in the proximal and middle versus distal colon. At 2 years, goblet cells are lost and progenitors increase in the middle and distal crypt mid MROI, whereas no significant cell composition changes are observed in the proximal region. Crypt mid MCP1 genes with neuroendocrine and metabolic functions (*Insl5* (ref. 132), *Slc18a1* (refs. 133,134) and *Rcan2* (refs. 135,136)) declined with aging in the proximal and middle colon but increased in the distal colon. This suggested aging-related anteriorization of the longitudinal colonic patterning, an event also associated with the emergence of malignant states^{137,138}. Such reprogramming genes were further coupled with the change in expression of genes involved in mucosal inflammation (*Cd74* and *Dync2h1*)^{139,140}, stress response (*Mapk14* and *Letm1*)^{141,142} and CRC initiation (*Wasl*)¹⁴³. The crypt mid MCP1 was also markedly enhanced in the middle and distal colon between 3 weeks and 4 weeks, at the time when juvenile mice undergo a drastic nutritional shift because of weaning, indicating that aging may aberrantly reactivate a stress state reminiscent of this important developmental transition.

Lastly, the activity of the crypt apex MCP1 program, which includes genes expressed in TAs, colonocytes and goblet cells, declined in the proximal and middle colon (Fig. 4c). MCP genes upregulated with age included inflammation genes related to pyroptosis (*Gsdmd*)¹⁴⁴, proteolysis (*Ctsl*)¹⁴⁵, membrane integrity (*Sgpp2*)¹⁴⁶, cell polarity (*Erbin*)¹⁴⁷ and the DNA damage response (*Jfit1*)¹⁴⁸. Crypt apex MCP1 genes downregulated with aging included cytoskeleton regulators (*Coro1b*, *Iqgap1* and *Tpm3*)^{131,149,150} expressed in TAs, colonocytes and goblet cells, *Lmna*, a gene encoding the nuclear lamin A/C (whose loss is linked to genomic instability, whereas mutant isoforms are found in congenital premature aging syndromes^{151–153}) and the cell-cycle repressor *Cdkn1a* (p21)¹⁵⁴ (Fig. 4d). Thus, the aging colonic crypt apex is in a diffuse inflammatory state, associated with the expression of canonical aging markers and the downregulation of cell-cycle regulators, indicating the acquisition of a senescent state in the apical colonocytes coupled with the loss of tumor suppressor mechanisms.

Discussion

Characterizing spatial patterning in large organs requires us to harmonize and relate molecular and morphological profiles measured by single-cell and spatial RNA-seq methods. Despite substantial progress to address spatial variance in expression and histological features, existing analysis methodologies are still mostly applied to a single tissue section at a time. As a result, they may fail to identify generalizable tissue function or identify how key features vary along anatomical, functional or temporal axes. Here, we developed a comprehensive experimental and computational framework and used it to present the first systematic atlas of colon aging.

Our analysis revealed changes in cell composition and cell-type-specific gene expression across multiple interleaving scales: gross anatomical scale (the longitudinal proximal–distal axis), fine histological scale (the crypt base to apex axis) and temporal scale (ages from newborn to old). Along the proximal–distal axis, these variations correlate with distinct digestive functions along the colonic tract. Along the

crypt axis, we revealed cell-type-specific positional patterns and genes with either pan-colonic or region-specific spatially restricted expression across the main crypt axis. The positionally oriented balance between progenitor self-renewal and lineage differentiation was associated with a base to apex gradient between biosynthetic and cytoarchitectural regulators, prioritizing pathways and genes for functional validation. Along the temporal axis, aging was associated with a pervasive loss of goblet cells and the establishment of a diffuse inflammatory state, spanning expression programs in multiple cells and across the colon, and revealed the specific impact of such a scenario at different levels of the crypt axis and in distinct regions of the colonic tract. In the upper parts of the crypts, the progressive displacement of goblet cells by apical colonocytes or the anteriorization of the colonic metabolic patterning can provide molecular insights toward a mechanistic understanding of common geriatric conditions, such as constipation and malnutrition, paving the way for interventions aimed to support the quality of life and preventing systemic consequences. At the bottom of the crypt, increasing levels of damage coupled with persistent inflammation are associated with loss of TSG expression, excessive expansion of the progenitor compartment and cellular dedifferentiation toward fetal-like states. Such temporal dynamics are consistent with the elevated incidence of CRC premalignant states in the elderly and illuminate specific cell types and genes for guiding preventive and diagnostic strategies.

Using the canonical system of the mouse colon, we demonstrated how contextual spatial and temporal information can help decipher large-scale molecular datasets and how statistical models such as cSpotch can be used to connect tissue architecture with the pathological alterations leading to aging and disease. In this study, histology annotations were performed manually; however, this could be circumvented in the future with novel image classification approaches¹⁵⁵. In the future, we plan to include additional modalities in our atlas, including protein expression to dissect immune cellular states and genome sequencing to measure mutational events and signatures, and increase temporal sampling frequency to further probe changing regenerative dynamics in elderly mice. We believe that this model framework has use in a wide range of tissue systems and hope that it may help to bridge the gap between single-cell and ST studies.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41587-025-02830-6>.

References

- Hounnou, G., Destrieux, C., Desmé, J., Bertrand, P. & Velut, S. Anatomical study of the length of the human intestine. *Surg. Radiol. Anat.* **24**, 290–294 (2002).
- Casteleyn, C., Rebecki, A., van der Aa, A., Simoens, P. & van den Broeck, W. Surface area assessment of the murine intestinal tract as a prerequisite for oral dose translation from mouse to man. *Lab. Anim.* **44**, 176–183 (2010).
- Gehart, H. & Clevers, H. Tales from the crypt: new insights into intestinal stem cells. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 19–34 (2019).
- Cummins, A. G. & Thompson, F. M. Effect of breast milk and weaning on epithelial growth of the small intestine in humans. *Gut* **51**, 748–754 (2002).
- Moeser, A. J., Pohl, C. S. & Rajput, M. Weaning stress and gastrointestinal barrier development: implications for lifelong gut health in pigs. *Anim Nutr* **3**, 313–321 (2017).
- Thevaranjan, N., et al. Age-associated microbial dysbiosis promotes intestinal permeability, systemic inflammation, and macrophage dysfunction. *Cell Host Microbe* **21**, 455–466 (2017).
- Nikolich-Zugich, J. The twilight of immunity: emerging concepts in aging of the immune system. *Nat. Immunol.* **19**, 10–19 (2018).
- Wang, Q. et al. The aged intestine: performance and rejuvenation. *Aging Dis.* **12**, 1693–1712 (2021).
- Haber, A. L. et al. A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333–339 (2017).
- Nowotschin, S. et al. The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature* **569**, 361–367 (2019).
- Pijuan-Sala, B. et al. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019).
- May-Zhang, A. A. et al. Combinatorial transcriptional profiling of mouse and human enteric neurons identifies shared and disparate subtypes in situ. *Gastroenterology* **160**, 755–770 (2021).
- Morarach, K. et al. Diversification of molecularly defined myenteric neuron classes revealed by single-cell RNA sequencing. *Nat. Neurosci.* **24**, 34–46 (2021).
- James, K. R. et al. Distinct microbial and immune niches of the human colon. *Nat. Immunol.* **21**, 343–353 (2020).
- Elmentaite, R., et al. Single-cell sequencing of developing human gut reveals transcriptional links to childhood Crohn's disease. *Dev. Cell* **55**, 771–783 (2020).
- Elmentaite, R. et al. Cells of the human intestinal tract mapped across space and time. *Nature* **597**, 250–255 (2021).
- Drokhlyansky, E. et al. The human and mouse enteric nervous system at single-cell resolution. *Cell* **182**, 1606–1622 (2020).
- Širvinskas, D. et al. Single-cell atlas of the aging mouse colon. *iScience* **25**, 104202 (2022).
- Hickey, J. W. et al. Organization of the human intestine at single-cell resolution. *Nature* **619**, 572–584 (2023).
- Fawcner-Corbett, D. et al. Spatiotemporal analysis of human intestinal development at single-cell resolution. *Cell* **184**, 810–826 (2021).
- Tabula Muris Consortium A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature* **583**, 590–595 (2020).
- Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
- Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nat. Methods* **11**, 360–361 (2014).
- Eng, C.-H. L. et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235–239 (2019).
- Lee, J. H. et al. Highly multiplexed subcellular RNA sequencing in situ. *Science* **343**, 1360–1363 (2014).
- Goltsev, Y. et al. Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. *Cell* **174**, 968–981 (2018).
- Keren, L. et al. A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging. *Cell* **174**, 1373–1387 (2018).
- Codeluppi, S. et al. Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat. Methods* **15**, 932–935 (2018).
- Moffitt, J. R., et al. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**, eaau5324 (2018).
- Ke, R. et al. In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).
- Ståhl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
- Rodriques, S. G. et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
- Vickovic, S. et al. High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* **16**, 987–990 (2019).

34. Dries, R. et al. Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol.* **22**, 78 (2021).
35. Hu, J. et al. SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nat. Methods* **18**, 1342–1351 (2021).
36. Pham, D., et al. Robust mapping of spatiotemporal trajectories and cell–cell interactions in healthy and diseased tissues. *Nat. Commun.* **14**, 7739 (2023).
37. Zhao, E. et al. Spatial transcriptomics at subspot resolution with BayesSpace. *Nat. Biotechnol.* **39**, 1375–1384 (2021).
38. Xu, H., et al. Unsupervised spatially embedded deep representation of spatial transcriptomics. *Genome Med.* **16**, 12 (2024).
39. Svensson, V., Teichmann, S. A. & Stegle, O. SpatialDE: identification of spatially variable genes. *Nat. Methods* **15**, 343–346 (2018).
40. Kats, I., Vento-Tormo, R. & Stegle, O. SpatialDE2: fast and localized variance component analysis of spatial transcriptomics. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.10.27.466045> (2021).
41. Zeira, R., Land, M., Strzalkowski, A. & Raphael, B. J. Alignment and integration of spatial transcriptomics data. *Nat. Methods* **19**, 567–575 (2022).
42. Liu, X., Zeira, R. & Raphael, B. J. PASTE2: partial alignment of multi-slice spatially resolved transcriptomics data. *Genome Res.* **33**, 1124–1132 (2023).
43. Kiemen, A. L. et al. CODA: quantitative 3D reconstruction of large tissues at cellular resolution. *Nat. Methods* **19**, 1490–1499 (2022).
44. Long, Y. et al. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nat. Commun.* **14**, 1155 (2023).
45. Andersson, A. et al. Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. *Commun Biol* **3**, 565 (2020).
46. Lopez, R., et al. DestVI identifies continuums of cell types in spatial transcriptomics data. *Nat. Biotechnol.* **40**, 1360–1369 (2022).
47. Kleshchevnikov, V., Shmatko, A., Dann, E. & Aivazidis, A. Comprehensive mapping of tissue cell architecture via integrated single cell and spatial transcriptomics. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.11.15.378125> (2020).
48. Elosua-Bayes, M., Nieto, P., Mereu, E., Gut, I. & Heyn, H. SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Res.* **49**, e50 (2021).
49. Biancalani, T. et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat. Methods* **18**, 1352–1362 (2021).
50. Li, B. et al. Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. *Nat. Methods* **19**, 662–670 (2022).
51. Chen, J. et al. A comprehensive comparison on cell-type composition inference for spatial transcriptomics data. *Brief. Bioinform.* **23**, bbac245 (2022).
52. Rood, J. E. et al. Toward a common coordinate framework for the human body. *Cell* **179**, 1455–1467 (2019).
53. Daly, A. C. et al. Tissue and cellular spatiotemporal dynamics in colon aging. ST Dataset. *GEO* <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE284137> (2025).
54. Daly, A. C. et al. Tissue and cellular spatiotemporal dynamics in colon aging. snRNA-seq Dataset. *GEO* <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE285985> (2025).
55. Burclaff, J. et al. A proximal-to-distal survey of healthy adult human small intestine and colon epithelium by single-cell transcriptomics. *Cell Mol. Gastroenterol. Hepatol.* **13**, 1554–1589 (2022).
56. Daly, A. C., Aijo, T. & Smith-Erb, M. cSplotch. Source code. *GitHub* <https://github.com/adaly/cSplotch> (2025).
57. Gillis, N. *Nonnegative Matrix Factorization* (SIAM, 2020).
58. Uhlén, M., et al. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
59. Jasso, G. J. et al. Colon stroma mediates an inflammation-driven fibroblastic response controlling matrix remodeling and healing. *PLoS Biol.* **20**, e3001532 (2022).
60. Zhu, J., Sun, S. & Zhou, X. SPARK-X: non-parametric modeling enables scalable and robust detection of spatial expression patterns for large spatial transcriptomic studies. *Genome Biol.* **22**, 1–25 (2021).
61. Nguyen, T. L. A., Vieira-Silva, S., Liston, A. & Raes, J. How informative is the mouse for human gut microbiota research? *Dis. Model. Mech.* **8**, 1–16 (2015).
62. Okumura, R. et al. Lypd8 promotes the segregation of flagellated microbiota and colonic epithelia. *Nature* **532**, 117–121 (2016).
63. Pelaseyed, T. et al. The mucus and mucins of the goblet cells and enterocytes provide the first defense line of the gastrointestinal tract and interact with the immune system. *Immunol. Rev.* **260**, 8–20 (2014).
64. Barker, N. et al. Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449**, 1003–1007 (2007).
65. Oost, K. C. et al. Specific labeling of stem cell activity in human colorectal organoids using an ASCL2-responsive minigene. *Cell Rep.* **22**, 1600–1614 (2018).
66. Merlos-Suárez, A. et al. The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. *Cell Stem Cell* **8**, 511–524 (2011).
67. Sanders, K. M., Koh, S. D., Ro, S. & Ward, S. M. Regulation of gastrointestinal motility—insights from smooth muscle biology. *Nat. Rev. Gastroenterol. Hepatol.* **9**, 633–645 (2012).
68. Muhl, L. et al. Single-cell analysis uncovers fibroblast heterogeneity and criteria for fibroblast and mural cell identification and discrimination. *Nat. Commun.* **11**, 3953 (2020).
69. Hu, Z. et al. Small proline-rich protein 2A is a gut bactericidal protein deployed during helminth infection. *Science* **374**, eabe6723 (2021).
70. Boland, M. Human digestion—a processing perspective. *J. Sci. Food Agric.* **96**, 2275–2283 (2016).
71. Farin, H. F. et al. Visualization of a short-range Wnt gradient in the intestinal stem-cell niche. *Nature* **530**, 340–343 (2016).
72. Kim, Y. S. & Ho, S. B. Intestinal goblet cells and mucins in health and disease: recent insights and progress. *Curr. Gastroenterol. Rep.* **12**, 319–330 (2010).
73. Gregorieff, A., Grosschedl, R. & Clevers, H. Hindgut defects and transformation of the gastro-intestinal tract in *Tcf4^{-/-}/Tcf1^{-/-}* embryos. *EMBO J.* **23**, 1825–1833 (2004).
74. Sherwood, R. I., Maehr, R., Mazzoni, E. O. & Melton, D. A. Wnt signaling specifies and patterns intestinal endoderm. *Mech. Dev.* **128**, 387–400 (2011).
75. Maimets, M. et al. Mesenchymal–epithelial crosstalk shapes intestinal regionalisation via Wnt and Shh signalling. *Nat. Commun.* **13**, 715 (2022).
76. Bergeron, K.-F. et al. Male-biased aganglionic megacolon in the TashT mouse line due to perturbation of silencer elements in a large gene desert of chromosome 10. *PLoS Genet.* **11**, e1005093 (2015).
77. Moran, C. M. et al. Expression of the fast twitch troponin complex, fTnT, fTnI and fTnC, in vascular smooth muscle. *Cell Motil. Cytoskeleton* **65**, 652–661 (2008).
78. Campagnolo, L. et al. EGFL7 is a chemoattractant for endothelial cells and is up-regulated in angiogenesis and arterial injury. *Am. J. Pathol.* **167**, 275–284 (2005).

79. Byrne, A. B. et al. Pathogenic variants in *MDFIC* cause recessive central conducting lymphatic anomaly with lymphedema. *Sci. Transl. Med.* **14**, eabm4869 (2022).
80. Sphyris, N., Hodder, M. C. & Sansom, O. J. Subversion of niche-signalling pathways in colorectal cancer: what makes and breaks the intestinal stem cell. *Cancers (Basel)* **13**, 1000 (2021).
81. Crosnier, C., Stamatakis, D. & Lewis, J. Organizing cell renewal in the intestine: stem cells, signals and combinatorial control. *Nat. Rev. Genet.* **7**, 349–359 (2006).
82. Morral, C. et al. Zonation of ribosomal DNA transcription defines a stem cell hierarchy in colorectal cancer. *Cell Stem Cell* **26**, 845–861 (2020).
83. Niec, R. E. et al. Lymphatics act as a signaling hub to regulate intestinal stem cell activity. *Cell Stem Cell* **29**, 1067–1082 (2022).
84. Goto, N. et al. Lymphatics and fibroblasts support intestinal stem cells in homeostasis and injury. *Cell Stem Cell* **29**, 1246–1261 (2022).
85. Palikuqi, B. et al. Lymphangiocrine signals are required for proper intestinal repair after cytotoxic injury. *Cell Stem Cell* **29**, 1262–1272 (2022).
86. Zeng, W., et al. Dysregulated hepatic UDP-glucuronosyltransferases and flavonoids glucuronidation in experimental colitis. *Front. Pharmacol.* **13**, 1053610 (2022).
87. Huang, X., Zhou, Y., Sun, Y. & Wang, Q. Intestinal fatty acid binding protein: a rising therapeutic target in lipid metabolism. *Prog. Lipid Res.* **87**, 101178 (2022).
88. Perez-Frances, M. et al. Pancreatic Ppy-expressing γ -cells display mixed phenotypic traits and the adaptive plasticity to engage insulin production. *Nat. Commun.* **12**, 4458 (2021).
89. Fujii, M. et al. Human intestinal organoids maintain self-renewal capacity and cellular diversity in niche-inspired culture condition. *Cell Stem Cell* **23**, 787–793 (2018).
90. Ashekyan, O., et al. Transcriptomic maps of colorectal liver metastasis: machine learning of gene activation patterns and epigenetic trajectories in support of precision medicine. *Cancers (Basel)* **15**, 3835 (2023).
91. Zhu, Q. et al. Sexual dimorphism in lipid metabolism and gut microbiota in mice fed a high-fat diet. *Nutrients* **15**, 2175 (2023).
92. Sakuma, T. et al. Regulation of the expression of two female-predominant CYP3A mRNAs (*CYP3A41* and *CYP3A44*) in mouse liver by sex and growth hormones. *Arch. Biochem. Biophys.* **404**, 234–242 (2002).
93. Vazquez Roque, M. & Bouras, E. P. Epidemiology and management of chronic constipation in elderly patients. *Clin. Interv. Aging* **10**, 919–930 (2015).
94. Tursi, A. et al. Colonic diverticular disease. *Nat. Rev. Dis. Primers* **6**, 20 (2020).
95. Rémond, D. et al. Understanding the gastrointestinal tract of the elderly to develop dietary solutions that prevent malnutrition. *Oncotarget* **6**, 13858–13898 (2015).
96. Dekker, E., Tanis, P. J., Vleugels, J. L. A., Kasi, P. M. & Wallace, M. B. Colorectal cancer. *Lancet* **394**, 1467–1480 (2019).
97. Steegenga, W. T. et al. Structural, functional and molecular analysis of the effects of aging in the small intestine and colon of C57BL/6J mice. *BMC Med. Genomics* **5**, 38 (2012).
98. Biragyn, A. & Ferrucci, L. Gut dysbiosis: a potential link between increased cancer risk in ageing and inflammaging. *Lancet Oncol.* **19**, e295–e304 (2018).
99. Jurk, D. et al. Chronic inflammation induces telomere dysfunction and accelerates ageing in mice. *Nat. Commun.* **2**, 4172 (2014).
100. Egge, N. et al. Age-onset phosphorylation of a minor actin variant promotes intestinal barrier dysfunction. *Dev. Cell* **51**, 587–601 (2019).
101. Pentimikko, N. et al. Notum produced by Paneth cells attenuates regeneration of aged intestinal epithelium. *Nature* **571**, 398–402 (2019).
102. Omrani, O. et al. IFN γ -Stat1 axis drives aging-associated loss of intestinal tissue homeostasis and regeneration. *Nat. Commun.* **14**, 6109 (2023).
103. Sun, T. et al. Aging-dependent decrease in the numbers of enteric neurons, interstitial cells of Cajal and expression of connexin43 in various regions of gastrointestinal tract. *Ageing* **10**, 3851–3865 (2018).
104. Sovran, B. et al. Age-associated impairment of the mucus barrier function is associated with profound changes in microbiota and immunity. *Sci. Rep.* **9**, 1437 (2019).
105. Moog, F. The differentiation and redifferentiation of the intestinal epithelium and its brush border membrane. *Ciba Found. Symp.* **70**, 31–50 (1979).
106. Henning, S. J. Postnatal development: coordination of feeding, digestion, and metabolism. *Am. J. Physiol.* **241**, G199–G214 (1981).
107. Ménard, D., Dagenais, P. & Calvert, R. Morphological changes and cellular proliferation in mouse colon during fetal and postnatal development. *Anat. Rec.* **238**, 349–359 (1994).
108. Jerby-Arnon, L. & Regev, A. DIALOGUE maps multicellular programs in tissue from single-cell or spatial transcriptomics data. *Nat. Biotechnol.* **40**, 1467–1477 (2022).
109. Kuwahara, M. et al. The transcription factor Sox4 is a downstream target of signaling by the cytokine TGF- β and suppresses T β 2 differentiation. *Nat. Immunol.* **13**, 778–786 (2012).
110. van de Wetering, M. et al. The β -catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. *Cell* **111**, 241–250 (2002).
111. Shah, Y. M., et al. Hypoxia-inducible factor augments experimental colitis through an MIF-dependent inflammatory signaling cascade. *Gastroenterology* **134**, 2036–2048 (2008).
112. Kobayashi, H. et al. The balance of stromal BMP signaling mediated by GREM1 and ISLR drives colorectal carcinogenesis. *Gastroenterology* **160**, 1224–1239 (2021).
113. Kraiczky, J. et al. Graded BMP signaling within intestinal crypt architecture directs self-organization of the Wnt-secreting stem cell niche. *Cell Stem Cell* **30**, 433–449 (2023).
114. Yaeger, R. et al. Genomic alterations observed in colitis-associated cancers are distinct from those found in sporadic colorectal cancers and vary by type of inflammatory bowel disease. *Gastroenterology* **151**, 278–287 (2016).
115. Chatila, W. K. et al. Integrated clinical and genomic analysis identifies driver events and molecular evolution of colitis-associated cancers. *Nat. Commun.* **14**, 110 (2023).
116. Jones, R. G. et al. Conditional deletion of beta1 integrins in the intestinal epithelium causes a loss of Hedgehog expression, intestinal hyperplasia, and early postnatal lethality. *J. Cell Biol.* **175**, 505–514 (2006).
117. Wiener, Z. et al. Prox1 promotes expansion of the colorectal cancer stem cell population to fuel tumor growth and ischemia resistance. *Cell Rep.* **8**, 1943–1956 (2014).
118. Marincola Smith, P. et al. Colon epithelial cell TGF β signaling modulates the expression of tight junction proteins and barrier function in mice. *Am. J. Physiol. Gastrointest. Liver Physiol.* **320**, G936–G957 (2021).
119. Akitake-Kawano, R. et al. Inhibitory role of Gas6 in intestinal tumorigenesis. *Carcinogenesis* **34**, 1567–1574 (2013).
120. Nguyen, J. Q. & Irby, R. B. TRIM21 is a novel regulator of Par-4 in colon and pancreatic cancer cells. *Cancer Biol. Ther.* **18**, 16–25 (2017).
121. Tian, Z.-Q., Shi, J.-W., Wang, X.-R., Li, Z. & Wang, G.-Y. New cancer suppressor gene for colorectal adenocarcinoma: filamin A. *World J. Gastroenterol.* **21**, 2199–2205 (2015).
122. Rasool, R. U. et al. A journey beyond apoptosis: new enigma of controlling metastasis by pro-apoptotic Par-4. *Clin. Exp. Metastasis* **33**, 757–764 (2016).

123. Yui, S. et al. YAP/TAZ-dependent reprogramming of colonic epithelium links ECM remodeling to tissue regeneration. *Cell Stem Cell* **22**, 35–49 (2018).
124. Nusse, Y. M. et al. Parasitic helminths induce fetal-like reversion in the intestinal stem cell niche. *Nature* **559**, 109–113 (2018).
125. Nair, S., Bist, P., Dikshit, N. & Krishnan, M. N. Global functional profiling of human ubiquitome identifies E3 ubiquitin ligase DCST1 as a novel negative regulator of type-I interferon signaling. *Sci. Rep.* **6**, 36179 (2016).
126. Feng, M., et al. Inducible guanylate-binding protein 7 facilitates influenza A virus replication by suppressing innate immunity via NF- κ B and JAK-STAT signaling pathways. *J. Virol.* **95**, e02038-20 (2021).
127. Wang, X.-Q. et al. Epithelial but not stromal expression of collagen alpha-1(III) is a diagnostic and prognostic indicator of colorectal carcinoma. *Oncotarget* **7**, 8823–8838 (2016).
128. Li, J. et al. Elastin is a key factor of tumor development in colorectal cancer. *BMC Cancer* **20**, 217 (2020).
129. Ghosh, S. Sialic acid and biology of life: an introduction. In *Sialic Acids and Sialoglycoconjugates in the Biology of Life, Health and Disease* (ed. Ghosh, S.) (Academic Press, 2020).
130. Sanders, R. D., Sefton, J. M. I., Moberg, K. H. & Fridovich-Keil, J. L. UDP-galactose 4' epimerase (GALE) is essential for development of *Drosophila melanogaster*. *Dis. Model. Mech.* **3**, 628–638 (2010).
131. Ardini, E. et al. The TPM3-NTRK1 rearrangement is a recurring event in colorectal carcinoma and is associated with tumor sensitivity to TRKA kinase inhibition. *Mol. Oncol.* **8**, 1495–1507 (2014).
132. Grosse, J. et al. Insulin-like peptide 5 is an orexigenic gastrointestinal hormone. *Proc. Natl Acad. Sci. USA* **111**, 11133–11138 (2014).
133. Lawal, H. O. & Krantz, D. E. SLC18: Vesicular neurotransmitter transporters for monoamines and acetylcholine. *Mol. Aspects Med.* **34**, 360–372 (2013).
134. Zhang, D. et al. Deletions at *SLC18A1* increased the risk of CRC and lower *SLC18A1* expression associated with poor CRC outcome. *Carcinogenesis* **38**, 1057–1062 (2017).
135. Lee, S.-K. & Ahnn, J. Regulator of calcineurin (RCAN): beyond Down syndrome critical region. *Mol. Cells* **43**, 671–685 (2020).
136. Niitsu, H. et al. *KRAS* mutation leads to decreased expression of regulator of calcineurin 2, resulting in tumor proliferation in colorectal cancer. *Oncogenesis* **5**, e253 (2016).
137. Gao, N., White, P. & Kaestner, K. H. Establishment of intestinal identity and epithelial–mesenchymal signaling by *Cdx2*. *Dev. Cell* **16**, 588–599 (2009).
138. Dalerba, P. et al. *CDX2* as a prognostic biomarker in stage II and stage III colon cancer. *N. Engl. J. Med.* **374**, 211–222 (2016).
139. Farr, L. et al. *CD74* signaling links inflammation to intestinal epithelial cell regeneration and promotes mucosal healing. *Cell Mol. Gastroenterol. Hepatol.* **10**, 101–112 (2020).
140. Wang, W. et al. *RAI16* maintains intestinal homeostasis and inhibits NLRP3-dependent IL-18/CXCL16-induced colitis and the progression of colitis-associated colorectal cancer. *Clin. Transl. Med.* **12**, e993 (2022).
141. Paillas, S. et al. *MAPK14/p38 α* confers irinotecan resistance to TP53-defective cells by inducing survival autophagy. *Autophagy* **8**, 1098–1112 (2012).
142. Natarajan, G. K., Mishra, J., Camara, A. K. S. & Kwok, W.-M. *LETM1*: a single entity with diverse impact on mitochondrial metabolism and cellular signaling. *Front. Physiol.* **12**, 637852 (2021).
143. Morris, H. T. et al. Loss of N-WASP drives early progression in an *Apc* model of intestinal tumorigenesis. *J. Pathol.* **245**, 337–348 (2018).
144. Liu, X., Xia, S., Zhang, Z., Wu, H. & Lieberman, J. Channelling inflammation: gasdermins in physiology and disease. *Nat. Rev. Drug Discov.* **20**, 384–405 (2021).
145. Menzel, K. et al. Cathepsins B, L and D in inflammatory bowel disease macrophages and potential therapeutic effects of cathepsin inhibition in vivo. *Clin. Exp. Immunol.* **146**, 169–180 (2006).
146. Huang, W.-C. et al. Sphingosine-1-phosphate phosphatase 2 promotes disruption of mucosal integrity, and contributes to ulcerative colitis in mice and humans. *FASEB J.* **30**, 2945–2958 (2016).
147. Shen, T. et al. Erbin exerts a protective effect against inflammatory bowel disease by suppressing autophagic cell death. *Oncotarget* **9**, 12035–12049 (2018).
148. Rospo, G. et al. Evolving neoantigen profiles in colorectal cancers with DNA repair defects. *Genome Med.* **11**, 42 (2019).
149. Hayashi, H. et al. Overexpression of *IQGAP1* in advanced colorectal cancer correlates with poor prognosis—critical role in tumor invasion. *Int. J. Cancer* **126**, 2563–2574 (2010).
150. Cai, L., Makhov, A. M., Schafer, D. A. & Bear, J. E. Coronin 1B antagonizes cortactin and remodels Arp2/3-containing actin branches in lamellipodia. *Cell* **134**, 828–842 (2008).
151. Goldman, R. D. et al. Accumulation of mutant lamin A causes progressive changes in nuclear architecture in Hutchinson–Gilford progeria syndrome. *Proc. Natl Acad. Sci. USA* **101**, 8963–8968 (2004).
152. Ho, C. Y., Jaalouk, D. E., Vartiainen, M. K. & Lammerding, J. Lamin A/C and emerin regulate MKL1–SRF activity by modulating actin dynamics. *Nature* **497**, 507–511 (2013).
153. Wang, A. S., Kozlov, S. V., Stewart, C. L. & Horn, H. F. Tissue specific loss of A-type lamins in the gastrointestinal epithelium can enhance polyp size. *Differentiation* **89**, 11–21 (2015).
154. Ogino, S. et al. p21 expression in colon cancer and modifying effects of patient age and body mass index on prognosis. *Cancer Epidemiol. Biomarkers Prev.* **18**, 2513–2521 (2009).
155. Daly, A. C., Geras, K. J. & Bonneau, R. A convolutional neural network for common coordinate registration of high-resolution histology images. *Bioinformatics* **37**, 4216–4226 (2021).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

Methods

Image and ST data preprocessing

Hematoxylin and eosin (H&E) images were processed with SpoTteR (version 0.0.1)¹⁵⁶. Briefly, original H&E images were scaled to approximately 500 × 500 pixels. Then, the tissue section was masked generously from the image through 10% quantile thresholding in a user-defined color channel. To detect probable spot centers, the image Hessian was computed. The spot centers then acted as potential grid points that were likely part of a regular grid structure and were selected by calculating the *x* and *y* distances between all detected centers. A regular grid was then fitted to these potential grid points using a custom optimizer based on the 'nlminb' function of the R package stats, which minimizes the distance of potential grid points to the suggested regular grid while assuming angles of 90° and 42 starting grid points per row and column. Through an iterative process, in which the 0.1% potential grid points that least fit the grid were removed in each iteration, the number of grid points per row and column was updated and a new grid was fitted until the target number of grid points per row (here, 35) and column (here, 33) was reached. Finally, those grid points that overlapped the tissue sections were identified by building a mask that represented the tissue area and registering all grid points that were present in this mask. In case a sectioning artifact was present, the corresponding ST spot was removed from all subsequent analyses.

ST spot annotation

To assign each ST spot with a single (mutually exclusive) corresponding histological tag, a cloud-based interface³³ was used to assign each spot (*x*, *y*) with one or more regional tags. A total of 14 tags (MROIs) based on established major gross morphology were used: crypt apex, subcrypt, crypt mid, crypt base, CM, epithelium and muscularis mucosae (EMM), EMM and submucosa (EMMSUB), epithelium and muscle and submucosa, ME, MEI, MI, muscularis mucosae and interna, muscle and submucosa, and PP.

Detection of tissue sections in histology images

In most cases, more than one tissue section was placed on the active area of one ST array. To distinguish between different tissue sections, a two-dimensional integer lattice was assumed so that labeled ST spots that were connected were assigned the same tissue section. Next, ST spots were filtered on the basis of their sequencing data quality, such that tissue sections labeled with fewer than five (ages 0 days–3 weeks) or ten (ages 4 weeks–2 years) spots in total were discarded from further analysis. ST spots with fewer than 800 unique molecular identifiers (UMIs) were also discarded from further analysis. To account for spots without four neighbors, each spot was mapped after filtering to match the same two-dimensional integer lattice $[[0,1,0], [1,1,1], [0,1,0]]$ and spots not matching this patterns were also discarded.

Training a cell density classifier for segmentation

To train a cell density classifier to segment individual objects in the histology images, each whole-slide image (WSI) was first subset into smaller patches while retaining patches at the same resolution for training, selected such that each patch would contain all the major colon layers from at least one tissue cross section from the original WSI. Overall, ~200 patches were selected for training, with at least ten replicate patches from each of the different ages. To count the number of cell segments present in each ST spot, a density classifier was first trained using ilastik (version 1.4.1)¹⁵⁷. This workflow estimated the density of blob-like structures usually present as overlapping instances, decreasing the chance of underestimating the number of objects because of undersegmentation, which we reasoned was the most appropriate approach for counting cell-dense areas in the colon. To ensure reproducibility across all density conditions in the dataset, in each training patch, at least three separate tissue areas (that is, training

squares) were used. In each training square, two classes of objects were labeled: cells and background.

Processing segments per ST spot

Each WSI processed with the SpoTteR ST processing tool¹⁵⁶ was split into image patches of 200 × 220 pixels representing the size of an ST spot capture area. The cell-counting workflow described above was then used to extract density predictions for each ST spot. The following image processing and segmentation steps were performed with skimage¹⁵⁸ (version 0.18.1). First, an ellipse shape (radius = 100 pixels) was used to mask the true ST capture area in each patch. If no cells were left in the patch (mean image intensity < 0.05), the patch was discarded from further processing. Next, the multi-Otsu¹⁵⁹ thresholding algorithm (cutoff > 50) was used to separate objects detected in the patch. Local maxima were found for each object and used to estimate distances between the same. These were then approximated by the watershed algorithm¹⁶⁰ into segments that were further labeled into individual objects used in all downstream analyses.

Training an object classifier to obtain superclass cell type labels from histology images

MROI-specific object classifiers were trained using ilastik (version 1.4.1)¹⁵⁷, with binary segmentation images and their corresponding H&E patches as input. In this way, each segment in the H&E patch was assigned a cell type superclass label. In each classifier, depending on the cells present in the corresponding MROI, up to five superclasses were labeled: colonocyte, immune, interstitial, muscle and epithelial. MROI-specific classifiers (colonocyte, PP, muscle, noncolonocyte and epithelium) with corresponding cell type superclass labels (colonocyte, immune, interstitial, muscle and epithelial), MROI labels and snRNA-seq cell type labels are presented in Supplementary Table 3. Because of differences in cell and tissue morphology, separate superclass classifiers were created for juvenile (0–4 weeks) and adult/aged (6 weeks–2 years) groups, for a total of ten image-based classifiers representing five MROI-specific classifiers across two age groups. To train each classifier, ~150 patches were randomly selected from all three regions of the colon (proximal, middle and distal) and from each of the following MROIs: CM, EMMSUB, subcrypt, crypt mid, crypt apex, MEI and PP. The object classifiers took into consideration object-level characteristics, such as object shape and work, to predict similar objects in the nearby space. Algorithm features used in training included two-dimensional convex hull and skeleton descriptors in a neighborhood size of 30 × 30 pixels for each object and used a simple threshold (0.5) with a small smoothing factor (1.0). Properties attributed to standard object features such as shape, size, channel intensity and location were also selected in the training process. In total, 1,540 patches and 83,721 segments were labeled during training.

Processing cell type superclass from histology images per ST spot

Each H&E image patch (200 × 220 pixels) and corresponding segmentation predictions were used in ilastik batch processing to predict cell type superclasses using the object classifier described above. Cell label predictions were used in the following image processing workflow implemented using skimage (version 0.18.1)¹⁵⁸. Each pixel class in the image was assigned one of the cell type superclasses. Then, small objects (<50 pixels) were removed from each patch and the remaining small segments abutting each other were merged if belonging to the same cell type class. The fractions of foreground pixels belonging to each object class were used as estimates of the abundance of each cell type in each patch.

Testing the object classifier used for obtaining superclass cell type labels from histology images

To evaluate the performance of the cell superclass classifiers, a test set of 781 patches (50% size of training set) spanning the five adult/

aged (colonocyte, immune, interstitial, muscle and epithelial) and five juvenile (colonocyte, immune, interstitial, muscle and epithelial) cell superclasses was set aside. Foreground objects were detected using the binary segmentation workflow, after which all objects were manually assigned to one of the five superclasses (colonocyte, immune, interstitial, muscle and epithelial). The same images were then input into the respective MROI-specific object classifiers and performance was evaluated by calculating the confusion (fraction of pixels labeled class A but predicted class B) for all pairs of cell types.

Morphology-informed deconvolution using SPOTlight

The SPOTlight model was used for ‘bottom-up’ deconvolution of ST data⁴⁸, which takes as input two matrices of count data: V , a (genes \times cells) matrix containing the snRNA-seq count data (in which each cell is assigned to one of k_{sn} types), and V' , a (genes \times spots) matrix containing the ST count data. Expression matrices were preprocessed in the following manner: (1) genes were subset to the set shared across both modalities; (2) data were depth-normalized to 10,000 UMI counts per cell or spot; and (3) data were scaled gene-wise to unit variance. Next, genes were further subset to cellular marker genes ($\log_2(\text{fold change}) > 1$, Benjamini–Hochberg false discovery rate (FDR) < 0.05 ; likelihood ratio test) and balanced across the k_{sn} cell types, selecting the top $m = 23$ genes by FDR for each cell type, where m was chosen as the minimum number of significant marker genes across all cell types (23 for T cells). This resulted in a total of 334 unique genes used in deconvolution (Supplementary Table 4).

V was factored into component matrices W (genes \times topics) and H (topics, cells) by NMF:

$$V = W \times H \tag{1}$$

where the number of topics is assumed to be equal to the number of snRNA-seq cell types k_{sn} . Before NMF, all values of $W_{g,t}$ were initialized to the probability of gene g being a marker gene for cell type t (quantified as Benjamini–Hochberg-corrected P_{adj} of t -test on $\log(\text{count})$ data) and H was initialized as a binary matrix denoting the class assignment for each cell in the dataset. These initialization conditions—in which topics were treated equivalently to cell types—are meant to bias the optimization toward the discovery of biologically meaningful topic profiles.

Next, topic profiles W were fixed, and the following equation was minimized over H' (topics, spots) by non-negative least squares (NNLS):

$$V' = W \times H' \tag{2}$$

In this manner, the expression profile of each spot in the ST data was mapped to a combination of topics inferred from snRNA-seq.

Third, Q , a (topics, k_{sn}) matrix from H was derived by selecting all cells from the same cell type and computing the median of each topic for a consensus cell-type-specific topic signature. This topic matrix was used in a final NNLS minimization to find P , the (k_{sn} , spots) matrix denoting the inferred cellular composition of each ST spot:

$$H' = Q \times P \tag{3}$$

A modification to Eq. 3 was implemented that allows the incorporation of morphology-informed composition information derived from the image segmentation workflow, by providing two additional inputs: L , a (k_{morph} , spots) matrix containing the composition of each ST spot in terms of the k_{morph} morphological cell types defined in the segmentation model, and S , a (k_{morph} , k_{sn}) binary matrix mapping each expression cell type to a morphological cell type. Any proposed compositional matrix P should additionally satisfy the following to reconstruct morphology-informed compositional data:

$$L = S \times P \tag{4}$$

Morphology-aware SPOTlight decomposition was then achieved by solving the following optimization problem:

$$\min_{P \geq 0} (H' - Q \times P) + \alpha(L - S \times P) \tag{5}$$

where α controls the relative importance of the morphological composition loss (second term) and the expression loss (first term). This optimization problem was solved using the PyTorch implementation of the Adam optimizer with a learning rate of 0.01 run for 100,000 iterations from a random initialization.

cSpotch model specification

Genes i , tissue sections j and spots k were indexed as follows: $i \in [1, \dots, N_{genes}]$, $j \in [1, \dots, N_{tissues}]$, $k \in [1, \dots, N^{(j)}_{spots}]$. Each tissue j was registered to a common coordinate system, such that each spot k was assigned to one of N_{MROI} distinct MROIs, denoted $D_k^{(j)} \in [1, \dots, N_{MROI}]$, as described above. In the compositional mode, each spot was additionally assigned a simplex vector $F_k^{(j)} \in \mathbb{R}_{\geq 0}^{N_{cell\ types}}$, $\sum_x F_k^{(j)} = 1 \forall j, k$, which describes its proportional composition in terms of all $N_{cell\ types}$ unique cell types across the dataset.

For each gene and each spot, the observed counts $y_{i,j,k}$ were considered to be realizations of random variable with an expected value equal to $s_{j,k} \lambda_{i,j,k}$, where $s_{j,k}$ is a size factor (total number of UMIs observed at spot k) and $\lambda_{i,j,k}$ is the rate of expression of gene i (events per exposure), such that gene expression is modeled independently of sequencing depth. In practice, $s_{j,k}$ was further normalized by the median depth across all spots in the dataset to facilitate comparisons of results across analyses. Thus, cSpotch offers the user two choices for modeling count data: the Poisson distribution or the negative binomial (NB) distribution. Either may be supplemented with zero inflation to account for dropout events (technical zeros), yielding the zero-inflated Poisson (ZIP) or zero-inflated NB (ZINB) distributions:

$$y_{i,j,k} \sim \begin{cases} \text{Pois}(s_{j,k} \lambda_{i,j,k}) & \text{nb} = 0, \text{zi} = 0 \\ \text{ZIP}(s_{j,k} \lambda_{i,j,k}, \theta_i^p) & \text{nb} = 0, \text{zi} = 1 \\ \text{NB}(s_{j,k} \lambda_{i,j,k}, \phi_i) & \text{nb} = 1, \text{zi} = 0 \\ \text{ZINB}(s_{j,k} \lambda_{i,j,k}, \phi_i, \theta_i^p) & \text{nb} = 1, \text{zi} = 1 \end{cases} \tag{6}$$

where $s_{j,k} \lambda_{i,j,k}$ represents the expected mean of all distributions, ϕ_i represents the gene-specific overdispersion parameter for the NB family and θ_i^p represents the gene-specific probability of technical zeros or dropout. The zero-inflated model accounts for an overabundance of zeros by introducing a second zero-generating process gated by a Bernoulli random variable:

$$y_{i,j,k} \sim \begin{cases} 0 & \text{if } \theta_i = 0 \\ \text{Pois}(s_{j,k} \lambda_{i,j,k}) & \text{if } \theta_i = 1 \end{cases} \tag{7}$$

$$\theta_i \sim \text{Bernoulli}(\theta_i^p), \theta_i^p \sim \beta(1, 2)$$

where the Poisson process can be replaced by NB without loss of generality. This mixture model allows for ‘true’ biological zeros to be generated by the Poisson or NB process describing the expression model while ‘shunting’ technical zeros into a separate, technical process, preventing abundant dropout events from lowering the estimated mean expression $\lambda_{i,j,k}$. Because the Poisson process does not allow for overdispersion (variance exceeding the mean), ZIP should be preferred to Poisson in most situations, whereas the use of NB or ZINB may depend on data quality.

While cSpotch considered a separate random variable to describe gene expression in each spot, the rate parameters $\lambda_{i,j,k}$ were described in terms of a GLM that separates variation into shared and individual components. Specifically, the rate of gene expression was informed by three components:

$$\log(\lambda_{i,j,k}) = B_{i,j,k} + \psi_{i,j,k} + \epsilon_{i,j,k}, \tag{8}$$

where $B_{i,j,k}$ describes the characteristic expression of gene i within the tissue context of spot k , $\psi_{i,j,k}$ describes the neighborhood effects and $\epsilon_{i,j,k}$ describes spot-specific effects. $B_{i,j,k}$ is calculated as a weighted sum

of cellular expression rates β_i in proportions $E_k^{(j)}$. Cellular expression was allowed to vary both across MROIs and sample conditions. As such, a characteristic expression matrix $\beta_i \in \mathbb{R}^{N_{\text{MROI}} \times N_{\text{celltypes}}}$ was defined (when compositional data are unavailable, each MROI may be treated as composed of a single ‘average’ cell type and $\beta_i \in \mathbb{R}^{N_{\text{MROI}}}$ is defined instead). Inferring a posterior over β_i allowed the quantification of expression changes across regions or cell types by comparing relevant entries.

Because characteristic expression is expected to vary across conditions (for example, age, colon region and sex), region-specific expression β_i was modeled in a hierarchical fashion defined by sample covariates. Up to three levels were explicitly modeled in the hierarchy, each of which split the sample to distinct groups along some covariate. At the top level, the dataset was split along an important covariate (for example, age) and a separate $\beta_i^{(l_1)}$ was modeled for each unique set $l_1 \in \{1, \dots, L_1\}$. At the next level, each set was further partitioned along another covariate (for example, colon region). $\beta_i^{(l_1, l_2)}$ was assumed to be centered around its corresponding top-level estimate $\beta_i^{(l_1)}$, with some additional variance associated with the new covariate $\sigma_i^{(l_2)}$. This encoded knowledge about the experimental system and separated out sources of variation associated with each covariate (without propagation of uncertainty, as all levels are inferred simultaneously). A three-level hierarchical model for β_i was, thus, specified as:

$$\begin{aligned} \beta_{i, l_1} &\sim \mathcal{N}\left(\mu_i^{(l_1)}, \left(\sigma_i^{(l_1)}\right)^2 \mathbf{I}\right) \\ \beta_{i, l_1, l_2} &\sim \mathcal{N}\left(\beta_{i, l_1}, \left(\sigma_i^{(l_2)}\right)^2 \mathbf{I}\right) \\ \beta_{i, l_1, l_2, l_3} &\sim \mathcal{N}\left(\beta_{i, l_1, l_2}, \left(\sigma_i^{(l_3)}\right)^2 \mathbf{I}\right) \\ \sigma_i^{(l_2)}, \sigma_i^{(l_3)} &\sim \mathcal{N}_{\geq 0}(0, 1) \end{aligned} \tag{9}$$

where, in the compositional mode, prior hyperparameters $\mu_i^{(l_1)}(\cdot, m)$ and $\sigma_i^{(l_1)}(\cdot, m)$ are set to the empirical mean and s.d., respectively, over the expression of gene i in cell type m in the snRNA-seq data for all MROIs; in the noncompositional mode (one ‘average’ cell type per MROI), $\mu_i^{(l_1)}(\cdot)$ and $\sigma_i^{(l_1)}(\cdot)$ are set to 0 and 2, respectively, for all MROIs. Variation parameters $\sigma_i^{(l_2)}, \sigma_i^{(l_3)}$ are assumed to have truncated Gaussian priors reflecting our limited knowledge of the effects of covariate-driven variation and are inferred separately for each level 2 and 3 covariate group. For convenience, because each tissue j belongs to one covariate group at each level, the inverse mapping function $\rho^{-1}(j)$ was introduced to map j to the appropriate l_1, l_2, l_3 indices for $\beta_i, B_{i,j,k}$ was formally defined in the noncompositional model:

$$B_{i,j,k} = \begin{cases} x_{j,k}^T \beta_{i, \rho^{-1}(j)} E_k^{(j)} & \text{if compositional} \\ x_{j,k}^T \beta_{i, \rho^{-1}(j)} & \text{else} \end{cases} \tag{10}$$

where $x_{j,k}^T$ is a one-hot encoding of $D_k^{(j)}$, the MROI annotation for spot k . With this framework for integrating multiple sections or experiments, the posterior distributions of the latent parameters β_i were studied at different levels of the hierarchical experimental design and expression changes were quantified across conditions, tissue contexts or individual cell types.

The second component of Eq. 8, $\psi_{i,j,k}$, describes the effects of the local neighborhood of spot k on the expression of gene i . This was modeled using the conditional autoregressive (CAR) prior, which assumes that the value at a given location (spot) is conditional on the values of neighboring locations (spots). $\psi_{i,j}$ was defined as a Markov random field over the spots on each array:

$$\begin{aligned} \psi_{i,k} | a_i, \tau_i, \mathbf{W}_j &\sim \mathcal{N}\left(\mathbf{0}, \left(\tau_i \mathbf{K}_j \left(\mathbf{I} - a_i \mathbf{K}_j^{-1} \mathbf{W}_j\right)\right)^{-1}\right) \\ a_i &\sim \mathcal{U}(0, 1) \\ \tau_i &\sim \Gamma^{-1}(1, 1) \end{aligned} \tag{11}$$

where a_i is a spatial autocorrelation parameter, τ_i is a conditional precision parameter, \mathbf{K}_j is a diagonal matrix containing the number of neighbors for each spot in tissue j and \mathbf{W}_j is the adjacency matrix (with zero diagonal). If the classic ST methodology of using cartesian arrays is applied, each spot is assumed to have a four-spot neighborhood; if the Visium platform using hexagonal arrays is applied, each spot is assumed to have a six-spot neighborhood. The level of spatial autocorrelation (a_i) and conditional precision (τ_i) was inferred separately for each gene, although weak hyperpriors on both parameters are shared across all genes to enforce the same physiological constraints (Eq. 11). Taken together, the B and ψ terms capture spatial autocorrelation on two different scales: tissue context (across samples) and local neighborhood (within samples).

The final component of Eq. 8, $\epsilon_{i,j,k}$, captures variation at the level of individual spots. This variation was assumed to be independent and identically distributed for each gene:

$$\begin{aligned} \epsilon_{i,j,k} &\sim \mathcal{N}(0, \sigma_i^2) \\ \sigma_i &\sim \mathcal{N}_{\geq 0}(0, 0.3^2) \end{aligned} \tag{12}$$

where σ_i is the inferred level of variability for gene i . The degree of variance in the truncated Gaussian prior on σ_i is set to 0.3 following the work of Maniatis et al.¹⁶¹, who chose the value to reflect the relatively low levels of spot-specific variation (relative to other model terms).

BF estimation for cSpotch DE analysis

To quantify difference in expression between two conditions using cSpotch, the BF between posterior distributions over characteristic expression coefficients β estimated by the model was examined. Without loss of generality, the difference in expression was quantified between conditions represented by $\beta^{(1)}$ and $\beta^{(2)}$, which may differ across any combination of genes, sample covariates (for example, distal versus proximal colon), tissue regions (for example, crypt apex versus muscle) or cell types (for example, neuron versus myocyte). A random variable $\Delta_\beta = \beta^{(1)} - \beta^{(2)}$ was defined, which captures the difference between $\beta^{(1)}$ and $\beta^{(2)}$. If Δ_β is tightly centered around zero, then the two distributions are very similar to each other and the null hypothesis of identical expression cannot be rejected. To quantify this similarity, the posterior distribution $\Delta_\beta | \mathcal{D}$ (where \mathcal{D} represents the data used to train the model) was compared to the prior distribution Δ_β using the Savage–Dickey density ratio that estimates the BF between the conditions:

$$\text{BF} \approx \frac{p(\Delta_\beta = 0)}{p(\Delta_\beta = 0 | \mathcal{D})} \tag{13}$$

where the probability density functions were evaluated at zero. If expression is different between the two conditions, then the posterior $\Delta_\beta | \mathcal{D}$ will have very little mass at 0 and the estimated BF will be large (by convention, $\text{BF} > 5$ indicates substantial support). Conversely, for similar expression regimes, the posterior will place a mass equal to or greater than that of the prior at zero and the BF will be ≤ 1 . While $p(\Delta_\beta)$ can be derived analytically (the prior distributions over all β are normally distributed and the difference between two normally distributed random variables is in turn normally distributed), $p(\Delta_\beta | \mathcal{D})$ must be approximated using the posterior samples obtained below. When we executed a comparison between sets of conditions (for example, neurons versus all other cells), we pooled the posterior samples from all component conditions together.

Parameter inference for cSpotch

cSpotch was implemented in Stan¹⁶². For all analyses, Bayesian inference was performed over the parameters using Stan’s adaptive Hamiltonian Monte Carlo (HMC) sampler with default parameters. Four independent chains were sampled, each with 250 warmup iterations and 250 sampling iterations, and convergence was monitored using

the \hat{R} statistic. On average, sampling of a given gene was completed in ~3 h parallelized over four CPUs (<2 GB of total RAM used) and the sampling time for each chain was linear in both the number of spots in the dataset and the number of samples drawn, while multiple genes can be run in parallel. For high-resolution technologies (for example, Visium HD), the number of spots per tissue is dramatically increased and users may consider using variational inference to initialize HMC sampling (provided as ‘spotch-vinit’ executable) to reduce sampling time and/or suppressing the CAR model component (accounting for ~60% of computational time) with the ‘--no-car’ option.

Simulated ST data generation

Simulated ST data were generated from the snRNA-seq profiles in the two regimes described in the subsequent sections. For all simulation studies, 12 ST arrays were generated, each containing 2,000 spots. For data in which distinct MROIs were simulated, the two regions were considered to exist in a 1:1 ratio. Cell clusters comprising each region are detailed in Supplementary Table 5.

Cluster-based simulation. Average expression profiles for unique mouse colon cell types obtained from snRNA-seq $G \in \mathbb{Z}_{\geq 0}^{N_{\text{genes}} \times N_{\text{celltypes}}}$ ($N_{\text{genes}} = 22,986, N_{\text{celltypes}} = 30$) were normalized column-wise, such that the total expression within each cluster summed to 1. For each spot k from tissue j , a true composition vector $E_k^{(j)} \in \mathbb{R}_{\geq 0}^{N_{\text{celltypes}}}$ was drawn such that a random subset of cell types present in the current region ($D_k^{(j)}$; Supplementary Table 5) were represented in uniformly random proportions. For each cell type t , reads $y_{k,t}$ were drawn from a multinomial distribution:

$$y_{k,t} \sim \text{Mult}\left(M_k^{(j)} E_{k,t}^{(j)}, G_{.,t}\right) \tag{14}$$

where $M_k^{(j)}$ is the total number of reads in the current spot and $G_{.,t}$ is the expression profile for cell type t . In practice, $M_k^{(j)} = 1,000$ was used for all j, k . Read counts were then pooled across all cell types, yielding spot-level reads $y_k = \sum_t y_{k,t}$. Composition vectors were then (optionally) pooled within high-level annotation categories (such as the histological superclasses in Supplementary Table 3) to simulate cases of limited observability. Finally, Gaussian noise was (optionally) added to $E_k^{(j)}$ to generate ‘observed’ composition vectors $\tilde{E}_k^{(j)}$. Resulting negative entries were removed by element-wise maximization against the zero vector ($\tilde{E}_k^{(j)} \leftarrow \max(\tilde{E}_k^{(j)}, 0)$), following which $\tilde{E}_k^{(j)}$ was renormalized to produce a valid simplex. y, D, \tilde{E} served as inputs to cSpotch. Randomized compositions (highest level of noise) were generated by drawing from an $N_{\text{celltypes}}$ -dimensional uniform (0, 1) distribution and then unit-normalizing sums to produce valid simplexes.

For all cluster-based data, cSpotch was run with a Poisson likelihood (without zero inflation) to meet the expected form of the marginals over the generating distribution (multinomial). To focus on the compositional module of cSpotch, the effects of local spatial autocorrelation were removed by simulating each spot independently and, as such, suppressed the ψ_i term of the GLM. Local neighborhood effects can be simulated in this regimen by passing a spatial smoothing filter over the simulated array, blending the transcriptome of each spot with those of its neighbors in a defined proportion.

Cell-based simulation. Individual snRNA-seq cell profiles $G \in \mathbb{Z}_{\geq 0}^{N_{\text{genes}} \times N_{\text{celltypes}}}$ ($N_{\text{genes}} = 22,986, N_{\text{cells}} = 419,334$), where each cell was assigned to one of $N_{\text{celltypes}} = 30$ cell types, were normalized cell-wise to a target depth of 1,000 counts using Scanpy’s (version 1.11.1)¹⁶³ ‘normalize_total’ function. Each spot k on tissue j comprised 30 cells, partitioned uniformly at random among the cell types present within the current region ($D_k^{(j)}$). The integer vector containing the number of cells belonging to each type was normalized to produce a simplex serving as the true composition vector $E_k^{(j)}$. For each cell type t , $30E_{k,t}^{(j)}$ cells were drawn at random (without replacement) from $G^{(t)}$, the subset

of cell profiles annotated as type t . Their expression profiles were summed to yield $y_{k,t}$ and then rounded to integer values to remove fractional reads introduced by ‘normalize_total’. As with the cluster-based approach, read counts were then pooled across all cell types, yielding spot-level reads $y_k = \sum_t y_{k,t}$, and composition vectors were then (optionally) pooled within high-level annotation categories. Observational noise was added in an identical fashion to the cluster-based approach. For all cell-based data, cSpotch was run with an NB likelihood (without zero inflation) to account for overdispersion present over gene counts in the snRNA-seq profiles. As with the cluster-based approach, spots were assumed to be fully independent; thus, ψ_i was suppressed in the GLM.

Power sampling and its effect on estimation

ST data of distal colon sections from 12-week-old mice were chosen for subsampling analysis given the large number of samples from these mice (six mice (three males and three females) with 13, 8, 9, 8, 8 and 6 tissue sections per mouse). To subsample the data, a given number of mice were randomly selected and a given number of tissue sections were selected from each mouse (if a selected mouse had fewer than the given number of tissue sections, all tissue sections belonging to that mouse were taken). Each subsampled dataset (nine combinations of one, two and six mice with two, four and eight tissue sections per mouse) and the full dataset (52 tissue sections and six mice) was analyzed separately by cSpotch. To compare the estimated posterior distributions of β_i at the gene and tissue context levels, the posterior means and s.d. were calculated, which were then used to obtain normal distribution approximations of the posterior distributions. The KL divergence between the posteriors derived from the full and subsampled data (that is, how much information is lost when using the distribution estimated from the subsampled data) was used to quantify the differences between the normal distributions.

Characterizing MCPs of gene expression

The cSpotch model output of posterior mean estimates of cellular gene expression $\{\beta_1, \dots, \beta_k\}$ for k cell types in a given spatial niche (for example, the crypt apex MROI), where each β_i is a matrix of dimensions $N_{\text{genes}} \times N_{\text{conditions}}$ that was used as input to identify gene sets spanning multiple cell types that show k -way correlation across the measured conditions. First, for each spatial niche, cell types and genes were selected for MCP analysis. Only cell types that are found in at least 5% of spatial spots, on average, were included, to focus on cell types with sufficient certainty in posterior estimates of gene expression. Across the k included cell types, the top 5% of gene signatures were considered on the basis of the coefficient of variation of β across the conditions to focus on genes with spatiotemporal variation.

Next, penalized matrix decomposition (PMD) was used to find linear combinations of gene signatures in each cell type that are maximally correlated across all conditions¹⁶⁴. PMD seeks to find sparse canonical variates $\{w_1, \dots, w_k\}$ that transform the original gene feature space into a new MCP feature space of dimension $M < k$:

$$\text{argmax}_{w_1, \dots, w_k} \sum_{i < j} w_i \beta_i \beta_j^T w_j \tag{15}$$

such that $\|w_i\| \leq 1, \rho_i(w_i) < c_i \forall i$, where $\rho_i(w_i)$ represent least absolute shrinkage and selection operator (LASSO) regression penalties and c_i controls the degree of sparsity. These tuning parameters were chosen by R’s MultiCCA.permute function¹⁰⁸. For each pair of cell types i and j ,

$$\text{argmax}_{w_i, w_j} w_i \beta_i \beta_j^T w_j = \text{argmax}_{w_i, w_j} \text{corr}(\beta_i^T w_i, \beta_j^T w_j) \tag{16}$$

Therefore, the canonical covariates identified a space in which each MCP signal in cell type i is highly correlated with the corresponding signal in all other cell types. Next, given canonical variates from PMD, each of the

M MCP signals was characterized by identifying at most $n = 250$ genes across all cell types that showed the strongest positive or negative contribution as measured by $\text{abs}(w) > 0.1$. The list of correlated genes for each MCP, both 'up' genes with positive weight and 'down' genes with negative weight, was then used as input to a Fisher's exact test to identify Kyoto Encyclopedia of Genes and Genomes (KEGG) functional gene sets enriched in the MCP. Each MROI was allowed at most k MCPs under the constraint that the bases for each MCP are orthogonal. As MCCA will always output programs in the same order, the maximum of k MCPs was calculated for a given spatial niche and latter programs that show low self-correlation among member genes (mean Pearson's $r < 0.3$) or high cross-correlation with genes from other programs (mean Pearson's $r > 0.05$) we optionally removed. Total MCP activity across conditions was calculated as a weighted sum of member gene activity ($y_m = \sum_i^k (\sum_g^G \beta_{i[g, \cdot]} w_{i[g, m]}))$), whereas cellular contributions to each MCP were calculated by binning positive and negative weights by cellular origin.

Spatiotemporal cellular composition analysis

To identify significant differences in cellular composition, each tissue section was treated as an independent sample and compositional estimates of all spots for a given MROI were pooled from the same section. Statistical significance across groups (for example, proximal–middle–distal samples) was assessed using Welch's t -test. A crypt gradient gene was defined as one that showed a significant ($\text{BF} > 2$, \log_2 fold change > 0.5) difference in expression between crypt base and apex and a monotonic change in expression from base to base and mid, mid and apex.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All processed data were deposited to the Single Cell Portal under accession number [SCP2595](#). All raw data were deposited to the National Center for Biotechnology Information Gene Expression Omnibus repository under accession numbers [GSE285985](#) and [GSE284137](#).

Code availability

All code is deposited on GitHub (<https://github.com/adaly/cSplotch>). A Google Cloud-enabled workflow and data exploration tool are available on Terra Firecloud (<https://app.terra.bio/#workspaces/techinno/cSplotch%20Workflow>).

References

156. Vickovic, S. et al. SM-Omics is an automated platform for high-throughput spatial multi-omics. *Nat. Commun.* **13**, 795 (2022).
157. Berg, S. et al. ilastik: interactive machine learning for (bio)image analysis. *Nat. Methods* **16**, 1226–1232 (2019).
158. van der Walt, S., et al. scikit-image: image processing in Python. *PeerJ* **2**, e453 (2014).
159. Liao, P.-S., Chen, T.-S. & Chung, P.-C. A fast algorithm for multilevel thresholding. *J. Inf. Sci. Eng.* **17**, 713–727 (2001).
160. Neubert, P. & Protzel, P. Compact watershed and preemptive SLIC: on improving trade-offs of superpixel segmentation algorithms. In *Proceedings of the 22nd International Conference on Pattern Recognition* (ed. O'Connor, L.) (IEEE, 2014).
161. Maniatis, S. et al. Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis. *Science* **364**, 89–93 (2019).
162. Carpenter, B. et al. Stan: a probabilistic programming language. *J. Stat. Softw.* **76**, 1 (2017).
163. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

164. Witten, D. M., Tibshirani, R. & Hastie, T. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* **10**, 515–534 (2009).

Acknowledgements

Work was supported by the Knut and Alice Wallenberg Foundation, Beijer Laboratory for Gene and Neuro Research, Royal Swedish Academy of Sciences, Swedish Society for Medical Research, Science for Life Laboratory, National Institutes of Health (NIH) 1U54 AG076040-01 and 1RM1 HG011014-01 (S.V.), Klarman Cell Observatory, Manton Foundation and Howard Hughes Medical Institute (A.R.). S.V. was supported as a Wallenberg Fellow at the Broad Institute of MIT and Harvard and as a Wallenberg Academy Fellow and SciLifeLab Fellow at Uppsala University. A.R. was an investigator of the Howard Hughes Medical Institute. We would like to thank the Flatiron Institute for providing computing resources that enabled the completion of this work. The Flatiron Institute is a division of the Simons Foundation.

Author contributions

S.V. conceptualized and designed the study with guidance from A.R. and R.B. S.V. performed the experiments with help from B.L., N.M., S.F., N.v.W. and E.D. A.D. analyzed the data with guidance from S.V. and R.B. and help from T.Ä., M.S.-E. and B.L. S.V. annotated the histological sections with help from O.K. and N.v.W. and guidance from G.G. S.V., A.D., F.C. and A.R. interpreted the data and wrote the manuscript with input from all the authors. All authors discussed the results. T.Ä., B.L., N.M. and S.F. contributed equally to this work.

Funding

Open access funding provided by Uppsala University.

Competing interests

A.R. was a founder and equity holder of Celsius Therapeutics, is an equity holder in Immunitas Therapeutics and, until August 31, 2020, was a scientific advisory board member of Syros Pharmaceuticals, Neogene Therapeutics, Asimov and Thermo Fisher Scientific. Since August 1, 2020, A.R. has been an employee of Genentech and equity holder in Roche. Since August 5, 2021, R.B. has been an employee of Genentech. S.V. is an author on patents applied for by Spatial Transcriptomics AB (10X Genomics). S.V. and A.R. are coinventors on PCT/US2020/015481 relating to this work. The remaining authors declare no competing interests.

Additional information

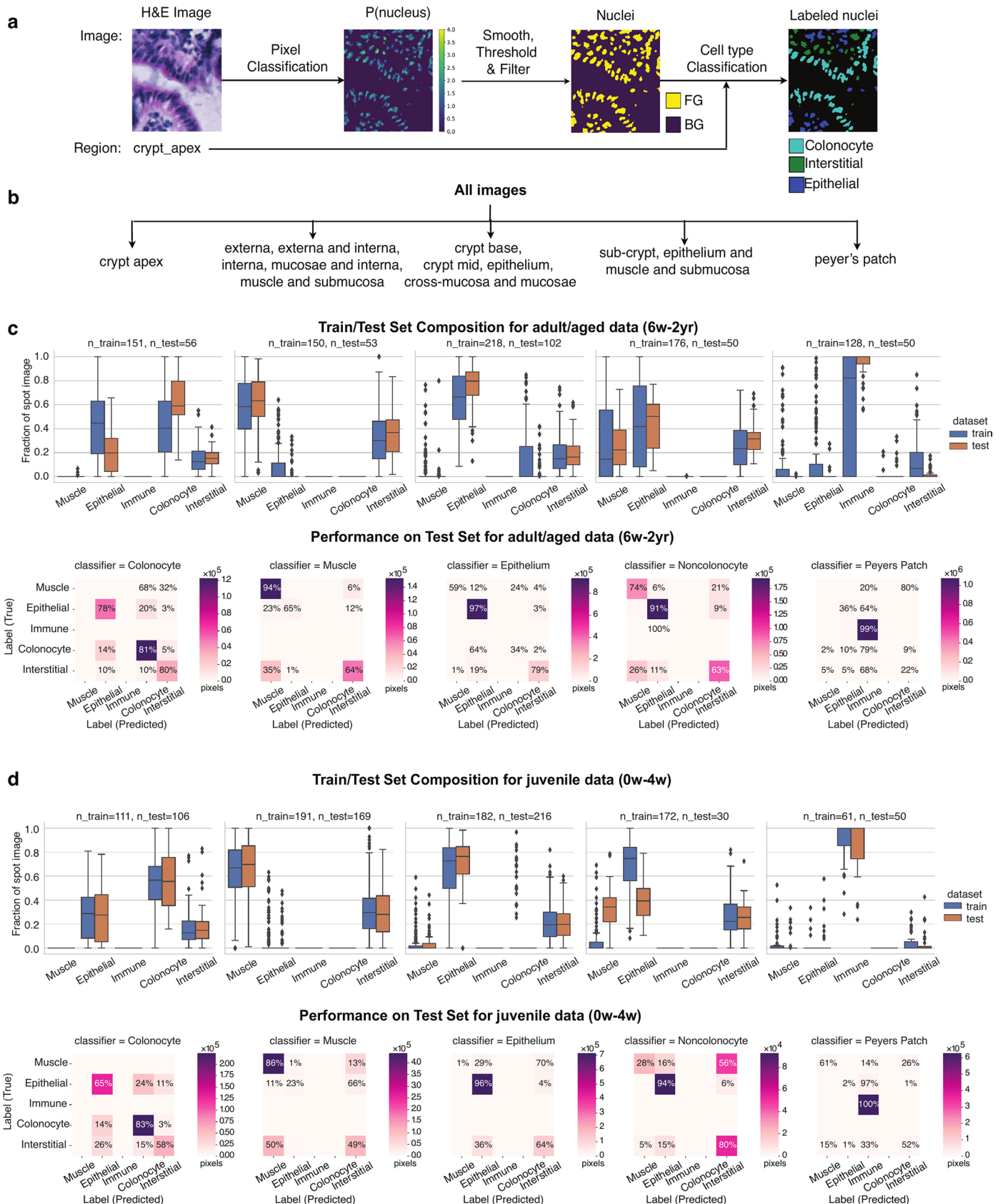
Extended data is available for this paper at <https://doi.org/10.1038/s41587-025-02830-6>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41587-025-02830-6>.

Correspondence and requests for materials should be addressed to Richard Bonneau, Aviv Regev or Sanja Vickovic.

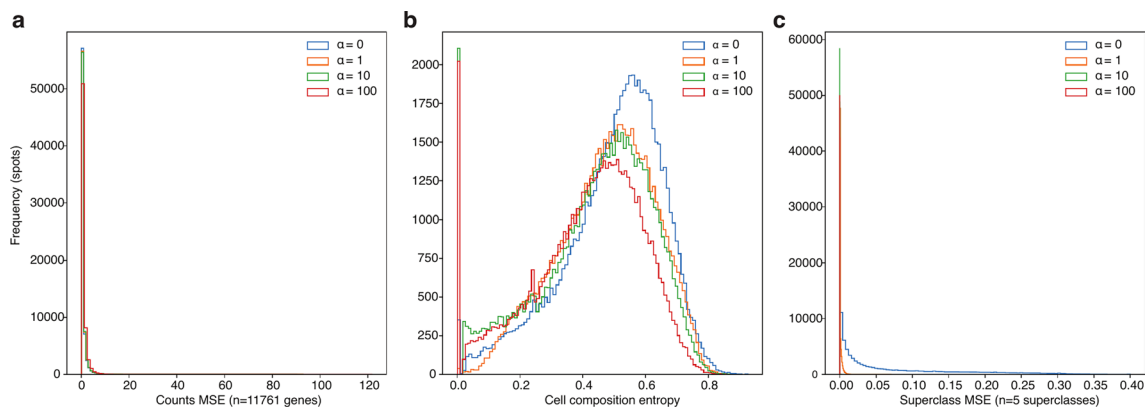
Peer review information *Nature Biotechnology* thanks Christopher Mason and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.



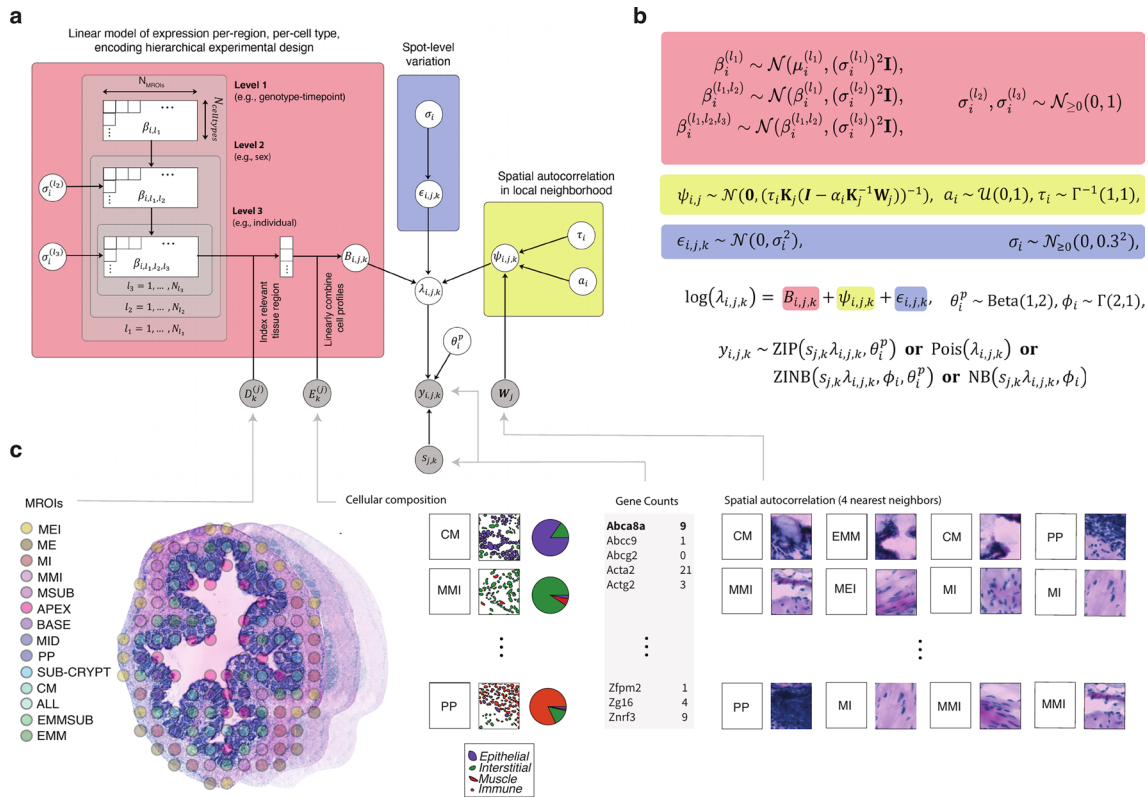
Extended Data Fig. 2 | Validation of semantic segmentation of morphological cell types (“super-classes”). (a,b) Semantic segmentation workflow. Semantic segmentation of cell nuclei (c), conditioned on five broad annotation groups for spots (b). (c,d) Accuracy of cell type composition by semantic segmentation. Fractions of FG pixels in each H&E image (y axis) assigned to each morphological superclass (x axis) in test and train sets (color code) in adult (c, top) and young

(d, top) mice, and percent and number (color scale) of pixels with a ground truth supercell label (rows) classified to each label (columns) in in adult (c, bottom) and young (d, bottom) mice. In Box plots: center black line, median; color-coded box, interquartile range; error bars, 1.5x interquartile range; black dots; outliers. Numbers of images used in train and test sets are denoted by “n_train” and “n_test”, respectively, above each subplot.



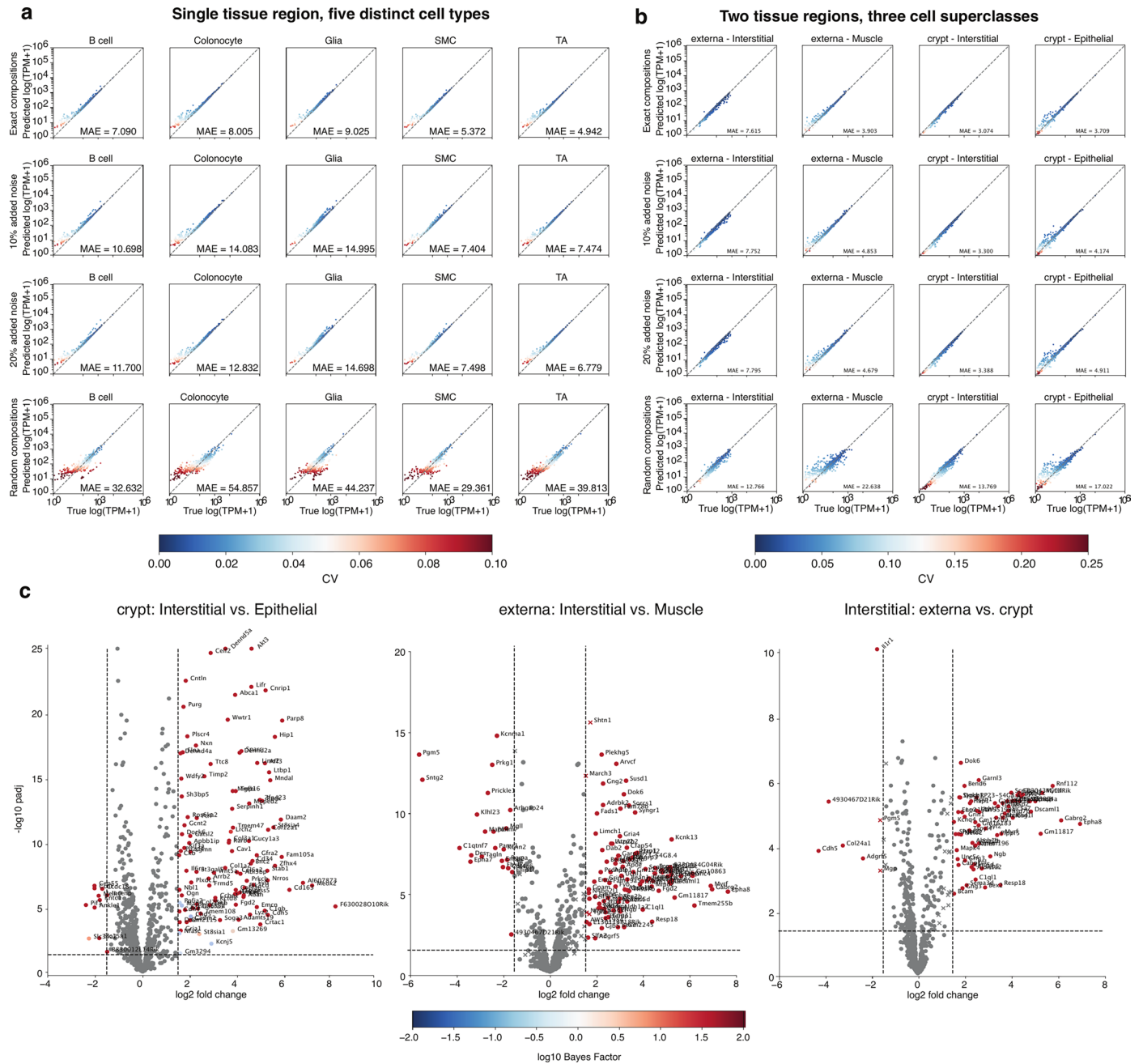
Extended Data Fig. 3 | Effect of morphological constraints on count-based deconvolution of ST data with NMF. (a) Reconstruction of expression by NMF⁴⁸ is unaffected by constrained weight parameter (α). Distribution (y axis) of mean squared error (MSE, x axis) between observed ($N = 11761$ genes) and SPOTlight-reconstructed expression (Methods) for $N = 69,721$ spots at different values of α (color code). (b) Cell compositions proposed by NMF decrease in entropy with increased α . Distribution (y axis) of entropy of cell composition vectors predicted by constrained NMF as in (a). Entropy (x axis) is calculated across all 17

snRNA-seq cell types ($k_{sn} = 17$), and thus ranges between 0 (one cell type present) and 1 (all cell types present in equal proportion). (c) Reconstruction of morphological cell compositions is greatly enhanced with increased α . Distribution of residual MSE between predicted and observed morphological cell type (“superclass”) composition vectors from SPOTlight as in (a). Morphological cell type vectors are calculated from the output of SPOTlight by pooling snRNA-seq cell types from the same morphological class (k_{morph} , as in Supplementary Table 3), reducing the space from $R_{0 \leq x \leq 1}^{k_{sn}=17}$ to $R_{0 \leq x \leq 1}^{k_{morph}=5}$.



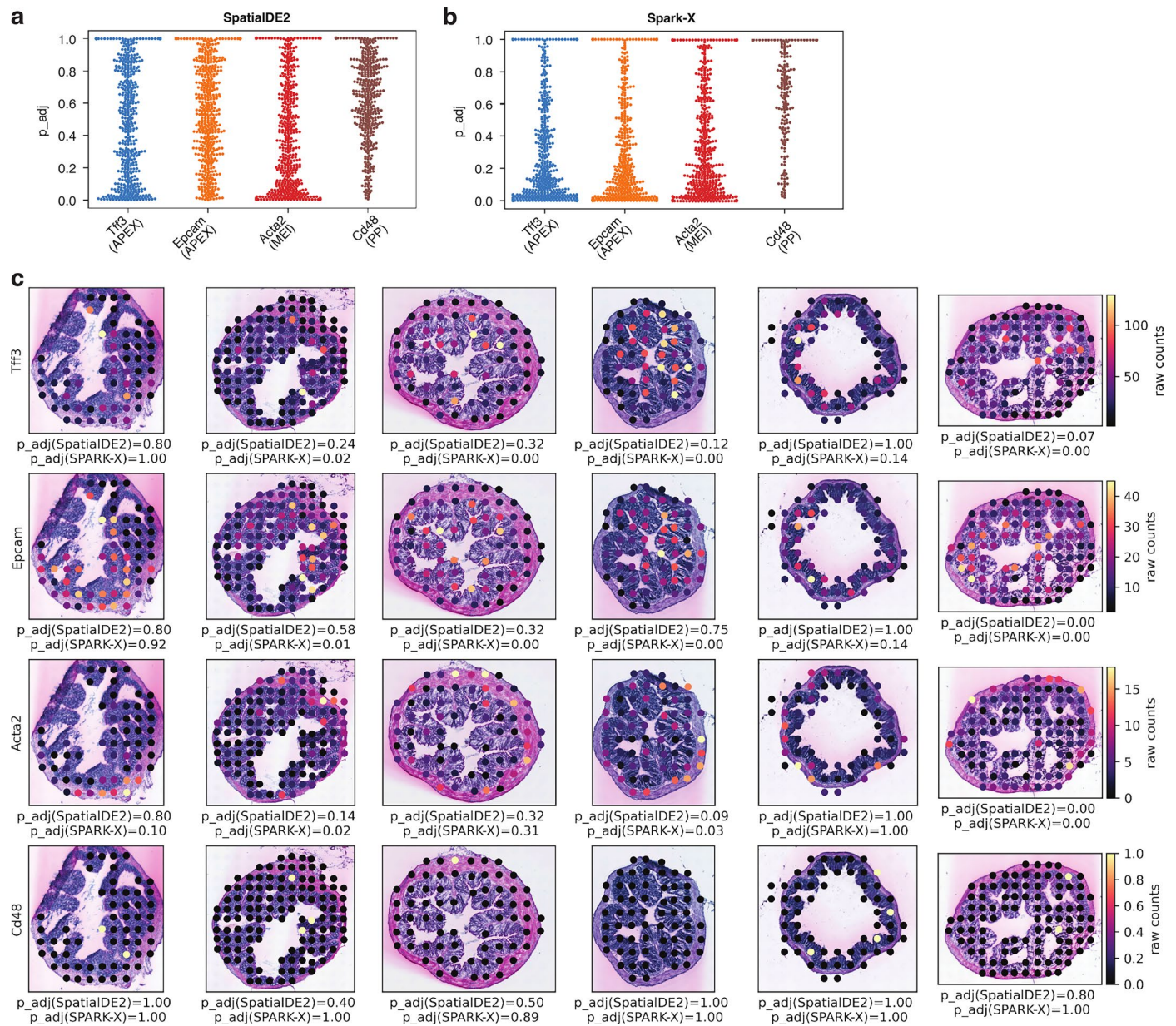
Extended Data Fig. 4 | cSpotch statistical model for ST data. (a, b) Underlying statistical model of spatial expression for a gene i at spot k on tissue j . (a) Characteristic expression rate β (red shaded area) for gene i is inferred separately in each cell type and MROI (white rectangles). Stacked gray boxes: three-level hierarchical formulation of β (l_1, l_2, l_3), where each level inherits from the one before it. Random variables ψ (yellow shaded area) and ϵ (blue shaded area) account for spatial autocorrelation and spot-level variation components, respectively. Gray and white circles: observed and latent variables, respectively. (b) Distributions over random variables of the statistical model in (a). ZIP, NB,

ZINB: zero-inflated Poisson, binomial, and zero-inflated negative binomial distributions, respectively. (c) Model inputs. A multi-tissue colon dataset (left) is annotated at the spot level with MROI tags (left color key) and cellular compositions (segmentation masks, pie charts; lower color key), encoded as observed variables $D_k^{(j)}$ and $E_k^{(j)}$, respectively. The expression measurement at spot k ("Gene Counts") yields the total sequencing depth of spot k ($s_{j,k}$), and the observed counts ($y_{i,j,k}$) of gene i (for example, *Abca8a*) therein. The (4-)neighborhood of each spot on tissue j (right) is encoded in the adjacency matrix W_j .



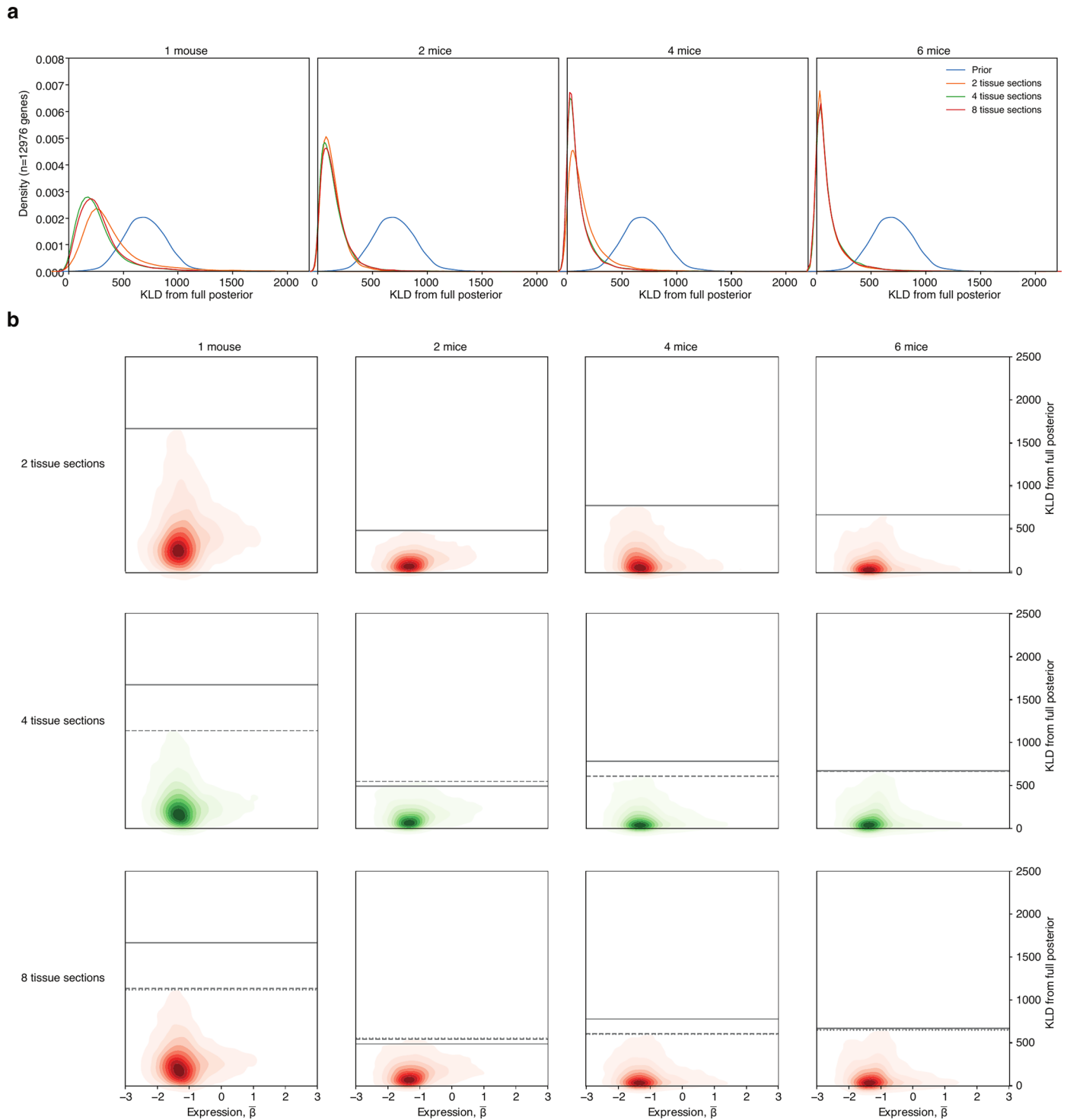
Extended Data Fig. 5 | Validation of cSpotch deconvolutional capabilities on simulated ST data. (a,b) Correlations between true and predicted counts for simulated cell type clusters. Recovered (y axis, $10^6 \exp(\hat{\beta})$, where $\hat{\beta}$ is the posterior mean expression in a given cell type) and observed (x axis, TPM, snRNA-Seq data) expression for each of 1,000 genes (dots) with real input (top row) or at different levels of corruption of cellular composition input (other three rows) in cells from each of five superclasses (columns) from a single tissue region (a) or in each of two cell types from each of two different tissue regions (b). Bottom right: Mean absolute error (MAE). Dotted line: $x = y$. Color scale: coefficient of variation (CV; standard deviation / mean) for Spotch posterior

estimates. (c) Differential expression effect size (log fold change; x axis) and significance ($-\log_{10}$ BH-adjusted p-value; one-sided t-test; y axis) for each gene (dot or X) between two cell types in one MROI (as labeled on top) based on snRNA-Seq data. Dots: genes showing agreement on DE assignment between snRNA-seq and Spotch X's: genes showing disagreement on DE assignment. Dots/X's for genes deemed DE from snRNA-seq ($|\log_{2}fc| > 1.5$, $p_{adj} < 0.05$) colored by \log_{10} Bayes factor (BF) of an analogous DE analysis between deconvolved cell profiles from cSpotch on simulated ST. Dashed vertical lines: $LFC = |1|$; dashed horizontal lines: $p_{adj} = 0.05$.



Extended Data Fig. 6 | Comparison to SpatialDE2 and Spark-X. (a–b) Swarm plots showing spatial variability scores (p_{adj}) output by (a) SpatialDE2 (two-sided LR test (asymptotically Chi-square w/one degree of freedom); Benjamini-Hochberg FDR correction) and (b) SPARK-X (one-sided weighted sum of Chi-squares; Benjamini-Yekutieli FDR correction) methods for four genes (x-axis) across $N = 734$ tissue sections. Each gene is known to be localized to

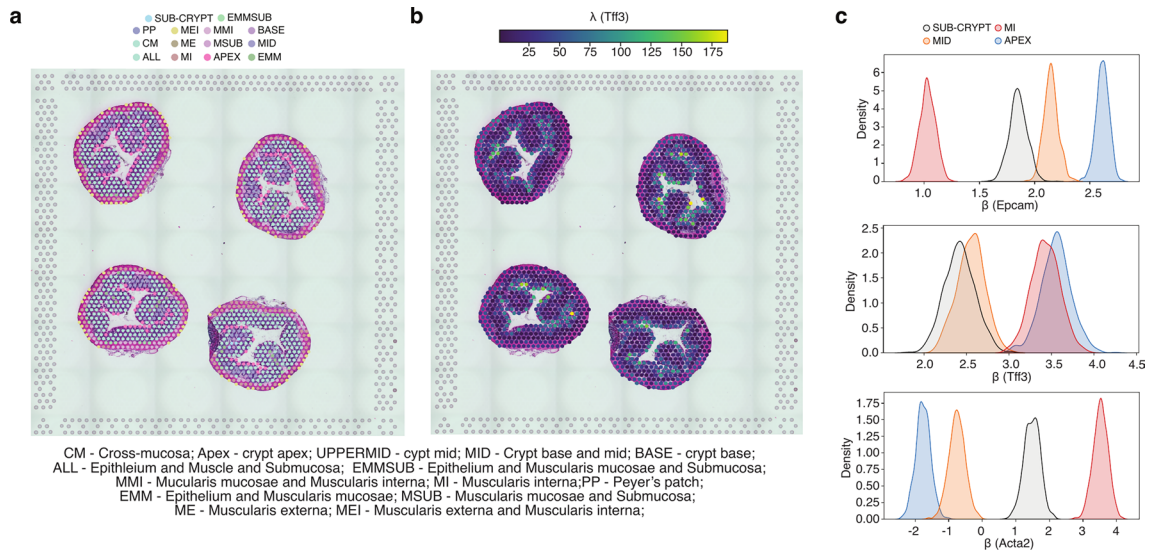
a specific MROI (parentheses) from IF imaging studies (Fig. 2g–i), and as such significant spatial DE is expected. (c) Observed spatial expression (color) of the same four genes (rows) on six tissue sections (columns) chosen at random from the dataset. For each gene on each tissue, adjusted p -values from SpatialDE2 and SPARK-X are denoted below the plot as $p_{\text{adj}}(\text{SpatialDE2})$ and $p_{\text{adj}}(\text{SPARK-X})$, respectively (identical test statistics and FDR correction to (a) and (b)).



Extended Data Fig. 7 | Effect of experimental design on study power.

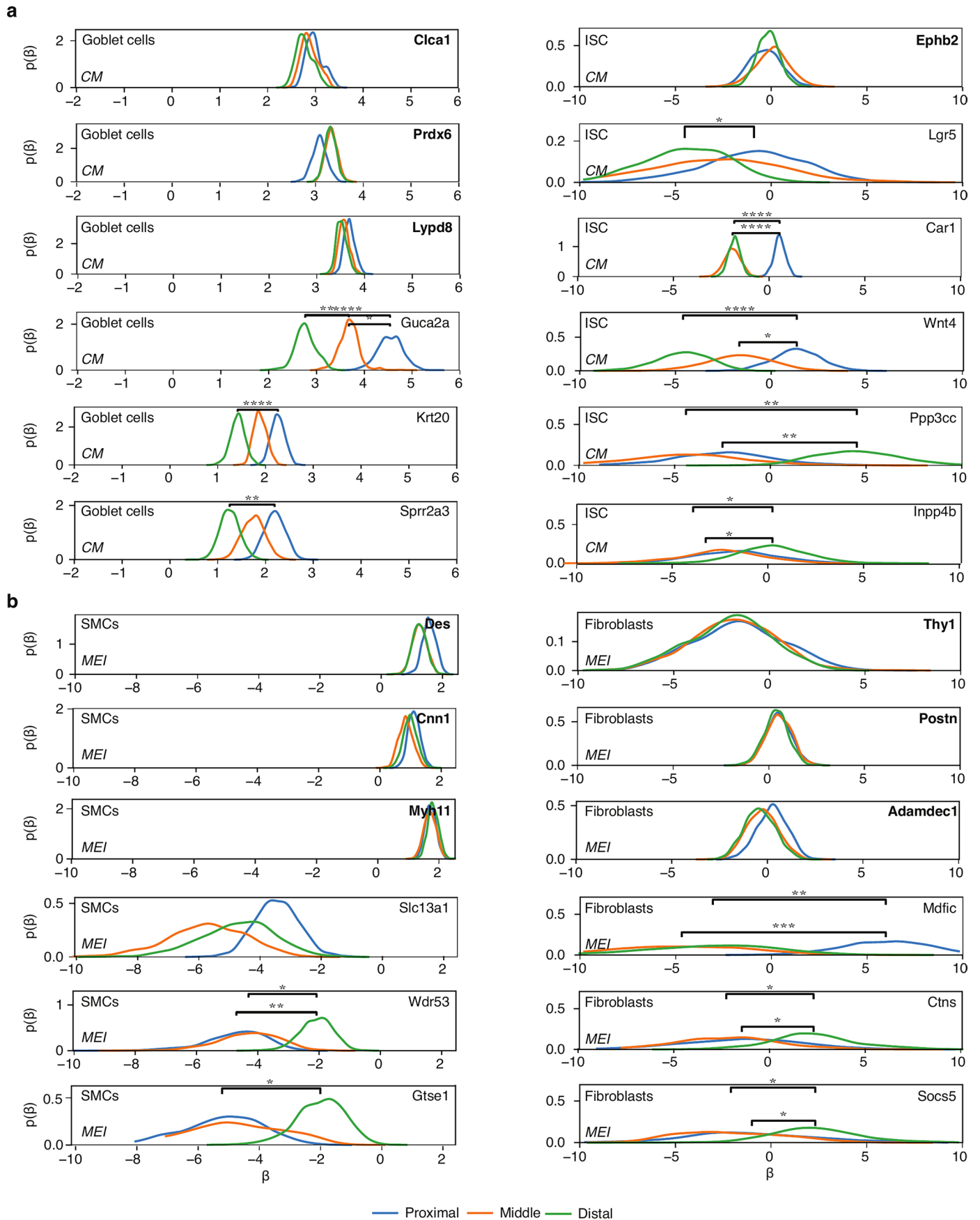
(a) Characteristic expression rates (β) in the cross-mucosa of the distal colon for 12-week-old mice. Distribution of Kullback-Leibler divergence (KLD, **Methods**) between posterior distributions of expression rates for $n = 12,976$ genes estimated from sub-sampled data (1, 2 and 4 mice; 2, 4 and 8 tissue sections per mouse) *vs.* the full data (6 mice, 53 tissue sections). Lower KLD values indicate greater agreement

between full and subsampled data. **(b)** Mean expression $\bar{\beta}$ (posterior mean estimated from the full dataset) (x axis) and KL divergence (y axis) for each gene between the estimate from a sub-sampling of mice (columns) and tissue sections (rows) *vs.* the full data for each of $n = 12,976$ genes. For each number of subsampled mice, solid, dashed, and dotted lines represent maximum KLD value attained across all genes for increasing numbers of tissue sections.



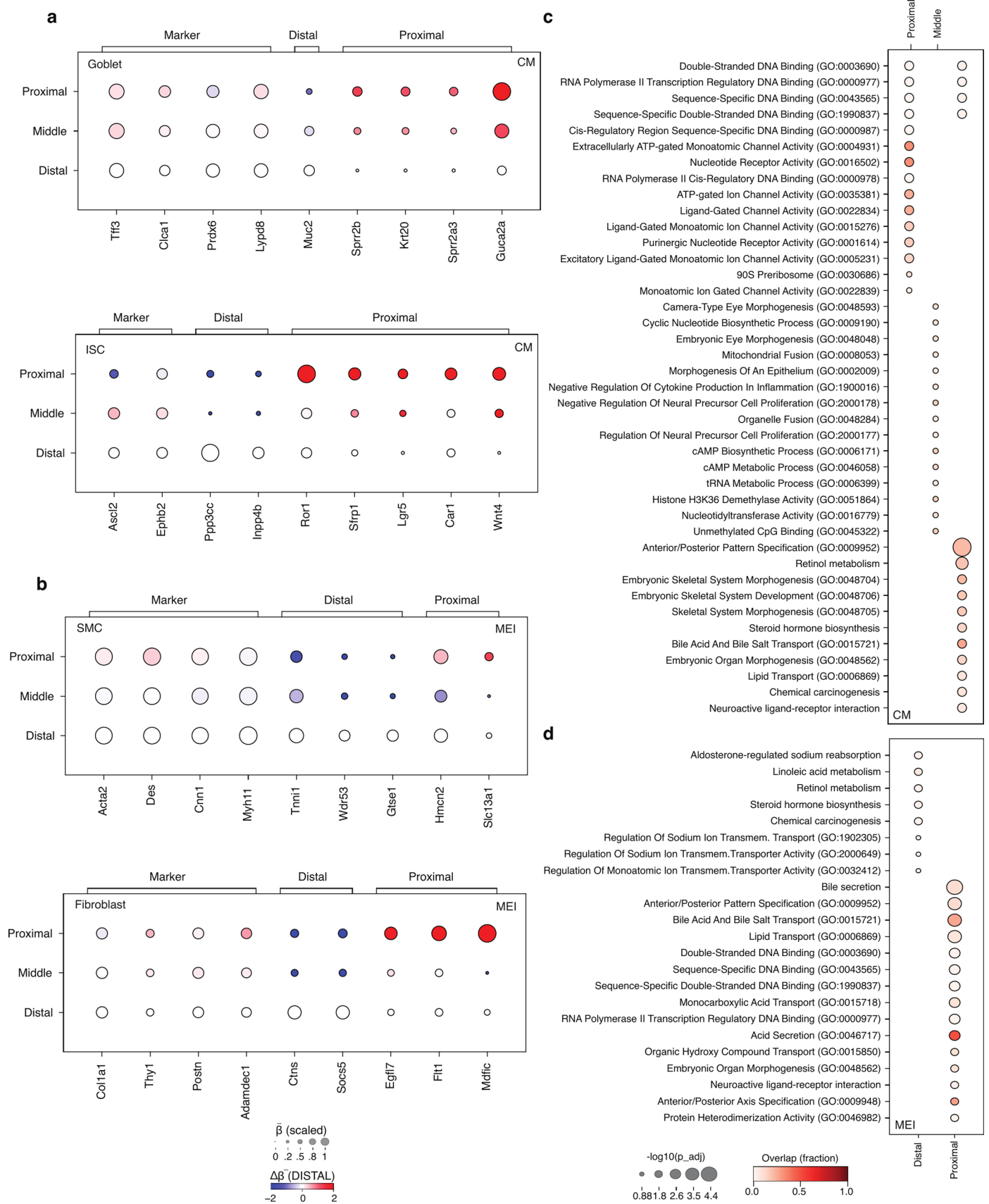
Extended Data Fig. 8 | cSpotch applied to Visium from 10X Genomics.
(a) Example Visium (10X Genomics) dataset of the proximal colon (age = 8 weeks). Spatial spots (color code) overlaid on top of the H&E image. **(b)** Inferred

Tff3 gene expression (λ ; color scale) overlaid on top of the H&E image as in **(a)**. **(c)** Characteristic expression rates (β) for three spatial markers as in Fig. 2g-i for proximal 8-week-old mice.



Extended Data Fig. 9 | Effect of colon region on inferred cell-type specific expression distributions. Posterior distribution of characteristic expression rate (β ; x-axis) of individual genes (labeled on top right) in each of the three

regions (color) in goblet cells (left) and ISCs (right) in CM (**a**), and in SMCs (left) and fibroblasts (right) in MEI (**b**). **Bold**: Canonical markers. Brackets: significant differential expression (*: BF > 2; **: BF > 10; ***: BF > 30; ****: BF > 100).



Extended Data Fig. 10 | Effect of colon region on inferred cell-type specific expression and associated pathways. Scaled (dot size) and absolute (dot color) change in mean log expression (β) of each gene (columns) in each colonic region (rows) relative to the DISTAL colon in goblet cells and ISCs in CM (a), and in SMCs

and fibroblasts in MEI (b). Brackets denote markers and region-specific genes. (c-d) GSEA analysis for regional DE genes marked by cSpotch in (c) CM and (d) MEI. Dot size: $-\log_{10}(p_{val})$. Color scale: Overlap between regional DE genes and GSEA process.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection We employed the following existing software in our data collection: Cell Ranger (v3.0), Spaceranger (v3.0), Vslide (v1.0.0), bcl2fastq2 (v2.20), ST Pipeline (1.7.6), STAR, HTseq-count (0.7.1), TagGD (0.3.6)

Data analysis We employed cSplotch (v.0.0.1), a custom code base for Bayesian analysis of ST/snRNA-seq data, in our analysis (<https://github.com/adaly/cSplotch>). We additionally employed the following existing software: Harmony (0.1.4), SpoTteR (0.0.1), Ilastik (1.4.1), scikit-image (0.18.1), GSEAPy (1.1.3), scanpy (1.11.1), Phenograph (1.5.3)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All transcriptomic data used in this study are available on GEO under accession numbers GSE284137 (spatial transcriptomics) and GSE285985 (snRNA-seq)

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	n/a
Reporting on race, ethnicity, or other socially relevant groupings	n/a
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	A minimum of five mice were samples per time point with a minimum of 9 tissue replicates per mouse. Minimum numbers of mice and replicates were derived by first over-sampling at a given time point, then incrementally downsampling and comparing the results of Bayesian inference between full and down-sampled posteriors using Kullback-Leibler (KL) Divergence.
Data exclusions	Both ST and snRNA-seq measurements were filtered using pre-established thresholds. ST spots with fewer than 100 UMIs were discarded, as were snRNA-seq nuclei with fewer than 800 UMIs, fewer than 800 unique genes, or more than 30% mitochondrial or ribosomal transcripts.
Replication	Select differential expression findings from modeled spatial transcriptomics (ST) data were successfully reproduced using both quantitative IF imaging and an alternative ST platform (10x Genomics' Visium technology). No attempts at reproduction were unsuccessful.
Randomization	All tissue-level covariates -- age, sex, and colon region from which sample was drawn -- are included sources of variation within our hierarchical Bayesian model.
Blinding	Blinding was unnecessary for this study, as all gene expression measurements were gathered in an objective, automated manner and all covariate groups are clearly defined by age, sex, and physical location.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	EPCAM (Biolegend, Alexa-647 labeled primary, clone G8.8, 1:100 dilution), TFF3 (Abcam, unlabeled primary, EPR26048-14, 1:100 dilution), SMA (Abcam, unlabeled primary, clone EPR5368, 1:100 dilution), FABP4 (Abcam, Alexa-647 labeled, clone EPR3579, 1:250 dilution), FABP7 (Abcam, unlabeled primary, clone EPR24033-13, 1:100 dilution), CD48 (Biolegend, APC labeled, clone HM48-1, 1:100 dilution), PRDX6 (Abcam unlabeled primary antibody, clone EPR3754, 1:500 dilution), CD74 (Biolegend unlabeled primary antibody, clone In1/CD74, 1:250 dilution), PDGFRFA (Abcam unlabeled primary antibody, clone EPR22059-270, 1:1000 dilution), goat anti-rat IgG (Life Technologies, Alexa Fluor 488 Goat anti-rat IgG, 1:100 dilution), donkey anti-rabbit IgG (Biolegend, Alexa 647 labeled, clone Poly6064, 1:100 dilution)
Validation	<p>All primary antibodies used in this study were validated for specificity and application by the manufacturers, as stated on their respective product datasheets. Where available, validation information was further confirmed by consulting antibody profiles in publicly accessible databases (e.g., UniProt, Human Protein Atlas, CiteAb) and by reviewing published literature citing the same clone in immunofluorescence or immunohistochemistry applications relevant to our tissue type. For each antibody, the clone, host species, labeling, and dilution are specified in the Methods section. In addition to the manufacturer-provided validation, each antibody was tested in-house by performing a serial dilution series on representative tissue sections. Dilutions spanned a range bracketing the manufacturer's recommended concentration; in most cases, dilutions of 1:20, 1:50, 1:100, 1:250, 1:500, and 1:1000 were tested. The optimal dilution was selected based on signal intensity, specificity of staining pattern, and minimal background. This ensured that each antibody produced the expected localization pattern for its target antigen under our staining conditions.</p> <p>EPCAM (BioLegend, Alexa-647, clone G8.8): Manufacturer validation for flow cytometry and immunofluorescence; documented in Human Protein Atlas with epithelial membrane staining. Literature reports in intestinal epithelia. (Gracz et al. 2010)</p> <p>TFF3 (Abcam, clone EPR26048-14): Validated by Abcam for IHC and IF, specific staining on goblet cells of mouse colon; literature reports in intestinal and glandular epithelia.</p> <p>SMA (Abcam, clone EPR5368): Validated by Abcam for IHC/IF in smooth muscle; abundant literature support.</p> <p>FABP4 (Abcam, Alexa-647, clone EPR3579): Validated for IHC/IF, positive staining in 3T3L1 cells; documented adipocyte-specific staining in Human Protein Atlas.</p> <p>FABP7 (Abcam, clone EPR24033-13): Validated for IHC/IF; literature support in astrocytes and glial cells.</p> <p>CD48 (BioLegend, APC, clone HM48-1): Manufacturer validation for flow cytometry and IF; reported in hematopoietic cells.</p> <p>PRDX6 (Abcam, clone EPR3754): Validated for IHC/IF; documented cytoplasmic pattern in relevant tissues, including muscle.</p> <p>CD74 (BioLegend, clone In1/CD74): Manufacturer validation for flow cytometry and IF; literature supports the staining in antigen-presenting cells.</p> <p>PDGFRFA (Abcam, clone EPR22059-270): Validated for IHC/IF; strong literature support for staining mesenchymal cells.</p> <p>Goat anti-rat IgG (Life Technologies, Alexa Fluor 488) and donkey anti-rabbit IgG (BioLegend, Alexa 647, clone Poly6064): Validated by the manufacturer for immunofluorescent detection of rat or rabbit primaries, respectively.</p> <p>In all cases, the final working dilution reported in the manuscript reflects the optimal concentration determined from our in-house titration experiments, which balances a high signal-to-noise ratio with the preservation of tissue morphology.</p> <p>Gracz, Adam D et al. "Sox9 expression marks a subset of CD24-expressing small intestine epithelial stem cells that form organoids in vitro." <i>American journal of physiology. Gastrointestinal and liver physiology</i> vol. 298,5 (2010): G590-600. doi:10.1152/ajpgi.00470.2009</p>

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	C57BL/6J mice sacrificed at 11 distinct time points: 0w, 1w, 2w, 3w, 4w, 6w, 8w, 12w, 6m, 1yr, and 2yr
Wild animals	n/a
Reporting on sex	Both male and female mice were collected for each age group (31 total male mice, 34 total female mice across 11 age groups), and sex was accounted for as a covariate within our hierarchical Bayesian model of spatial gene expression. All genes were assessed for sex-driven differential expression at each age group and spatial niche; these results are included in and commented on in our analysis.
Field-collected samples	n/a
Ethics oversight	Mice were maintained in accordance with ethical guidelines monitored by the Institutional Animal Care and Use Committees (IACUC) established by the Division of Comparative Medicine at the Broad Institute of MIT and Harvard and Columbia University, and consistent with the Guide for Care and Use of Laboratory Animals, National Research Council, 1996 (institutional animal welfare assurance no. A4711-01), with protocols 0122-10-16 and AABI3617, respectively.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks

n/a

Novel plant genotypes

n/a

Authentication

n/a