

Diverse infections transcriptionally reprogram the intestinal epithelium and epithelial–immune cell interactions

Received: 23 December 2025

Accepted: 20 August 2026

Published online: 07 October 2026

 Check for updates

Andrew Hart¹, Maria Merolle¹, Christian Howard¹, Breanne E. Haskins¹, Ian S. Cohn¹, Suhas Bobba^{1,2}, Rui Xiao¹, Yi Yang³, Ken Cadwell^{1,3}, Junjie Ma⁴, Hiroshi Yano^{5,6,7,8}, Xiaoxiao Hou^{5,6,7,8}, Bethan A. Wallbank¹, Daniel Cutillo¹, Ivaylo I. Ivanov⁴, Boris Striepen¹, Sunny Shin², Igor E. Brodsky¹, David Artis^{5,6,7,8,9,10,11}, Christopher A. Hunter¹ & Daniel P. Beiting¹ ✉ on behalf of the MIST Consortium*

The distal small intestine has a vital role in host physiology through its regulation of nutrient and fluid homeostasis. Despite being affected in Crohn's disease and serving as a major target for a range of infections, we know relatively little about the complexity of cellular responses and cell–cell communication in the ileum during infection. Here we present GutPath, an atlas of over 500,000 single cells with RNA and protein expression profiles for 91 cell states in the ileum across diverse infectious archetypes. We show that GutPath accurately captures established immune responses to infection while revealing pathogen-specific responses in enterocytes. To highlight the discovery potential of this atlas, we identify a new enterocyte transcriptional state present during *Yersinia pseudotuberculosis* infection that is spatially linked to bacterial load and tissue pathology. GutPath establishes a much-needed resource for the immunology community that will accelerate the study of transcriptional diversity in the cellular landscapes of the small intestine.

The small intestine constitutes a massive surface area tasked with carrying out essential digestive functions while simultaneously contending with a battery of immunological challenges, including tolerance to food-derived antigens, regulation of responses to commensals, and sensing and restricting infections^{1–3}. To meet these physiological and immune demands, the small intestine has evolved complex cellular environments with dozens of different cell types and unique metabolic and immune niches^{4–6}. The ileum is distinct from other parts of the gut in that it has a crucial role in nutrient and fluid balance as the primary site for the absorption of vitamin B₁₂ and bile acids, as well as being highly efficient at absorbing fat-soluble vitamins (A, D, E and K), water, sodium and chloride^{7,8}. The ileal epithelium regulates nutrient transport and serves as a physical barrier against infection. It comprises

absorptive enterocytes, goblet cells and their secreted mucins, Paneth cells, tuft cells and other specialized populations that collectively maintain gut homeostasis. Epithelial cell diversity in the gut varies between the small intestine and the colon, among different anatomical regions of the small intestine, and even within microanatomical structures such as villi^{4,8,9}. Intestinal stem cells located in the crypts of the villi give rise to nascent absorptive enterocytes that produce antimicrobial peptides. As these cells mature and advance toward the villus tips, tightly regulated transcriptional programs alter their functional capacity to perform nutrient transport^{4,10}. Interspersed with and underlying the epithelial barrier in the lamina propria (LP) are diverse immune cells that orchestrate immune tolerance and pathogen defense⁶. Finally, stromal populations and the enteric nervous system are essential for

A full list of affiliations appears at the end of the paper. ✉ e-mail: beiting@upenn.edu

tissue architecture, cell–cell communication and gut physiology, and they further contribute to cellular heterogeneity in the small intestine.

Because of the vital role of the ileum in nutrient and fluid balance, infection and inflammation of this tissue can lead to major disruptions in host physiology. For example, in patients with Crohn's disease, the ileum is a commonly affected part of the gastrointestinal tract, with one-third of all patients presenting with inflammation exclusively in the ileum¹¹—a condition associated with malabsorption of vitamin B₁₂ and bile acids^{12,13}. The ileum is also a common target for pathogenic infections that can have profound consequences, ranging from diarrhea and growth delays to death^{3,14,15}. The protozoan parasite *Cryptosporidium* infects the ileum and is a leading cause of diarrheal mortality and growth stunting in children globally¹⁴. Bacterial pathogens, including *Salmonella*, *Shigella*, *Campylobacter*, *Listeria*, *Yersinia* and some pathogenic *Escherichia coli*, gain access through Peyer's patches, which are abundant in the ileum. Likewise, many helminth parasites establish infection in the ileum, causing substantial morbidity worldwide¹⁵. Despite the complex and multifaceted roles of the ileum in host biology and defense, relatively little effort has been devoted to single-cell analysis of this tissue. Studies focused on small intestinal infections, including *Salmonella*, astrovirus and *Cryptosporidium* infections, have informed our understanding of how tissue-specific immunity is regulated in the gut; however, these studies are often limited to one or a few cell types, making their findings difficult to extrapolate or compare to those from other infections^{4,16–18}. Thus, large-scale single-cell studies that interrogate the small intestine across a wide range of pathogenic phyla are needed to capture how distinct organisms alter cellular heterogeneity and how pathogen-specific immunity is coordinated through transcriptional regulation.

The development of single-cell technologies, including single-cell RNA sequencing (scRNA-seq), that capture transcriptomic, proteomic and epigenetic data has expanded our understanding of mucosal immunology. Recent scRNA-seq studies of the gut have led to the discovery of new cell states^{4,7,9,10,19–23}, including newly identified enteroendocrine cell subsets and inflammation-associated changes to colonic intestinal stem cells during colitis^{4,23}. To date, intestinal scRNA-seq studies often deeply profile specific cell populations^{10,24–26} or catalog the steady state and developmental conditions of humans^{7,9,20,21} or mice^{4,19}. Intensive research efforts have focused on the large intestine because ulcerative colitis and colon cancer drive interest, biopsy specimens are readily available from human participants^{23,26–32}, and mouse models of inflammation are routinely used^{32,33}. Although data from these studies capture gastrointestinal biology at steady state and during various diseases, infection-induced immune processes are underrepresented^{19,20,23,34}. Consequently, we need to understand how immune responses are tailored to distinct classes of pathogens.

To address this critical knowledge gap, we created GutPath (<https://gutpath.org>), a single-cell transcriptomic and proteomic atlas of the ileum constructed through cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) across a range of viral, bacterial, fungal, protozoan and helminth mouse models of infection and colonization. This resource contains more than 500,000 cells from the ileum and the ileal-draining mesenteric lymph node (MLN). A total of 91 annotated cell states in the ileum and 49 cell states in the MLN were organized into multiple layers of reference annotation, ranging from broad cell lineages to refined cell states. We demonstrate that GutPath accurately captures the acute infection biology of six model pathogens affecting the distal small intestine. Furthermore, we identify pathogen-specific responses, including unique enterocyte transcriptional responses and cell–cell signaling programs that arise during *Nippostrongylus brasiliensis* and *Yersinia pseudotuberculosis* infections. Finally, using spatially resolved single-cell transcriptomics, we illustrate that unique enterocyte responses develop locally, regionally and throughout the tissue during infection.

Results

Creating a single-cell atlas of the small intestinal response to diverse microorganisms

To create GutPath, we focused on microorganisms that infect, colonize or influence the small intestine, while sampling a diverse range of species from distinct phyla, including segmented filamentous bacteria (SFB) (*Candidatus savagella* spp., a gram-positive commensal bacterium), *Y. pseudotuberculosis* (a gram-negative bacterium), *N. brasiliensis* (a parasitic helminth), *Candida albicans* (an opportunistic fungal pathogen), *Cryptosporidium parvum* (an apicomplexan parasite) and the murine norovirus (MNV) strain CR6 (a positive-sense single-stranded RNA virus in the family *Caliciviridae*) (Supplementary Fig. 1). All organisms were orally administered, except for subcutaneously injected *Nippostrongylus* third-stage (L3) larvae, and tissues were collected during the first week of infection for sequencing and infection confirmation (Supplementary Fig. 1 and Supplementary Table 1). Single-cell RNA and ~100 surface proteins were simultaneously profiled using CITE-seq of the MLN, the small intestinal LP and the intestinal epithelial cells (IECs) of the small intestine for at least two replicate mice per condition. *Nippostrongylus* is largely confined to the proximal small intestine. Thus, duodenal LP and IEC samples were also collected from the infected mice to capture the primary site of infection³⁵.

Across the MLN, IEC and LP fractions, mean reads per cell exceeded 30,000 for mRNA and 8,000 for antibody-derived tags (ADTs) (Supplementary Fig. 2a,b, dotted lines). After quality control analysis (Supplementary Fig. 2c), we retained 505,956 high-quality cells that were subsequently separated by tissue type for downstream analyses—184,973 and 320,983 cells from the MLN and ileum, respectively. These cells contained, on average, 5,867 UMI (unique molecular identifier) transcripts and 2,258 genes per cell in the MLN and 6,398 transcripts and 2,186 genes per cell in the ileum (Supplementary Figs. 2d, dotted line, and 3a). Between 40,000 and 100,000 cells were collected per condition across tissues (Supplementary Fig. 3b). To visualize these cells in two-dimensional space, mRNA and ADT data were integrated independently and combined through weighted nearest neighbor analysis (Fig. 1a,b). Cell types were classified using automated methods complemented by manual curation and validated by gene and protein expression profiles (Supplementary Fig. 3c–e). As expected, the MLN fraction predominantly comprised B cell, T cell and innate lymphoid cell (ILC) populations (collectively constituting 95% of MLN cells) (Fig. 1c and Supplementary Fig. 3f). Importantly, while epithelial cells comprised 80% of the IEC fraction (Fig. 1d and Supplementary Fig. 3g), further digestion of the small intestine to isolate the LP fraction yielded additional epithelial populations, particularly Paneth cells, stem cells and transit-amplifying cells (Fig. 1d and Supplementary Fig. 3g). Compared to the IEC fraction, the LP was enriched in mesenchymal and endothelial cells, cells of the enteric nervous system and B cells (Supplementary Fig. 3g). We also observed infection-specific alterations in the cellular composition of the LP fraction, with *Yersinia* inducing expansion of neutrophils, while *Nippostrongylus* infection led to expansion of tuft cells and ILCs (Supplementary Fig. 3g). In addition to expected patterns of protein expression (Supplementary Fig. 3e), we found that most proteins targeted by our ADT panel were either highly or moderately correlated with the expression of their corresponding genes (Supplementary Fig. 4a).

The size and cellular complexity of the small intestine make cell type and state annotation critical for fully capturing tissue heterogeneity. Yet, even with the advent of improved automated methods, high-resolution cell annotation remains a challenge^{4,9,23,36}. We used existing resources and extensive manual annotation to generate highly resolved cell labels for our atlas, defining 49 and 91 separate transcriptional cell states in the MLN (Supplementary Fig. 4b) and ileum (Fig. 1e), respectively. We identified early, middle and late stages of enterocyte differentiation¹⁰, as well as unique enterocyte states in mice infected with *Yersinia* (hereafter referred to as 'Yps enterocytes')

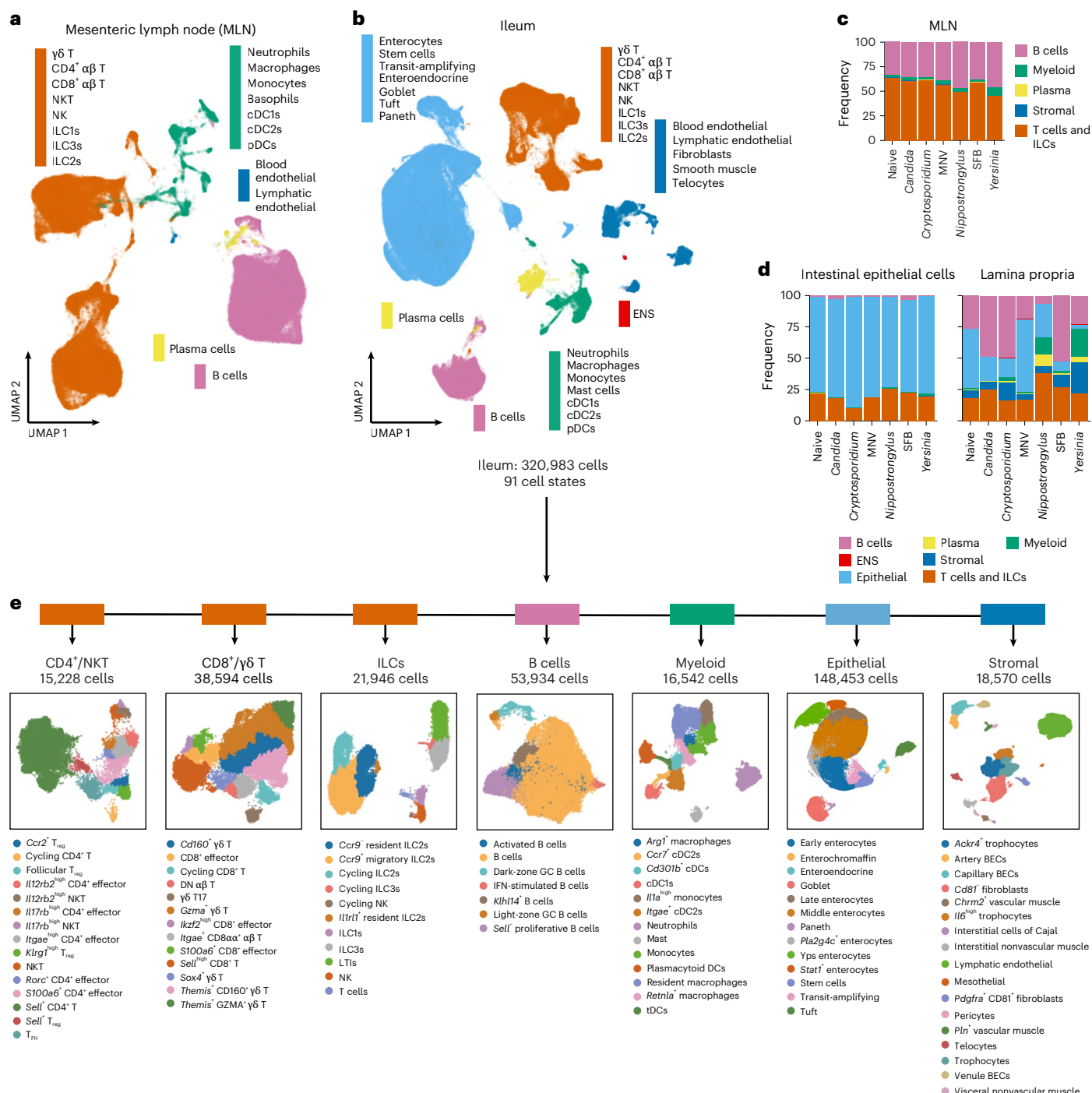


Fig. 1 | GutPath as a resource for mining mucosal responses to infection.

a,b, Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction of CITE-seq data, depicting major cell lineages (indicated by color) and annotated cell types in the MLN (**a**) and ileum (**b**) across all infections or colonizations. **c,d**, Frequencies of major cell lineages (indicated by color) in the MLN (**c**) for each condition and in the IECs (**d**, left) and LP (**d**, right) across infections or colonizations. **e**, UMAP plots and cell numbers of ileal cell subsets after reclustering and high-resolution cell annotation. Major cell lineages to which each subset belongs are indicated on the top edge by the same colors as

in **b**. Plasma cells and enteric nervous system lineages are not shown. Number of replicate mice for each condition: naive, $n = 2$; *Candida*, $n = 2$; *Cryptosporidium*, $n = 3$; MNV, $n = 2$; *Nippostrongylus*, $n = 2$; SFB, $n = 2$; *Yersinia*, $n = 3$. T_{reg}, regulatory T cells; T_{TH}, T follicular helper cells; T17, T cells characterized by their production of IL-17 and/or *Rorc* expression; DN, double negative; LTIs, lymphoid tissue inducer cells; GC, germinal center; DCs, dendritic cells; cDCs, conventional dendritic cells; pDCs, plasmacytoid dendritic cells; tDCs, transitional dendritic cells; BECs, blood endothelial cells; ENS, enteric nervous system.

or *Nippostrongylus* (*Pla2g4c*⁺ enterocytes) (Fig. 1e and Supplementary Fig. 4c). In the *Yersinia*-infected intestine, we also observed an inflammatory population of trophocytes (*Pdgfra*^{low} and expressing *Il6*, *Ccl2* and *Cxcl1*), previously reported to support intestinal stem cells and regulate their differentiation³⁷. Our atlas also captured several rare

cell types, including telocytes (*Pdgfra*^{high}) that coordinate intestinal stem cell Wnt signaling and differentiation³⁸, pericytes (*Cspg4*⁺ and *Pdgfrb*⁺) that stabilize vessels and support blood flow, enterochromaffin cells (*Lmx1a*⁺) that produce serotonin, and interstitial cells of Cajal (*Kit*⁺) that serve as the motility ‘pacemakers’ of the gut (Fig. 1e)^{39,40}.

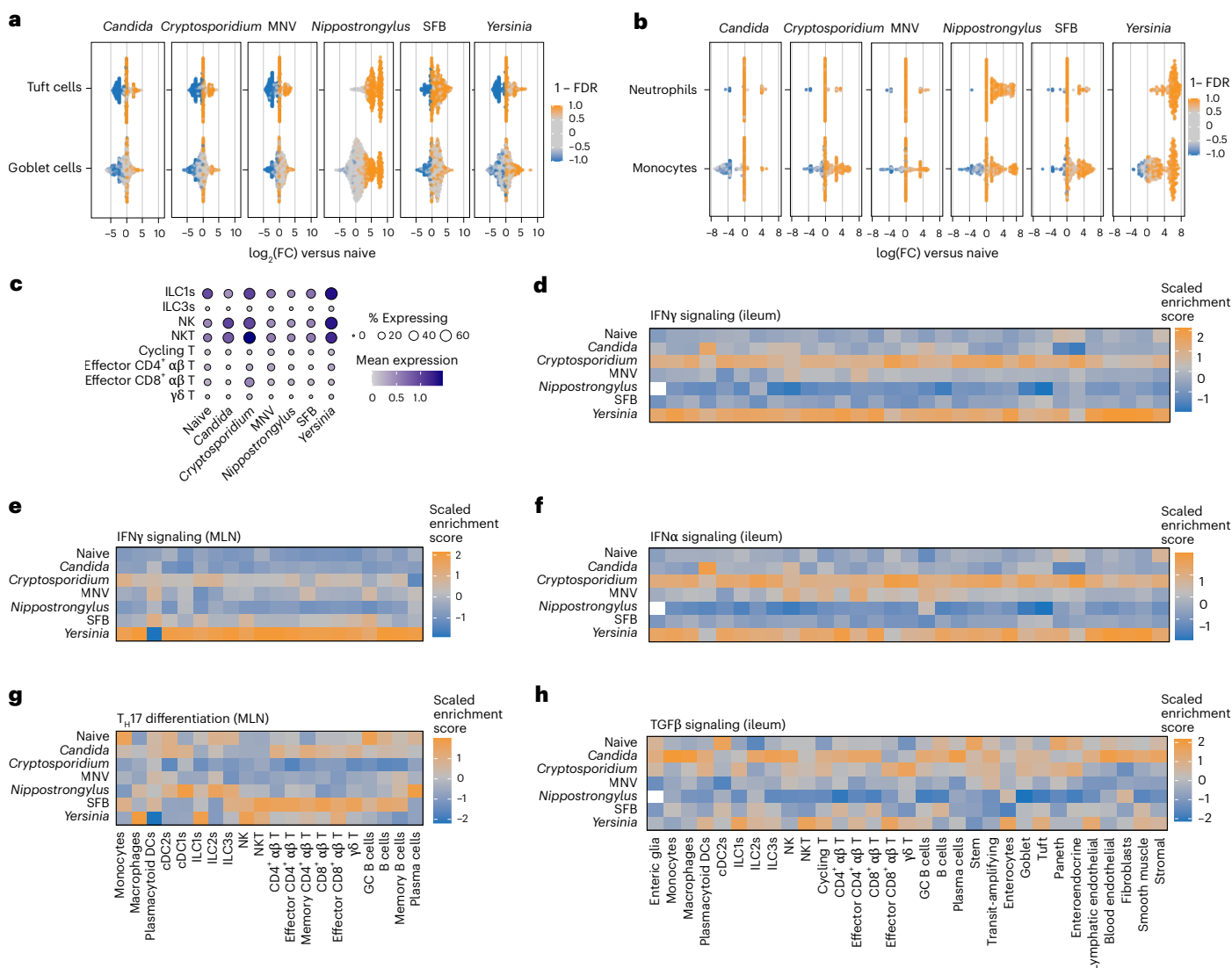


Fig. 2 | Canonical immune responses to infection are faithfully recapitulated in GutPath. a,b, Differential abundance analysis (two-sided quasi-likelihood *F* test implemented in *miR*) of defined cell neighborhoods for tuft and goblet cells (**a**) and neutrophils and monocytes (**b**) in the ileum. Significantly altered neighborhoods are indicated by color (false discovery rate (FDR)), with orange and blue indicating increased or decreased infection, respectively, compared to the naive condition. **c,** Frequency of cells expressing *Ifng* (indicated by bubble size) and the mean expression level of *Ifng* (indicated by color scale) across cell types and conditions in the ileum. **d,e,** Scaled UCell

signature scores of the Hallmark IFN γ signaling gene signature in the ileum (**d**) and MLN (**e**). **f–h,** Scaled UCell signature scores of the Hallmark IFN α signaling gene signature in the ileum (**f**), the Gene Ontology (GO) Biological Process (BP) T_H17 differentiation gene signature in the MLN (**g**) and the Hallmark TGF β signaling gene signature in the ileum (**h**). Panels **e** and **g**, and panels **d**, **f** and **h**, share the same x-axis labels. Number of replicate mice for each condition: naive, *n* = 2; *Candida*, *n* = 2; *Cryptosporidium*, *n* = 3; MNV, *n* = 2; *Nippostrongylus*, *n* = 2; SFB, *n* = 2; *Yersinia*, *n* = 3.

This detailed annotation of transcriptional states in GutPath constitutes a valuable resource that can be used for label transfer in mucosal studies, thus accelerating the analysis of scRNA-seq data in the field. To make these cellular descriptions accessible to researchers, complete marker profiles were generated for coarse, intermediate and refined cell state annotations (Supplementary Table 2) using a differential expression approach⁴¹. Additionally, we used random forest analysis to create condensed marker profiles for each cell state at each annotation level (Supplementary Table 3)⁴².

GutPath captures immune hallmarks of model infections

To test whether GutPath faithfully recapitulates pathogen-specific innate immune responses, we used differential abundance testing of cell ‘neighborhoods’ (defined by the *k* nearest neighbors) to examine both differential abundance of cell populations and transcriptional heterogeneity relative to naive samples. Differentially abundant

neighborhoods containing tuft cells and goblet cells were enriched in *Nippostrongylus* infection (Fig. 2a), but not in other infections, consistent with the tuft and goblet cell hyperplasia that is a feature of helminth infections (Supplementary Fig. 5a,b)⁴³. Rapid recruitment of monocytes, neutrophils and other myeloid effector cells is a dominant feature of many bacterial infection models, including *Y. pseudotuberculosis* infection⁴⁴, and we observed differentially abundant neighborhoods and increased frequencies of myeloid populations in the *Yersinia*-infected samples (Fig. 2b and Supplementary Fig. 5c,d). *Cryptosporidium* infection elicits interferon- γ (IFN γ) signaling and IFN γ -dependent immunity, and we observed increased *Ifng* expression in natural killer (NK) cells and natural killer T (NKT) cells at this early time point in the *Cryptosporidium*- and *Yersinia*-infected samples (Fig. 2c), as well as increased enrichment of the Hallmark IFN γ signaling gene set in cells from both *Cryptosporidium*- and *Yersinia*-infected mice (Fig. 2d)⁴⁵. Robust enrichment of this pathway was also observed in the

MLN of *Yersinia*-infected mice, likely due to the systemic nature of this infection (Fig. 2e)⁴⁶. The persistent CR6 strain of MNV is sensitive to STAT1-driven immunity and elicits a modest type I interferon response compared to other MNV strains^{47,48}. Consistent with this, we observed a modest enrichment of the type I interferon signaling gene set in NK cells and CD4⁺ T cells during MNV infection (Fig. 2f).

Colonization with *Candida* is complex, and the organism can behave as either a commensal or a pathogen in mice, depending in part on its interactions with the microbiome⁴⁹. *Candida* colonization has been linked to local induction of hypoxia in the intestine⁵⁰, and we observed moderate enrichment of hypoxia-related gene signatures in cells from *Candida*-colonized mice (Supplementary Fig. 5e). *Candida* is also known to promote the development of T helper 17 (T_H17) responses. We did not observe robust T_H17 responses to *Candida* in the ileum at 7 days post-infection (dpi). However, transforming growth factor- β (TGF β) signaling—a cytokine pathway that drives T_H17 differentiation and is systemically increased during *Candida* colonization^{51,52}—was enriched in cells from *Candida*-colonized mice (Fig. 2h). Additionally, *Candida* led to enrichment of the T_H17 differentiation gene set in T cell populations in the MLN (Fig. 2g). SFB colonization, another T_H17-inducing model, likewise showed enrichment of T_H17 differentiation in T cells compared to naive samples and other infection models (Fig. 2g)⁵³. Thus, these data show that GutPath accurately captures immunological archetypes of phylogenetically diverse model infections.

Infection alters enterocyte development and nutrient physiology

Given the clear induction of canonical immune pathways across the six infections (Fig. 2), we next performed an unbiased model-based differential gene expression analysis to identify unique responses to infection in the small intestine compared to the naive condition⁵⁴. We hypothesized that many pathways associated with tissue damage or general immune activation, including proliferation, might be conserved across different infections and cell types. Alternatively, as GutPath captures innate immune responses to phylogenetically diverse pathogens—and the immune system recognizes evolutionarily conserved molecular patterns—transcriptional responses to infection could be both cell type- and pathogen-specific. The magnitude of transcriptional changes varied greatly across infections and among cell types. While MNV had the lowest number of differentially expressed genes (DEGs), *Candida* colonization showed the weakest induction of DEGs as measured by the log-transformed fold change (log₂(FC)) (Fig. 3a and Supplementary Fig. 6a). Compared to MNV and *Candida*, *Cryptosporidium* and SFB both elicited more DEGs and a more pronounced induction of these DEGs, while *Nippostrongylus* and *Yersinia* infections resulted in the most robust transcriptional responses (Fig. 3a and Supplementary Fig. 6a). These responses mapped primarily to enterocytes and $\gamma\delta$ T cells, cell types that are particularly abundant in GutPath. However, when we set a more stringent cutoff for differential expression (FC > 1.5), IECs—including enterocytes, stem cells, transit-amplifying cells and goblet cells—as well as fibroblasts emerged as major drivers of the innate response across all infections (Fig. 3b).

Absorptive enterocytes provide the first cellular barrier to infections in the small intestine, but they are also critical regulators of host metabolism and exhibit a unique physiology that allows them to transport lipids, amino acids and nucleotides from the lumen of the intestine into the tissue¹⁰. The large number of IECs in GutPath, together with our finding that enterocytes, stem cells and transit-amplifying cells elicit robust transcriptional responses to infection, prompted us to further investigate pathogen-specific and conserved responses to infection in these populations. Strikingly, we observed that DEGs identified in stem cells, transit-amplifying cells and enterocytes showed little overlap between infections, suggesting substantial pathogen-specific regulation (Fig. 3c). We reasoned that pathogen-specific responses in IECs

might be explained by alterations in enterocyte differentiation. Stem cells at the base of the crypts progressively differentiate from stem cells to transit-amplifying cells and ultimately into immature enterocytes, which continue to mature and alter their gene expression as they move to the tip of the villus¹⁰. Using pseudotime trajectory inference, we captured this maturation process (Fig. 3d,e and Supplementary Fig. 6b). To validate our trajectory analysis, we examined enterocyte marker genes with well-established expression patterns, including *Lgr5* (stem cells), *Reg3b* (early enterocytes), *Slc5a1* (late enterocytes) and *Ada* (mature enterocytes at the villus tip). All four genes showed the anticipated ‘timing’ of expression in naive mice and across all infections combined (Supplementary Fig. 6c)¹⁰. Next, we evaluated the distribution of enterocytes by frequency along this pseudotime to gauge how each infection shapes enterocyte differentiation. *Candida* colonization and MNV infection did not shift enterocyte pseudotime compared to the naive condition (Fig. 3f). We observed a loss of mature enterocytes during infection with *Cryptosporidium*, a pathogen that exclusively infects enterocytes¹⁶, and a similar phenotype was observed during SFB colonization (Fig. 3f). In contrast, *Yersinia* and *Nippostrongylus* infections, which elicited the most DEGs in enterocytes (Fig. 3b), shifted pseudotime forward, indicating that these infections may drive enterocytes toward a more mature or distinct transcriptional state.

These results prompted us to explore enterocyte gene expression along pseudotime in greater detail. We used tradeSeq⁵⁵, a generalized additive model framework based on the negative binomial distribution, to analyze enterocyte gene expression along pseudotime during infection. We used naive samples as a baseline to define temporally regulated DEGs along the enterocyte pseudotime ($n = 4,278$ genes). Ontology analysis of these genes, as expected, pointed to cell cycling; however, pathways related to metabolism and protein transport were also highly enriched (Supplementary Fig. 6d–f and Supplementary Table 4), including many solute carriers⁵⁶. Solute carrier transporters are transmembrane proteins enriched on the apical and basal surfaces of enterocytes, and their dysregulation leads to metabolic disease^{10,57}. There is evidence that solute carriers are actively regulated during infection and participate in pathogenesis⁵⁸, but their role across different infections has not been explored. Among the 4,278 DEGs identified across pseudotime in the naive condition, 82 encode solute transporters (Fig. 3g), of which 29 (35%) were differentially expressed in at least one infection relative to naive samples (Fig. 3h) and generally followed one of two patterns. The first pattern is evident in genes such as *Slc28a2* (encoding a sodium-coupled nucleoside transporter) and *Slc6a6* (encoding a taurine transporter), which maintained the same timing across enterocyte development as observed in naive mice but were downregulated by *Yersinia* and *Nippostrongylus* infections (Fig. 3h). The second pattern occurred when infection altered the timing of expression, rather than its magnitude, as observed with genes such as *Slc20a1*, encoding a sodium-phosphate symporter that regulates phosphate absorption from interstitial fluid, and *Slc25a39*, encoding a mitochondrial protein responsible for glutathione uptake (Fig. 3h). Distinct expression patterns in pseudotime might indicate that these molecules are expressed in spatially unique niches along the villus or that their expression is induced within a larger shift of enterocyte cell state that alters pseudotime inference. Together, these data provide a detailed ontological map of genes involved in enterocyte physiology and underscore the impact of infection on the timing and magnitude of their expression.

Enterocytes exhibit pathogen-specific responses during infection

The changes we observed in solute transporter expression prompted us to assess how metabolic pathways downstream of intestinal nutrient transport might be altered in enterocytes during infection. We used scCellfie to score distinct metabolic tasks—specific functional activities that a cell performs or purposes that it serves through

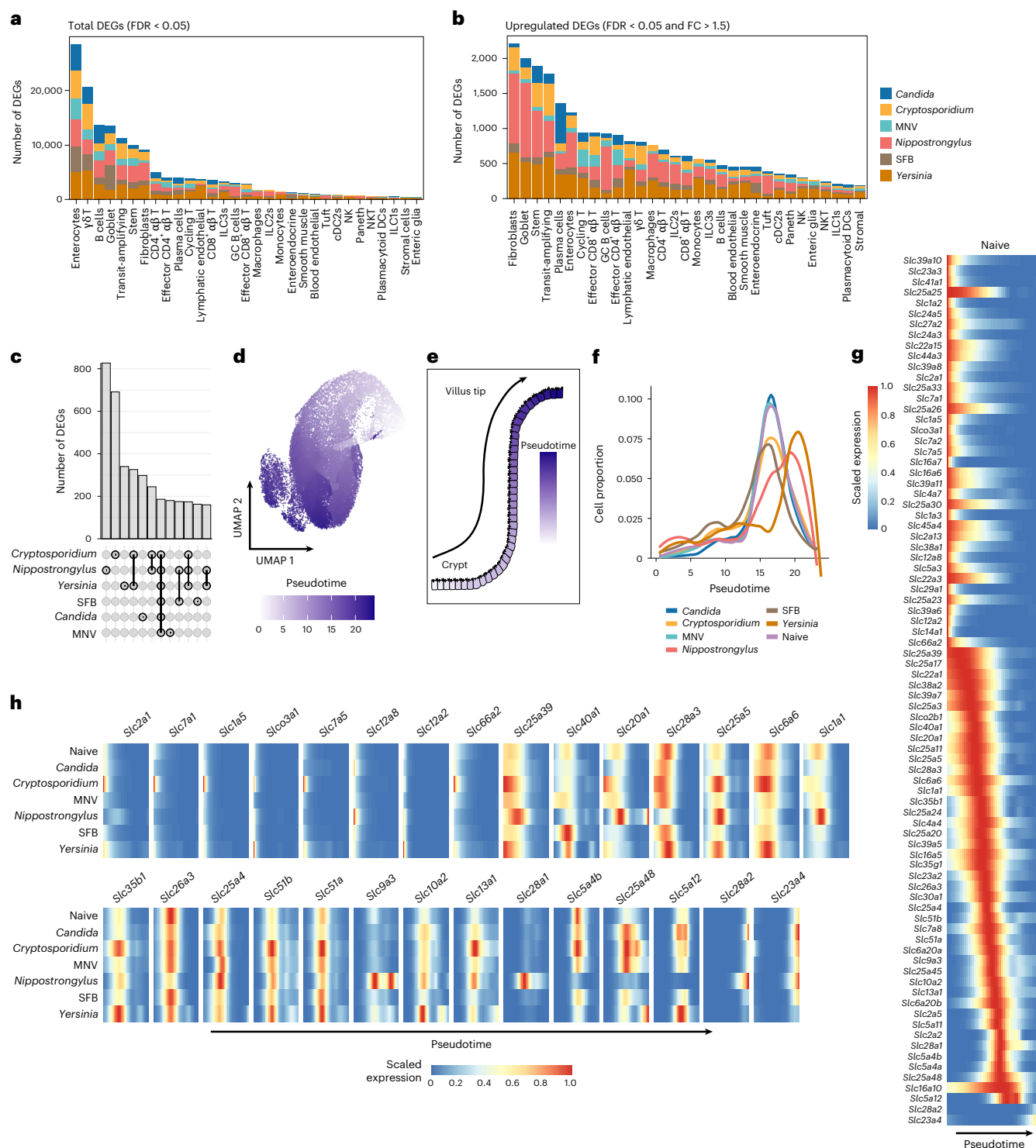


Fig. 3 | Disruption of enterocyte solute transport potential during infection.

a,b, Cumulative DEGs from MAST (Model-based Analysis of Single-cell Transcriptomics), relative to naive, for each cell type (represented by bars) and condition (indicated by colors) in the ileum at FDR < 0.05 (**a**) and FDR < 0.05 with an FC cutoff of >1.5 (**b**). **c**, UpSet plot showing the co-occurrence of upregulated DEGs from stem cells, transit-amplifying cells and enterocytes across all conditions. The number of DEGs for the 12 most numerous combinations is shown. **d**, UMAP reduction indicating the pseudotime progression (indicated by color hue) of reclustered stem cells, transit-amplifying cells and enterocyte populations from all conditions. **e**, Schematic showing the small intestinal villus, with pseudotime trajectory (indicated by color hue) representing stem cell-to-enterocyte differentiation. **f**, Histograms showing the proportional abundance of enterocyte lineage cells (stem cells, transit-amplifying cells and enterocytes)

across pseudotime for each condition (indicated by color). **g**, Pseudotime-scaled expression of selected solute transporter genes identified as significantly differentially expressed (tradeSeq generalized additive model followed by a two-sided Wald test for significance and Benjamini–Hochberg multiple-comparison correction) along pseudotime among stem cells, transit-amplifying cells and enterocytes from the ileum of naive mice. **h**, Pseudotime-scaled expression of solute transporter genes for each condition. Transporter ('Slc') genes differentially expressed (adjusted $P < 0.05$) in at least one infection relative to the naive condition are shown. Benjamini–Hochberg multiple-comparison correction was applied to the MAST differential expression analysis, and FDR < 0.05 was used as the significance threshold. Number of replicate mice for each condition: naive, $n = 2$; *Candida*, $n = 2$; *Cryptosporidium*, $n = 3$; MNV, $n = 2$; *Nippostrongylus*, $n = 2$; SFB, $n = 2$; *Yersinia*, $n = 3$.

metabolic pathways—across our cell populations and infection models⁵⁹. This analysis, restricted to the ileum, showed that enterocytes from *Nippostrongylus*-infected mice exhibited the most pronounced alterations in metabolic tasks, particularly in pathways involved in lipid metabolism (Fig. 4a). These alterations included reduced synthesis of secondary bile acids (Fig. 4b, gray) and altered metabolism of fatty acids such as elaidate, linoleate, palmitate and arachidonate (Fig. 4b, orange). The expression of genes associated with fatty acid processing tasks was increased in *Nippostrongylus* infection relative to all other conditions (Supplementary Fig. 7a), indicating a unique phenotype associated with helminth infection. Given the vital role of enterocytes in regulating solutes and metabolism, we hypothesized that this phenotype would be restricted to the epithelial compartment. Indeed, the median difference in activity score for select fatty acid-related tasks between *Nippostrongylus*-infected and naive samples across cell types was highest in enterocytes, irrespective of their maturity, and in goblet cells (Fig. 4c). We used cell–cell communication analysis to identify the signaling events that might drive this metabolic dysregulation. Compared to enterocytes in naive mice, enterocytes in *Nippostrongylus*-infected mice were predicted to receive interleukin-4 (IL-4) and IL-13 signals (Fig. 4d, right, arrow, and 4e), as expected in a T_H2 -dominant immune response. These T_H2 cytokines drive tuft cell and goblet cell expansion from intestinal stem cells, but their effects on functional capacity and signaling in absorptive enterocytes are not understood⁶⁰. Enterocytes from *Nippostrongylus*-infected mice were also predicted to produce unique ligands for cell–cell interactions not found in naive mice, including fibroblast growth factor (FGF) family ligands (Fig. 4d, left, arrow). FGF and CX3C family ligand-encoding genes, specifically *Fgf15* and *Cx3cl1*, are produced by enterocytes and were inferred to act on other cell populations (Fig. 4f). While CX3C chemokine ligand 1 (CX3CL1) is notable for its chemoattractant capacity for mast cells, FGF15 is a critical regulator of bile acid and cholesterol metabolism⁶¹.

To validate our findings, we first compared the observed in vivo enterocyte response during *Nippostrongylus* infection to a publicly available RNA-seq dataset from enteric organoids stimulated with IL-4 (ref. 62). Genes induced at the single-cell level in enterocytes of *Nippostrongylus*-infected mice were also upregulated by IL-4 in enteric organoids, including *Pla2g4c* (Supplementary Fig. 7b). Given that our data indicate marked metabolic remodeling in the small intestine during *Nippostrongylus* infection, we sought to confirm these findings using mass spectrometry-based metabolomic analyses. Ileal luminal contents showed altered levels of bile acids during *Nippostrongylus* infection, with the most pronounced reductions observed in cholic, chenodeoxycholic and β -muricholic primary bile acids (Supplementary Fig. 7c), a finding consistent with active production

and signaling of FGF15 (refs. 61,63). We also examined fatty acids and found that epithelial scrapings, but not luminal contents, had elevated levels of fatty acids in *Nippostrongylus*-infected mice compared to naive controls (Fig. 4g), suggesting that functional changes in the intestinal epithelium, as opposed to microbiota-driven metabolic changes, are responsible for this phenotype (Supplementary Fig. 7d). Fatty acid accumulation was more pronounced in mice with a greater worm burden (Fig. 4g, inset), linking pathogen load to metabolic disruption in the small intestine. Taken together, these data show that infection with *Nippostrongylus* reprograms ileal IEC metabolism and that T_H2 cytokines may have distinct effects on absorptive enterocytes that have yet to be explored.

***Yersinia* infection stimulates a unique proinflammatory program in enterocytes**

Given the pronounced impact of *Nippostrongylus* infection on enterocytes, and the fact that both *Nippostrongylus* and *Yersinia* infections resulted in the largest number of enterocyte-derived DEGs (Fig. 3a,b), we next questioned how *Yersinia* infection affects enterocytes. We examined the extent to which each infection produced a similar clustering pattern for enterocytes, measured as a scaled imbalance⁶⁴. Here, imbalance measures the local deviation from the expected cell distribution across conditions in the k nearest neighbors of an embedding and indicates where cells from one condition are preferentially abundant. This analysis showed that *Yersinia*, but not other infections, resulted in an imbalance in enterocyte clustering, suggesting that *Yersinia* may give rise to unique enterocyte transcriptional states (Fig. 5a). To explore this further, we applied tradeSeq again to determine differential expression along enterocyte pseudotime compared to naive animals and found that *Yersinia* elicited more temporally regulated DEGs than all other infections combined (adjusted $P < 0.05$ and $\log_2(FC) > 2$) (Fig. 5b). Hierarchical clustering of *Yersinia* DEGs identified a set of genes whose expression was typically high early in enterocyte pseudotime in naive animals but was altered in pattern by *Yersinia* infection—exhibiting either extended temporal expression or reactivation later in enterocyte pseudotime (Supplementary Fig. 8a, bracket). Expression of these genes was enriched in areas of higher imbalance and corresponded to a distinct *Yersinia*-driven cluster of mature enterocytes (Fig. 5c and Supplementary Fig. 7b). We refer to these cells as Yps enterocytes for their origin in *Y. pseudotuberculosis* infection (Fig. 5d,e and Supplementary Fig. 7c). Yps enterocytes represent only a subset of all enterocytes in *Yersinia*-infected mice. To identify the genes that uniquely define this subset, we compared Yps enterocytes to all other enterocytes during *Yersinia* infection, resulting in 132 Yps enterocyte-associated genes (Supplementary Table 5).

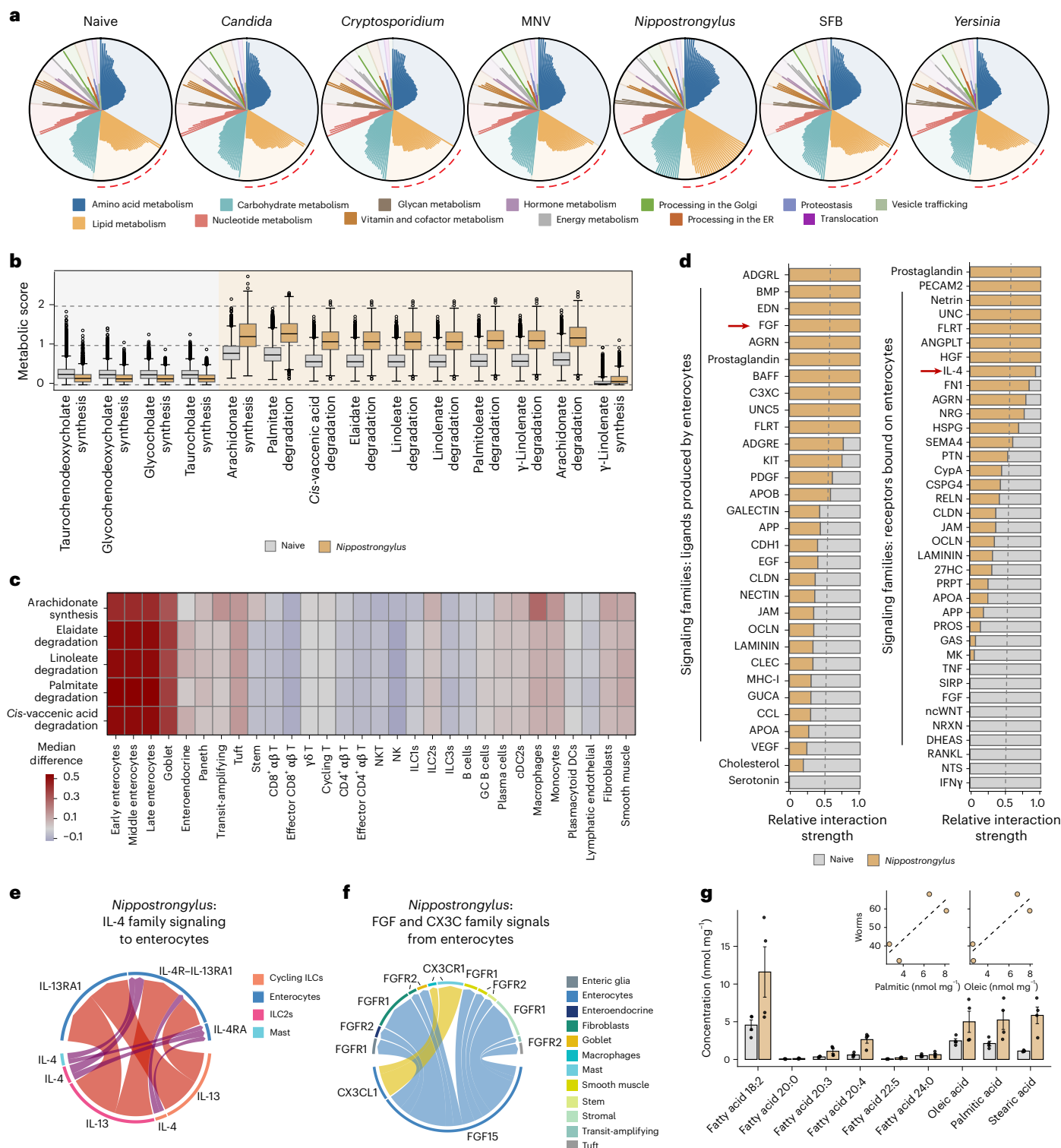
Fig. 4 | Metabolic reprogramming of enterocytes during *Nippostrongylus* infection. **a**, scCellfie metabolic task scores for enterocytes from each condition. Radial plots depict scaled activity scores across metabolic categories (indicated by colors) and individual metabolic tasks (represented by bars). Outer red dashed line highlights the 'lipid metabolism' metabolic task, which showed the most pronounced difference between *Nippostrongylus* infection and all other conditions. ER, endoplasmic reticulum. **b**, Significantly altered (adjusted $P < 0.05$, absolute Cohen's $d > 0.5$, absolute $\log_2(FC) > 0.585$ (-1.5 FC)) lipid metabolism task score distributions for enterocytes from *Nippostrongylus*-infected (orange) and naive (gray) samples. Outlier points (cells) are shown. Statistical significance of differential metabolic activity was evaluated using a two-sided Welch's t -test with multiple-testing correction through the Benjamini–Hochberg FDR method. Box plots display median values (center line), 25th (Q1) and 75th (Q3) percentiles (box boundaries), and whiskers extending to the minimum and maximum values within $1.5 \times$ the interquartile range. **c**, Heatmap of median difference showing the change in metabolic task score activity across cell types between *Nippostrongylus*-infected and naive samples. **d**, Cell–cell communication signaling predictions for all signals produced (left) or received (right) by enterocytes from *Nippostrongylus*-infected mice relative to signals

predicted in enterocytes from naive mice. Significantly different predictions are shown ($P < 0.05$), with signals from the FGF and IL-4 families indicated by red arrows. **e**, Chord diagram of IL-4 family signals produced by various cell types (circumference bar colors) and received by IL-4 family receptors on enterocytes during *Nippostrongylus* infection. **f**, Chord diagram of FGF and CX3C family signals produced by enterocytes and received by various cell types (circumference bar colors) during *Nippostrongylus* infection. **g**, Concentration of fatty acids determined by mass spectrometry in epithelial scraping samples of the ileum from naive or *Nippostrongylus*-infected mice (mean $\pm$ s.d.). Points represent individual mice. Data were normalized to sample weight (nmol mg^{-1}). Pairwise comparisons were performed using the Wilcoxon rank-sum test, with Benjamini–Hochberg correction for multiple comparisons. To test for global significance in fatty acid measurements across conditions, we used permutational multivariate analysis of variance ($P = 0.03$, $R^2 = 0.477$). Insets show the correlation between palmitic or oleic acid and worm burden in each mouse. Number of replicate mice for each condition in **a–f**: naive, $n = 2$; *Candida*, $n = 2$; *Cryptosporidium*, $n = 3$; MNV, $n = 2$; *Nippostrongylus*, $n = 2$; SFB, $n = 2$; *Yersinia*, $n = 3$. In **g**: naive, $n = 4$; *Nippostrongylus*, $n = 4$.

Among these, *Saa1* and *Saa2*—genes encoding acute-phase proteins—were previously shown to be upregulated in epithelial cells during *Salmonella* infection⁴ and SFB colonization⁶⁵. Other genes, including *Nos2* and *Reg3a*, encode effectors that are thought to directly limit bacterial survival and replication. Yps enterocyte genes were significantly associated with immune defense ontology terms, further indicating that this unique cell state may represent a direct adaptation of enterocytes to immune stimuli (Fig. 5f and Supplementary Table 6).

To identify potential upstream factors and signals that give rise to Yps enterocytes, we performed transcription factor motif enrichment

analysis using the 132 genes that define the Yps enterocyte phenotype. This analysis identified SREBF1 (normalized enrichment score (NES) = 4.69) (Supplementary Fig. 8d), a regulator of fatty acid synthesis, as well as STAT3 (NES = 3.30), a downstream effector of many cytokine signaling cascades (Supplementary Table 7). The expression of STAT3 target genes was compared across enterocyte subsets in *Yersinia*-infected and naive mice, revealing that the STAT3 signature was specific to Yps enterocytes (Fig. 5g). The separation of Yps enterocytes from other mature enterocytes in *Yersinia*-infected mice has interesting implications and may reflect spatially restricted bacterial



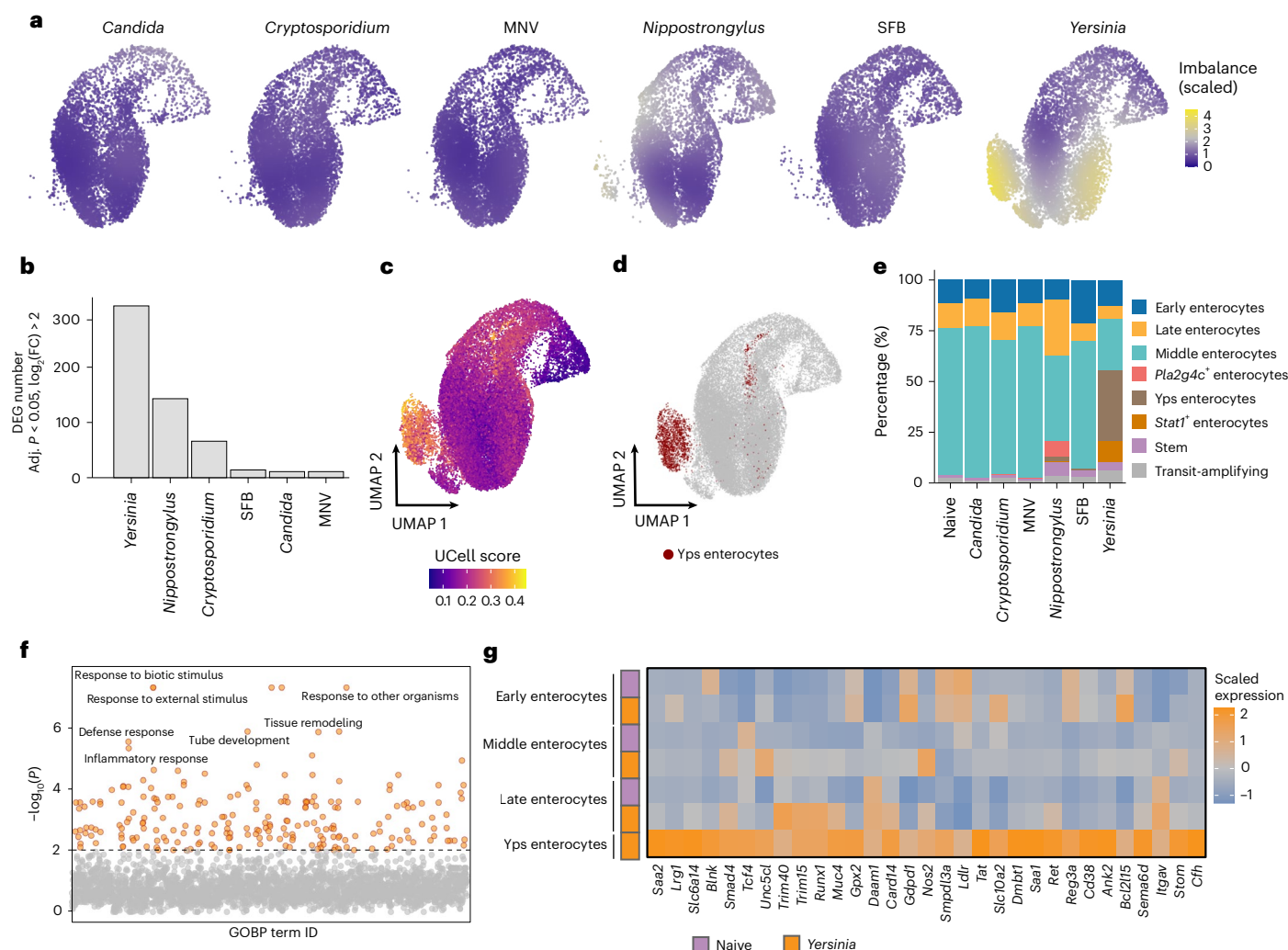


Fig. 5 | *Yersinia* infection gives rise to a unique enterocyte transcriptional state. **a**, UMAP embeddings depicting cell imbalance scores (indicated by color) for each infection. **b**, DEG counts identified from pseudotime-modeled comparisons of the enterocyte lineage (stem cells, transit-amplifying cells and enterocytes) within each infection relative to naive samples. A total of 5,000 highly variable genes were included for modeling and testing (adjusted $P < 0.05$, Benjamini–Hochberg; $\log_2(\text{FC}) > 2$). **c**, UCell single-cell enrichment (indicated by color hue) of select DEGs identified in enterocytes of *Yersinia*-infected samples, visualized on a UMAP dimensionality reduction plot. **d**, Yps enterocytes (red) visualized on a UMAP dimensionality reduction plot. **e**, Frequencies of enterocyte and stem cell subset populations (indicated by color) within each

condition. **f**, GOBP results of DEG analysis in Yps enterocytes compared to other enterocyte populations ($n = 132$ genes) from *Yersinia*-infected mice. The seven most significantly enriched ontology pathways are labeled. GO enrichment was performed using overrepresentation analysis (Fisher's exact test) with Benjamini–Hochberg multiple-testing correction. **g**, Gene expression heatmap of RcisTarget-defined genes containing an enriched STAT3 motif among Yps enterocyte-defining genes. Motifs 500 bp upstream and 1,000 bp downstream of the transcription start site were included for enrichment, and gene expression is shown for enterocyte populations in naive and *Yersinia*-infected mice. Number of replicate mice for each condition: naive, $n = 2$; *Candida*, $n = 2$; *Cryptosporidium*, $n = 3$; MNV, $n = 2$; *Nippostrongylus*, $n = 2$; SFB, $n = 2$; *Yersinia*, $n = 3$.

growth or immune signaling. Collectively, these data identify a distinct proinflammatory enterocyte population induced by *Yersinia* infection, potentially contributing to ongoing immune responses.

Spatially localized inflammation during *Yersinia* infection remodels the ileal epithelium

Y. pseudotuberculosis infection results in macroscopic, dense foci of inflammatory cells in the small intestine called pyogranulomas (PGs)⁴⁴. PGs are highly enriched in *Yersinia* relative to the surrounding tissue and accumulate neutrophils and monocytes, which promote host defense through TNF and IL-1 production^{44,66}. To investigate the relationship between Yps enterocytes and the formation of PGs, we performed spatial transcriptomic profiling of the small intestine of naive and *Yersinia*-infected mice (Supplementary Fig. 9a). Cell segmentation using Proseg⁶⁷ and quality control filtering yielded 680,410 cells across two biological replicates each from naive and *Yersinia*-infected mice

(Supplementary Fig. 9b,c). Using a panel of 5,000 gene probes, we obtained an average of 566 reads per cell and 126,422–207,476 cells per sample (Supplementary Fig. 9c). Using GutPath to guide annotation of this data (Supplementary Fig. 9d), we observed differential clustering of cells indicative of cell state changes in the enterocyte compartment of *Yersinia*-infected mice (Fig. 6a).

To examine the fidelity of our spatial transcriptomic data, we first asked whether the localized expression of specific solute transporters—predicted based on the pseudotime analysis of enterocytes in our atlas (Fig. 3)—was recapitulated in situ in sections from the naive and infected ileum. For example, the crypt-specific marker genes *Lgr5* and *Mki67* and the villus tip marker gene *Ada* were differentially expressed at the beginning and end of our pseudotime, respectively (Supplementary Fig. 6c), a finding confirmed by spatial transcriptomics (Supplementary Fig. 9e). We also observed that many solute transporter genes had distinct patterns of expression along our

pseudotime axis, suggesting localized expression in crypts or along the villus (Fig. 3g). Exploring these genes in situ confirmed the crypt localization of *Slc12a2*, as well as the mid-villus and tip localization of *Slc26a3* and *Slc5a12*, respectively (Supplementary Fig. 9f). Finally, our scRNA-seq data highlighted *Slc40a1* as a solute transporter modulated during infection (Fig. 3h), which was confirmed in our spatial transcriptomic data from *Yersinia*-infected mice (Supplementary Fig. 9g). Taken together, these results provide robust validation of our scRNA-seq pseudotime results in spatial transcriptomes.

To understand the spatial relationships between Yps enterocytes and PGs, we identified putative PGs histologically (Supplementary Fig. 10a, boxes) and computationally ‘unrolled’ equal portions of small intestinal Swiss rolls to create a proximal-to-distal placement of cells along a linearized ileum (Supplementary Fig. 10b). This allowed us to reduce the spatial localization of genes to a simple linear distance across a length of the small intestine that included two distinct PGs (Fig. 6b). *Il1a* and *Tnf*—two genes known to be expressed in PGs—were enriched in proximity to the PGs compared to naive controls (Fig. 6b). Applying cell annotations from GutPath to our spatial data, we observed peaks of myeloid cell enrichment at PG locations (Fig. 6b). We reasoned that if the Yps enterocyte phenotype identified in our data was driven by bacterial signals, then the marker genes defining these enterocytes might localize to PGs where the bacterial burden is highest. Cross-referencing our 132-gene Yps enterocyte signature (Fig. 5) with the 5,000 gene panel used in our spatial transcriptomic assay identified 53 genes, which we then used to test for spatially localized enrichment of the Yps enterocyte signature. We observed statistically significant peaks of Yps enterocyte-associated expression precisely overlapping the histologically confirmed PGs (Fig. 6c).

To identify cell–cell signaling events that might mediate the activation of enterocytes in close proximity to PGs, we focused on a single PG and the adjacent tissue (Fig. 6b, purple bar) as well as a non-PG region (Fig. 6b, yellow bar) and then predicted cell–cell communications separately within these two regions. This analysis identified ligands enriched in the tissue closest to PGs (Fig. 6d), including the IL-1 family member IL-18 and its receptors. Closer inspection showed that *Il18* expression was greatly reduced within the PG compared to the adjacent tissue (Fig. 6e, top), whereas *Il18rap*, which encodes an IL-18 receptor component, showed peak expression in the region corresponding to the PG. These findings suggest that enterocyte-derived IL-18 may signal to inflammatory cells residing within the PG. Cell–cell communication analysis predicted early enterocytes to be the main producers of IL-18, which can signal to NK and NKT cells, as well as ILC1s and ILC3s (Fig. 6f). This result, taken together with our observation of a STAT3-driven program in Yps enterocytes (Fig. 5g), prompted us to ask whether the expression of IL-22—a STAT3-activating cytokine with pleiotropic effects, such as inducing IL-18 production in enterocytes⁶⁸—is increased at PGs. Indeed, we found that *Il22* expression peaked at the PG site

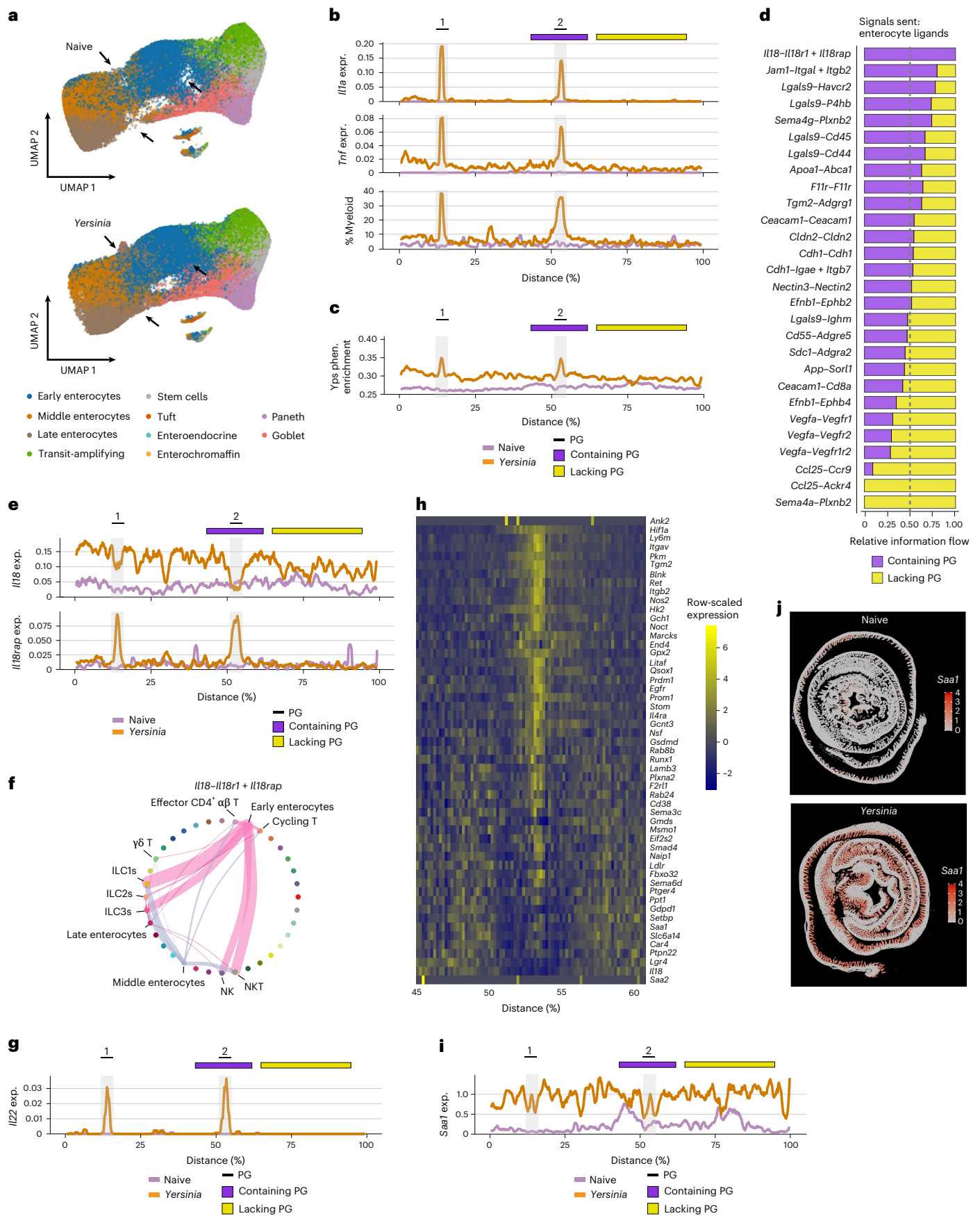
(Fig. 6g), thereby implicating an IL-22/STAT3/IL-18 signaling axis in the response of enterocytes at the site of *Yersinia* infection in the gut.

Although Yps enterocyte-associated gene expression peaked within PGs, it remained elevated relative to the naive control in tissue distant from the PGs (Fig. 6c), suggesting that the expression of individual enterocyte genes may vary spatially during infection. To examine this in more detail, we focused on the region spanning the second PG and including the adjacent tissue (Fig. 6b, purple bar) and visualized the expression of each of the 53 Yps enterocyte genes as a heatmap across that region (Fig. 6h). Our analysis showed that most of these genes—including *Hif1a*, *Nos2*, and the integrins *Itgav* and *Itgb2*—had highly restricted expression in enterocytes located immediately above the PG (Fig. 6h). However, *Saa1* expression remained elevated relative to the naive control across the entire region we examined (Fig. 6i) and was induced in enterocytes across the entire ileum during infection (Fig. 6j). Taking a broader approach to visualize spatially regulated patterns of gene transcription during *Yersinia* infection, we recalculated differential gene expression within all enterocytes from *Yersinia*-infected mice compared to the naive control using only Xenium data. This analysis identified 350 genes upregulated in enterocytes from *Yersinia*-infected mice, and we examined the expression patterns of these genes in the region surrounding a PG (Supplementary Fig. 10c). Some previously unidentified genes, including *Mboat1* and *Slc16a3*, were specifically enriched in enterocytes at the PG (Supplementary Fig. 10d). Other genes, such as *Stat1* and *Stat3*, were broadly elevated in enterocytes during *Yersinia* infection but peaked in PG regions (Supplementary Fig. 10e). Finally, this analysis revealed many genes—including *Slc40a1*, which encodes ferroportin—that were enriched in all enterocytes from *Yersinia*-infected mice, except those surrounding the PG (Supplementary Fig. 10f). Thus, there are overlapping and distinct patterns of gene expression induced by this infection that may influence local immune cell function. These data identify a *Y. pseudotuberculosis*-associated enterocyte phenotype that is spatially linked to the formation of PGs, with effects extending far beyond the sites of infection.

Given the proximity of Yps enterocytes to PGs, and therefore to foci of *Yersinia*, these cells could be influenced either directly by bacterial sensing or indirectly by cytokines and other secreted factors produced by immune cells during infection. *Y. pseudotuberculosis* is primarily extracellular, so we hypothesized that the Yps enterocyte phenotype does not result from direct infection of enterocytes by *Yersinia*. Indeed, using Stereo-seq (STOmics) spatial transcriptomic analysis, which profiles host and microbial RNA, we found that *Yersinia* transcripts were localized to the basal region of the PG and did not extend to the enterocytes overlaying the PG (Supplementary Fig. 10g). Moreover, *Nos2* expression, which might be induced in enterocytes, including those surrounding the PG, did not overlap with *Yersinia* transcripts (Supplementary Fig. 10g). In contrast, our finding that *Il22* expression

Fig. 6 | *Yersinia* PGs reshape enterocyte transcription locally and regionally in the small intestine. **a**, UMAP dimensionality reduction of the Xenium data of subclustered epithelial cells from naive (top) and *Yersinia*-infected (bottom) mice. Arrows indicate discrepancies between naive and *Yersinia* reductions, and cell types are indicated by color. **b**, *Il1a* and *Tnf* rolling mean expression and myeloid cell frequency. **c**, Rolling mean UCell enrichment of the Yps enterocyte signature along a bioinformatically informed distance in *Yersinia*-infected and naive mice. Distances of 0 and 100 indicate the proximal and distal ends, respectively, with the distal end adjacent to the colon. The relative positions of two PGs in the tissue section are shown by shaded areas and are numbered (top edge). Areas containing or lacking a PG are indicated by a purple or yellow bar, respectively (top edge), and were used for the analyses shown in **d**, **f** and **h**. PG positions show significantly increased enrichment relative to regions lacking a PG and compared to the naive sample (FDR < 0.0001, Wilcoxon test). Yps phen., Yps enterocyte phenotype. **d**, Cell–cell communication signaling predictions for all signals produced by enterocytes in the area containing (purple) versus lacking

(yellow) a PG during *Yersinia* infection. **e**, *Il18* and *Il18rap* rolling mean expression along a bioinformatically informed distance in *Yersinia*-infected and naive mice. **f**, Predicted IL-18-mediated cell–cell communication between enterocytes in PG-containing tissue and cells expressing IL-18 receptor components. The width of each interaction is proportional to its predicted strength, and the color indicates the cell type from which IL-18 originates. **g**, *Il22* rolling mean expression along a bioinformatically informed distance in *Yersinia*-infected and naive mice. **h**, Expression of Yps enterocyte-associated genes in epithelial cells along the proximal-to-distal distance, spanning a PG within the *Yersinia* sample. The bioinformatically unrolled distance shown corresponds to the selected area of study (purple bar in **b**). **i**, *Saa1* rolling mean expression along a bioinformatically informed distance in *Yersinia*-infected and naive mice. **j**, *Saa1* expression (indicated by color hue), determined by Xenium spatial analysis, overlaid on intact ileal sections from naive (top) and *Yersinia*-infected (bottom) mice. *n* = 2 replicate mice for each condition (naive, *Yersinia* infection). Representative data are shown from one animal.



is highly restricted to the PG (Fig. 6g), and that genes containing STAT3 motifs in their promoters (Fig. 5g) show elevated expression in enterocytes during *Yersinia* infection, suggests that immune cell recruitment and cytokine secretion have dominant roles in shaping the enterocyte phenotype near a PG. Consistent with this notion, *Stat1* expression in enterocytes is elevated at the PG, and *Stat3* expression is broadly upregulated during *Yersinia* infection (Supplementary Fig. 10e). As activation of these transcription factors is triggered by phosphorylation, we performed immunohistochemical staining for phospho-STAT1 and phospho-STAT3 on *Yersinia*-infected and naive ileal sections. Consistent with our single-cell data, we observed STAT1 activation during *Yersinia* infection, particularly at PGs and including epithelial cells, but not in naive samples (Supplementary Fig. 10h). In contrast, STAT3 was activated more broadly during *Yersinia* infection across the tissue (Supplementary Fig. 10i), again mirroring our scRNA-seq data for *Stat3* (Supplementary Fig. 10e). In summary, *Y. pseudotuberculosis* infection markedly alters the transcriptional state of enterocytes. IL-22, likely acting in concert with other signals, may influence enterocytes in a spatially restricted manner. However, other signals are clearly involved and can robustly modulate enterocyte gene expression across large areas of the intestine. How local versus regional expression patterns are governed and their relative importance in generating protective versus pathogenic responses remain important questions for further study.

Discussion

Infections of the small intestine are major causes of malnutrition, growth stunting and death worldwide. Leveraging intestinal pathogens and experimental infections, we created GutPath, a resource that comprises RNA and protein data for over 300,000 single cells (91 cell states) from the ileum and nearly 200,000 cells (49 cell states) from the ileal-draining MLN, across six different infections and naive controls. Responses to fungi, bacteria, parasitic helminths, protozoa and viruses are represented, thus capturing diverse immune states spanning T_H1 , T_H2 - and T_H17 -dominant immune archetypes, with all major cell populations and many rare cell types profiled. The breadth and depth of GutPath allow investigators to ask new questions about conserved patterns of biology in the ileum, to contextualize new and existing mucosal datasets, and to better interpret mucosal immune responses in other infectious or inflammatory diseases.

A major impediment to the efficient analysis of new single-cell datasets is accurate and detailed cell annotation, particularly in highly dynamic and complex tissues such as the gut. Although multiple resources exist to aid in this process^{34,69}, they do not reflect diverse immune cell activation states, particularly those elicited by infection, thus requiring investigators to undertake arduous manual annotation efforts. GutPath addresses this problem by providing multiple layers of cell state annotations, ranging from broad cell lineages to highly specific cell phenotypes, each accompanied by marker profiles. In addition to all raw sequencing data and code being publicly accessible, gene expression in different cell types and within various pathogen model systems contained in GutPath can be directly queried in a web browser through the CELLxGENE⁷⁰ Visualization In Plugin (VIP)⁷¹, available at <https://gutpath.org>. This atlas provides a foundation for leveraging rapidly evolving large language models to explore complex cellular data. Such approaches streamline analysis workflows, substantially improving research efficiency while accelerating the translation of high-resolution datasets into actionable insights.

GutPath captures the wide breadth of cellular heterogeneity induced by acute infection; however, it has several limitations. Although our time points were chosen to coincide with the peak of infection, differences in how each infection is established and progresses, as well as how adaptive immune components are recruited and activated, cannot be fully explained without further sampling along the time course of infection. Although our infectious doses are standard in the field, further work could be done to investigate the effects of

dose on immune responses in the small intestine. This is particularly relevant, as we revealed associations between worm burdens and fatty acid levels during *Nippostrongylus* infection. Some cell types, notably eosinophils and neutrophils, are more fragile and may require methodological adjustments to improve their representation in scRNA-seq datasets. We also note that, although we designed differential testing to limit bias caused by differences in cell number, bias toward epithelial cells in the data may have enhanced statistical power for DEG detection in this population (Fig. 3).

GutPath is a rich resource for understanding enterocyte biology, particularly because our sample preparation yielded the complete developmental ‘lifespan’ of an enterocyte, ranging from crypt-resident stem cells and transit-amplifying cells to mature absorptive enterocytes present along the length and tip of the villus (Fig. 1). This allowed us to use pseudotime-based trajectory analyses to assign transcripts to precise developmental stages of enterocytes, revealing a highly structured pattern of expression for 82 solute transporters from the crypts to the tips of the villi, 29 of which were dysregulated during one or more infections (Fig. 3g,h). Disruption of the intestinal epithelium in patients with Crohn’s disease and during infection can lead to a wide range of malabsorption phenotypes, but the underlying molecular mechanisms are poorly understood. The dysregulation of solute transporters observed in this study constitutes a potential molecular mechanism for pathogen disruption of ileal biology and merits further investigation. In addition to enterocytes, GutPath captured immune cells interspersed within the epithelial layer, as well as in the underlying LP and the draining MLN, thereby enabling researchers to predict cell–cell communication between immune cells or between immune and stromal cells. This is illustrated by our identification of a unique enterocyte phenotype induced by *Yersinia* infection (Fig. 5), which our data suggest may be driven by IL-22/STAT3 signaling to produce IL-18—a cytokine that then signals through the IL-18 receptor accessory protein (IL-18RAP) on ILCs, NK cells and NKT cells residing in the PG.

We also explored how mucosal immune responses are influenced by proximity to tissue pathology and pathogen load (Fig. 6). Our experiments with *Y. pseudotuberculosis* revealed that some enterocyte transcriptional states colocalized precisely with tissue pathology, while others were more regional, found adjacent to PGs or even distributed across the entire tissue. Similarly, our experiments with *Nippostrongylus*, which establishes infection in the proximal small intestine, showed marked metabolic reprogramming of the distal small intestinal epithelium (Fig. 4). Taken together, these results reveal a ripple effect of gene expression that can emanate from areas where pathogen load and pathology are highest, suggesting that even highly focal or well-contained infections may still influence enterocyte behavior far from the site of insult. As our experiments only covered a single, early time point for each infection, it is unclear to what extent these transcriptional programs are sustained and how they affect the development of effector or memory T cell responses. Nevertheless, this finding has important implications not only for gastrointestinal infections but also for inflammatory diseases such as Crohn’s disease, in which the development of intestinal lesions is a key component of disease pathogenesis. Our spatial data provide a detailed view of the genes most closely linked to tissue pathology and their relative localization across the tissue. These genes should be investigated as therapeutic targets, as well as drivers of tissue remodeling and bacterial control.

Although immune cell infiltrates and proinflammatory gene expression were concentrated around PGs in *Yersinia*-infected mice, we observed that *Saa1*, *Stat3*, *Slc40a1* and many other genes were widely expressed in enterocytes across the entire distal small intestine. SAA1 is known to be induced by IL-1 family members and TNF, and it has direct antimicrobial properties^{72,73} as well as promotes IEC migration during wound healing⁷⁴. However, SAA1 can also contribute to proinflammatory responses through the generation of pathogenic T_H17 cells⁷⁵ and has been linked to sites of active inflammation in patients with Crohn’s

disease⁷⁶. Most recently, *Saa1* and *Saa2* were shown to define a unique tooth-adjacent epithelium where cells are in constant contact with oral microorganisms⁷⁷. It is possible that the highly focal nature of proinflammatory cytokine induction around PGs is sufficient to elicit tissue-wide expression of downstream targets. Conversely, there may be a mechanism by which enterocytes propagate inflammatory signals over long distances. Delineating how inflammatory signaling events unfold over time and space to modulate gene expression in enterocytes can have important implications for identifying new therapeutic opportunities for gastrointestinal diseases.

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/s41590-026-02665-6>.

References

- Culver, R. N. et al. Improved mouse models of the small intestine microbiota using region-specific sampling from humans. Preprint at *bioRxiv* <https://doi.org/10.1101/2024.04.24.590999> (2024).
- Shalon, D. et al. Profiling the human intestinal environment under physiological conditions. *Nature* **617**, 581–591 (2023).
- Shealy, N. G., Baltagulov, M. & Byndloss, M. X. A long journey to the colon: the role of the small intestine microbiota in intestinal disease. *Mol. Microbiol.* **122**, 304–312 (2024).
- Haber, A. L. et al. A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333–339 (2017).
- Mowat, A. M. & Agace, W. W. Regional specialization within the intestinal immune system. *Nat. Rev. Immunol.* **14**, 667–685 (2014).
- Sun, H., Tan, J., Chen, H., Wu, N. & Su, B. Immune niches orchestrated by intestinal mesenchymal stromal cells lining the crypt-villus. *Front. Immunol.* **13**, 1057932 (2022).
- Wang, Y. et al. Single-cell transcriptome analysis reveals differential nutrient absorption functions in human intestine. *J. Exp. Med.* **217**, e20191130 (2020).
- Guéant, J. L., Guéant-Rodriguez, R. M. & Alpers, D. H. Vitamin B12 absorption and malabsorption. *Vitam. Horm.* **119**, 241–274 (2022).
- Hickey, J. W. et al. Organization of the human intestine at single-cell resolution. *Nature* **619**, 572–584 (2023).
- Moor, A. E. et al. Spatial reconstruction of single enterocytes uncovers broad zonation along the intestinal villus axis. *Cell* **175**, 1156–1167 (2018).
- Richard, N. et al. Crohn's disease: why the ileum? *World J. Gastroenterol.* **29**, 3222–3240 (2023).
- Vitek, L. Bile acid malabsorption in inflammatory bowel disease. *Inflamm. Bowel Dis.* **21**, 476–483 (2015).
- Ward, M. G. et al. Prevalence and risk factors for functional vitamin B12 deficiency in patients with Crohn's disease. *Inflamm. Bowel Dis.* **21**, 2839–2847 (2015).
- Khalil, I. A. et al. Morbidity, mortality, and long-term consequences associated with diarrhoea from *Cryptosporidium* infection in children younger than 5 years: a meta-analysis study. *Lancet Glob. Health* **6**, e758–e768 (2018).
- Li, T. et al. Global burden of enteric infections related foodborne diseases, 1990–2021: findings from the Global Burden of Disease Study 2021. *Sci. One Health* **3**, 100075 (2024).
- Pardy, R. D. et al. Analysis of intestinal epithelial cell responses to *Cryptosporidium* highlights the temporal effects of IFN- γ on parasite restriction. *PLoS Pathog.* **20**, e1011820 (2024).
- Cortez, V. et al. Astrovirus infects actively secreting goblet cells and alters the gut mucus barrier. *Nat. Commun.* **11**, 2097 (2020).
- Bossel Ben-Moshe, N. et al. *Salmonella* Typhi gut invasion drives hypoxic immune subsets associated with disease outcomes. *Nat. Commun.* **16**, 6755 (2025).
- Širvinskas, D. et al. Single-cell atlas of the aging mouse colon. *iScience* **25**, 104202 (2022).
- Oliver, A. J. et al. Single-cell integration reveals metaplasia in inflammatory gut diseases. *Nature* **635**, 699–707 (2024).
- Ofer, R. et al. A spatial and temporal transcriptomic atlas of mouse intestinal regeneration. Preprint at *bioRxiv* <https://doi.org/10.1101/2025.08.06.668727> (2025).
- Gu, W., Eke, C., Gonzalez Santiago, E., Olaloye, O. & Konnikova, L. Single-cell atlas of the small intestine throughout the human lifespan demonstrates unique features of fetal immune cells. *Mucosal Immunol.* **17**, 599–617 (2024).
- Karakasheva, T. A. et al. An epigenetic basis for sustained inflammatory epithelial progenitor cell states in Crohn's disease. *Cell. Mol. Gastroenterol. Hepatol.* **20**, 101665 (2026).
- Smith, C. A. et al. Single-cell transcriptomic atlas of enteroendocrine cells along the murine gastrointestinal tract. *PLoS ONE* **19**, e0308942 (2024).
- Beumer, J. et al. High-resolution mRNA and secretome atlas of human enteroendocrine cells. *Cell* **181**, 1291–1306 (2020).
- Gurtner, A. et al. Active eosinophils regulate host defence and immune responses in colitis. *Nature* **615**, 151–157 (2023).
- Garrido-Trigo, A. et al. Macrophage and neutrophil heterogeneity at single-cell spatial resolution in human inflammatory bowel disease. *Nat. Commun.* **14**, 4506 (2023).
- Boland, B. S. et al. Heterogeneity and clonal relationships of adaptive immune cells in ulcerative colitis revealed by single-cell analyses. *Sci. Immunol.* **5**, eabb4432 (2020).
- Wang, Q. et al. Landscape of cell heterogeneity and evolutionary trajectory in ulcerative colitis-associated colon cancer revealed by single-cell RNA sequencing. *Chin. J. Cancer Res.* **33**, 271–288 (2021).
- James, K. R. et al. Distinct microbial and immune niches of the human colon. *Nat. Immunol.* **21**, 343–353 (2020).
- Lee, H.-O. et al. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. *Nat. Genet.* **52**, 594–603 (2020).
- Yang, C. & Merlin, D. Unveiling colitis: a journey through the dextran sodium sulfate-induced model. *Inflamm. Bowel Dis.* **30**, 844–853 (2024).
- Kiesler, P., Fuss, I. J. & Strober, W. Experimental models of inflammatory bowel diseases. *Cell. Mol. Gastroenterol. Hepatol.* **1**, 154–170 (2015).
- Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).
- Croll, N. A. The location of parasites within their hosts: the influence of host feeding and diet on the dispersion of adults of *Nippostrongylus brasiliensis* in the intestine of the rat. *Int. J. Parasitol.* **6**, 441–448 (1976).
- Li, T. et al. An overview of computational methods in single-cell transcriptomic cell type annotation. *Brief. Bioinform.* **26**, bbaf207 (2025).
- McCarthy, N. et al. Distinct mesenchymal cell populations generate the essential intestinal BMP signaling gradient. *Cell Stem Cell* **26**, 391–402 (2020).
- Greicius, G. et al. Telocytes deliver essential Wnts directly to murine intestinal stem cells via synapse-like contacts. *Dev. Cell* **60**, 3102–3115 (2025).
- Klein, S. et al. Interstitial cells of Cajal integrate excitatory and inhibitory neurotransmission with intestinal slow-wave activity. *Nat. Commun.* **4**, 1630 (2013).
- Gross, S. et al. The novel enterochromaffin marker Lmx1a regulates serotonin biosynthesis in enteroendocrine cell lineages downstream of Nkx2.2. *Development* **143**, 2616–2628 (2016).

41. Hao, Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
42. Liu, A. et al. Discovery of optimal cell type classification marker genes from single cell RNA sequencing data. *BMC Methods* **1**, 15 (2024).
43. von Moltke, J., Ji, M., Liang, H.-E. & Locksley, R. M. Tuft-cell-derived IL-25 regulates an intestinal ILC2–epithelial response circuit. *Nature* **529**, 221–225 (2016).
44. Sorobetea, D. et al. Inflammatory monocytes promote granuloma control of *Yersinia* infection. *Nat. Microbiol.* **8**, 666–678 (2023).
45. Gullicksrud, J. A. et al. Enterocyte–innate lymphoid cell crosstalk drives early IFN- γ -mediated control of *Cryptosporidium*. *Mucosal Immunol.* **15**, 362–372 (2022).
46. Peterson, S. T. et al. TNF signaling maintains local restriction of bacterial founder populations in intestinal and systemic sites during oral *Yersinia* infection. *mBio* **16**, e0177925 (2025).
47. Nice, T. J., Robinson, B. A. & Van Winkle, J. A. The role of interferon in persistent viral infection: insights from murine norovirus. *Trends Microbiol.* **26**, 510–524 (2018).
48. Nice, T. J. et al. Interferon- λ cures persistent murine norovirus infection in the absence of adaptive immunity. *Science* **347**, 269–273 (2015).
49. Savage, H. P. et al. Epithelial hypoxia maintains colonization resistance against *Candida albicans*. *Cell Host Microbe* **32**, 1103–1113 (2024).
50. Wang, X. et al. *Candida albicans* accelerates atherosclerosis by activating intestinal hypoxia-inducible factor2 α signaling. *Cell Host Microbe* **32**, 964–979 (2024).
51. Wang, J., Zhao, X. & Wan, Y. Y. Intricacies of TGF- β signaling in Treg and Th17 cell biology. *Cell. Mol. Immunol.* **20**, 1002–1022 (2023).
52. Wang, F. et al. Mannan-binding lectin regulates the Th17/Treg axis through JAK/STAT and TGF- β /SMAD signaling against *Candida albicans* infection. *J. Inflamm. Res.* **15**, 1797–1810 (2022).
53. Flannigan, K. L. & Denning, T. L. Segmented filamentous bacteria-induced immune responses: a balancing act between host protection and autoimmunity. *Immunology* **154**, 537–546 (2018).
54. Finak, G. et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).
55. Van den Berge, K. et al. Trajectory-based differential expression analysis for single-cell sequencing data. *Nat. Commun.* **11**, 1201 (2020).
56. Mayassi, T. et al. Spatially restricted immune and microbiota-driven adaptation of the gut. *Nature* **636**, 447–456 (2024).
57. Lin, L., Yee, S. W., Kim, R. B. & Giacomini, K. M. SLC transporters as therapeutic targets: emerging opportunities. *Nat. Rev. Drug Discov.* **14**, 543–560 (2015).
58. Singh, K. et al. The L-arginine transporter solute carrier family 7 member 2 mediates the immunopathogenesis of attaching and effacing bacteria. *PLoS Pathog.* **12**, e1005984 (2016).
59. Armingol, E. et al. Atlas-scale metabolic activities inferred from single-cell and spatial transcriptomics. Preprint at *bioRxiv* <https://doi.org/10.1101/2025.05.09.653038> (2025).
60. McKenzie, G. J., Bancroft, A., Grecis, R. K. & McKenzie, A. N. J. A distinct role for interleukin-13 in Th2-cell-mediated immune responses. *Curr. Biol.* **8**, 339–342 (1998).
61. Bozadjieva-Kramer, N. et al. Intestinal FGF15 regulates bile acid and cholesterol metabolism but not glucose and energy balance. *JCI Insight* **9**, e174164 (2024).
62. Xi, R. et al. Up-regulation of gasdermin C in mouse small intestine is associated with lytic cell death in enterocytes in worm-induced type 2 immunity. *Proc. Natl Acad. Sci. USA* **118**, e2026307118 (2021).
63. Kong, B. et al. Fibroblast growth factor 15-dependent and bile acid-independent promotion of liver regeneration in mice. *Hepatology* **68**, 1961–1976 (2018).
64. Roux de Bézieux, H., Van den Berge, K., Street, K. & Dudoit, S. Trajectory inference across multiple conditions with condiments. *Nat. Commun.* **15**, 833 (2024).
65. Ivanov, I. I. et al. Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell* **139**, 485–498 (2009).
66. Matsuda, R. et al. A TNF–IL-1 circuit controls *Yersinia* within intestinal pyogranulomas. *J. Exp. Med.* **221**, e20230679 (2024).
67. Jones, D. C. et al. Cell simulation as cell segmentation. *Nat. Methods* **22**, 1331–1342 (2025).
68. Chiang, H.-Y. et al. IL-22 initiates an IL-18-dependent epithelial response circuit to enforce intestinal host defence. *Nat. Commun.* **13**, 874 (2022).
69. Aran, D. et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).
70. Abdulla, S. et al. CZ CELLxGENE Discover: a single-cell data platform for scalable exploration, analysis and modeling of aggregated data. *Nucleic Acids Res.* **53**, D886–D900 (2025).
71. Li, K. et al. Cellxgene VIP unleashes full power of interactive visualization and integrative analysis of scRNA-seq, spatial transcriptomics, and multiome data. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.08.28.270652> (2022).
72. Hari-Dass, R., Shah, C., Meyer, D. J. & Raynes, J. G. Serum amyloid A protein binds to outer membrane protein A of gram-negative bacteria. *J. Biol. Chem.* **280**, 18562–18567 (2005).
73. Shah, C., Hari-Dass, R. & Raynes, J. G. Serum amyloid A is an innate immune opsonin for Gram-negative bacteria. *Blood* **108**, 1751–1757 (2006).
74. Hinrichs, B. H. et al. Serum amyloid A1 is an epithelial prorestitutive factor. *Am. J. Pathol.* **188**, 937–949 (2018).
75. Lee, J.-Y. et al. Serum amyloid A proteins induce pathogenic Th17 cells and promote inflammatory disease. *Cell* **180**, 79–91 (2020).
76. Tang, M. S. et al. Integrated analysis of biopsies from inflammatory bowel disease patients identifies SAA1 as a link between mucosal microbes with T_H17 and T_H22 cells. *Inflamm. Bowel Dis.* **23**, 1544–1554 (2017).
77. Theofilou, V. I. et al. Distinct spatial organization governs oral mucosal immunity. *Nat. Immunol.* **27**, 624–635 (2026).

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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

© The Author(s) 2026

¹Department of Pathobiology, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA, USA. ²Department of Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ³Division of Gastroenterology and Hepatology, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ⁴Department of Microbiology and Immunology, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, USA. ⁵Jill Roberts Institute for Research in Inflammatory Bowel Disease, Weill Cornell Medicine, Cornell University, New York, NY, USA. ⁶Friedman Center for Nutrition and Inflammation, Weill Cornell Medicine, Cornell University, New York, NY, USA. ⁷Allen Discovery Center for Neuroimmune Interactions, New York, NY, USA. ⁸Division of Gastroenterology and Hepatology, Joan and Sanford I. Weill Department of Medicine at Weill Cornell Medicine, Cornell University, New York, NY, USA. ⁹Immunology and Microbial Pathogenesis Program, Weill Cornell Medicine, Cornell University, New York, NY, USA. ¹⁰Department of Microbiology and Immunology, Weill Cornell Medicine, Cornell University, New York, NY, USA. ¹¹Parker Institute for Cancer Immunotherapy, Weill Cornell Medicine, Cornell University, New York, NY, USA. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: beiting@upenn.edu

the MIST Consortium

Christopher A. Hunter¹, Boris Striepen¹, Iwaylo I. Ivanov⁴ & David Artis^{5,6,7,8,9,10,11}

Methods

Mouse care and use

All animal procedures and mouse husbandry were performed in accordance with the guidelines and approval of the University of Pennsylvania Institutional Animal Care and Use Committee (protocol nos. 805432, 804523 and 807356). Female specific-pathogen-free C57BL/6 mice aged 4–7 weeks were purchased from The Jackson Laboratory and housed at the University of Pennsylvania Animal Care Facilities for 1–2 weeks before the study. All mice were free of SFB colonization before the experiments.

Infections

Cryptosporidium parvum. Mouse-adapted *C. parvum* expressing nanoluciferase was propagated as described previously in *lfn^g^{-/-}* mice⁷⁸. Oocysts were purified from feces of the infected mice using sucrose flotation and cesium chloride gradient centrifugation. Mice aged 6–7 weeks were infected with 5×10^4 oocysts by oral gavage. Fecal shedding of oocysts was measured by luminescence using a GloMax reader (Promega). Mice were killed at 4 dpi for tissue collection. Tissues from three replicate mice were analyzed.

Nippostrongylus brasiliensis. L3 larvae of *N. brasiliensis* were purified and washed in PBS. Mice were infected subcutaneously with 200 L3 larvae and killed at 5 dpi. Additional mice were used to measure the worm burden at 6 dpi. Tissues from two replicate mice were used for CITE-seq analysis, including ileum- and duodenum-derived IEC and LP cell preparations. Four replicate mice were used for metabolomic experiments.

Yersinia pseudotuberculosis. *Y. pseudotuberculosis* (clinical isolate strain 32777, serogroup O1) was cultured at 28 °C for 16 h in 2× YT broth supplemented with $2 \mu\text{g ml}^{-1}$ triclosan (Millipore Sigma). Mice were fasted for 16 h and administered 2×10^8 colony-forming units (CFU) of *Yersinia* in 200 μl PBS, as previously described^{44,66}. Mice were killed at 5 dpi, and tissues were collected. Spleens from infected mice were collected in sterile PBS, weighed and homogenized for 40 s with 6.35-mm ceramic beads (MP Biomedicals) using a FastPrep-24 bead beater (MP Biomedicals). Tissue homogenates were serially diluted tenfold in PBS, plated in triplicate on LB agar supplemented with $2 \mu\text{g ml}^{-1}$ triclosan and incubated for 2 days at room temperature. Countable colonies at the lowest dilution step were used to calculate the CFU per gram of tissue. Tissues from three replicate mice were analyzed.

Murine norovirus. Molecular clones of the MNV strain CR6 were generated as previously described to produce working stocks of infectious virus⁷⁹. Plasmids encoding the viral genome were transfected into 293T cells (ATCC) to generate a P0 stock, which was subsequently amplified through two passages in BV2 cells at a multiplicity of infection of 0.05 to produce a P2 stock. For viral stock preparation, infected BV2 cells were freeze-thawed, and the cell lysates and supernatants were centrifuged at 1,200g for 5 min. Supernatants were filtered through a 0.22- μm filter and concentrated using Amicon Ultra centrifugal filters with a 100 kDa molecular weight cutoff. Viral titers were determined by plaque assay using BV2 cells overlaid with DMEM (Corning) containing 1% methylcellulose (Sigma-Aldrich). Plaques were evaluated at 3 dpi using crystal violet staining. Mice aged 7 weeks were infected orally with 3×10^6 plaque-forming units of the virus and killed at 7 dpi for tissue collection. Tissues from two replicate mice were analyzed.

Candida albicans. A single colony of *C. albicans* CHN1 was cultured in 5 ml Sabouraud dextrose medium at 30 °C for 16 h. Cells were pelleted by centrifugation at 1,000g for 5 min and resuspended in 2 ml PBS. At 6–7 weeks of age, mice were infected by oral gavage with 5×10^7 CFU of the fungus and were killed at 7 dpi for tissue collection. Tissues from two replicate mice were analyzed, but IECs were usable from only one of the animals.

Segmented filamentous bacterium. Fecal pellets from SFB-colonized mice were collected and stored at –80 °C. SFB-negative mice were colonized by oral gavage with homogenized fecal pellets, as previously described⁸⁰. Mice were killed at 7 dpi, and tissues were collected for analysis. Successful colonization was measured on day 5 after gavage using SFB-specific qPCR. Tissues from two replicate mice were analyzed. Primers for qPCR amplification were purchased from Integrated DNA Technologies and included the following pairs: SFB 16S common forward (5′-GACGCTGAGGCATGAGAGCAT-3′), SFB 16S common reverse (5′-GACGGCACGGATTGTATTCA-3′); universal 16S forward (5′-ACTCTACGGGAGGCAGCAGT-3′), universal 16S reverse (5′-ATTACCGCGGCTGCTGGC-3′); and secondary SFB 16S forward (5′-GTGTCCCTCCAGAACATACATTA-3′), secondary SFB 16S reverse (5′-CCCAATCCACCATTAGAAATACAA-3′).

Tissue processing

All samples were handled on ice, and centrifugation was performed using centrifuges prechilled to 4 °C. The ileal-draining MLN was removed and placed in PBS before dissociation. Mechanical and chemical dissociation steps included 20 min of incubation at 37 °C in 700 μl digest buffer (complete RPMI (RPMI + 10% FBS + 0.1% β -mercaptoethanol + 1% nonessential amino acids + 1% sodium pyruvate + 1% penicillin–streptomycin), 0.5 mg ml⁻¹ DNase and 0.167 mg ml⁻¹ Liberase TL). The distal third of the small intestine was isolated, Peyer's patches were excised and discarded, and the tissue was minced. The epithelium was isolated by incubating and shaking the tissue at 37 °C for 20 min in 20 ml of chelating and reducing buffer (1 liter of HBSS, 5% FBS, 5 mM EDTA and 0.154 mg ml⁻¹ DTT). Further shaking and washing were performed to separate and collect the epithelium. The LP was then digested in 20 ml of digest buffer for 30 min at 37 °C with shaking. All samples were filtered through 70- and 40- μm filters. Dead cell removal was performed for all IEC suspensions and for LP samples as needed based on cell viability (Dead Cell Removal Kit, Miltenyi Biotec, cat. no. 130-090-101).

Single-cell sequencing

Single-cell suspensions were processed according to the 10x Genomics and BioLegend protocols to create CITE-seq libraries. In brief, cells were stained with the TotalSeq-B Mouse Universal Cocktail v1.0, which includes oligonucleotide-conjugated antibodies for ~100 surface proteins (BioLegend). Cells and 10x Genomics-barcoded gel beads were then encapsulated using GEM-X 3′ chips and a Chromium Controller device. Standard library amplification and indexing were performed using either 10x Genomics GEM-X single-cell 3′ v3 or v4 chemistry. Paired-end libraries were sequenced on a NextSeq 2000 platform.

Analysis of single-cell data

Data preprocessing and integration. Cell Ranger (v8.0.0) was used with the mouse genome GRCm39 to generate raw and filtered feature-barcode matrices. Matrices were imported into R (v4.5.0) using Seurat (v5.3.0) and subjected to primary filtering to retain cells that met the following criteria: RNA and ADT library sizes within ± 2.5 median absolute deviations from their respective medians; >750 detected genes; and <5% mitochondrial RNA content for the MLN samples and <20% mitochondrial content for the ileum samples. ADT counts were normalized using the denoised and scaled by background (DSB) method (dsb v2.0.0)⁸¹, with isotype control features included for technical correction. Merged and processed Seurat objects were used for all downstream analyses.

RNA and ADT data were independently processed, integrated and visualized. Standard variable feature identification ($n = 3,000$), normalization and scaling were applied. Independent samples were integrated using Harmony with both Louvain and Leiden clustering (Monocle3 v1.3.1). RNA and ADT modalities were incorporated through weighted nearest neighbor analysis for further clustering and

subsequent dimensionality reduction. Doublet removal was performed using scDblFinder (v1.22.0), and regression of cell cycle expression⁸² was performed before the final dimensionality reduction visualization.

Clusters and cells were annotated using automated reference-based methods (SingleR v2.11.1, reference = ImmGenData), together with manual curation. Subsequent annotation relied on label transfer to newly added data and manual refinement. For most major populations, subsetting, reclustering and the generation of new dimensionality reductions were performed to describe and visualize cellular heterogeneity and to enable detailed annotation, as shown in Fig. 1 and Supplementary Fig. 4. Duodenal fractions collected from *Nippostrongylus*-infected mice were included in initial clustering, UMAP dimensionality reduction and cell annotation (Fig. 1), as well as in exploring tuft and goblet cell hyperplasia (Fig. 2a,b), but were excluded from analyses focused on the ileum.

DEG analysis using MAST. Differential gene expression between infection and naive control conditions was calculated for each cell type at an intermediate resolution level where there were more than ten cells per condition. All cell types and states were downsampled to $\leq 5,000$ cells to limit bias from uneven sampling depths between populations. MAST (Seurat v5.3.0) was used with the following latent parameters: genes per cell, UMI per cell, percentage of ribosomal reads and percentage of mitochondrial reads. *P* values were adjusted using the Benjamini–Hochberg method, and an FDR of < 0.05 was used for data exploration.

GO enrichment analysis. Select gene sets identified through differential expression analysis were subjected to GO enrichment analysis (gProfiler v0.2.3). The GOBP collection was used for most analyses, with FDR correction. GO terms associated with fewer than 20 or more than 500 genes were excluded. UCell analysis (escape v2.4.0 and UCell v2.6.2) was conducted to identify significantly enriched gene sets.

Metabolic analysis. To infer metabolic activity states across single cells, we used scCellfie and the included metabolic task database (v0.4.7) (Python v3.10.18, Scanpy v1.11.3). Individual task scores were calculated using the following parameters: smooth cells = true; $\alpha = 0.33$; number of nearest neighbors = 10; and batch indicator = mouse. Normalized cell type-specific metabolic activity was summarized by computing the trimean of task scores within each cell type and visualized in radial plots across infection conditions. Differential analysis of metabolic activity was performed for each cell type relative to the naive samples, with significance determined at FDR < 0.05 , Cohen's *d* > 0.5 and $\log_2(\text{FC}) > 0.585$ (1.5 FC). Significantly altered lipid metabolic pathways identified in enterocytes from *Nippostrongylus*-infected mice were visualized.

Pseudotime and trajectory-based differential expression. To model epithelial differentiation along the crypt–villus axis, intestinal stem cells, transit-amplifying cells and enterocytes were reclustered using new dimensionality reductions. Pseudotime trajectories were inferred using Monocle3, and intestinal stem cells were manually selected as the root node⁸³. Within the naive condition, pseudotime-associated genes were identified using a generalized additive model and trade-Seq (v1.13.12) to test for significant expression changes as a function of pseudotime. Genes with FDR < 0.05 were considered significantly dynamic and were filtered for predefined genes associated with solute transport. The same pseudotime ordering was used to compare expression between naive and infection conditions. Infection-induced changes in pseudotime expression were analyzed using pseudotime scores as a continuous covariate within a linear model framework. Genes exhibiting significant condition-by-pseudotime interactions were classified as differentially regulated along the trajectory. DEGs from enterocytes of *Yersinia*-infected mice were further analyzed using functional enrichment analysis.

Single-cell analysis: transcription factor activity inference. To identify transcription factors regulating the Yps enterocyte transcriptional module, motif enrichment analysis was performed using the RcisTarget package (v1.28.1)⁸⁴. The *Mus musculus* motif ranking database (mm10_refseq-r80_500bp_up_and_100bp_down_tss.mc9nr.genes_vs_motifs.rankings.feather) and the corresponding motif-to-transcription factor annotation table (motifs-v9-nr.mgi-m0.001-o0.0.tbl) were downloaded from the cisTarget resource repository. Motifs with NES > 3.0 were considered significantly enriched and were annotated to candidate transcription factors. Motif–gene relationships of the selected enriched motifs were then visualized.

Cell–cell communication analysis. Cell–cell communication analysis was performed using CellChat (v2.1.2)⁸⁵. Separate CellChat objects were created from Seurat objects. Ligand–receptor interactions were inferred from receptor–ligand pairs contained in the CellChatDB mouse database. Identification of overexpressed genes and interactions, computation of communication probabilities (bootstrap $n = 50$) and aggregation at the signaling pathway level were also performed. Putative interactions were compared across conditions using merged conditions and signaling centrality. RankNet plots (significance threshold $P < 0.05$) were used to visualize infection-specific changes in putative cell–cell interactions.

Correlation between ADT and gene expression. To assess RNA–protein coherence, RNA and ADT expression data from the ileum and MLN were pseudobulked by cell type and mouse to mitigate single-cell sparsity. Pearson correlation coefficients were then calculated for each predefined protein–transcript pair to quantify the relationship between transcript and protein abundance.

Spatial transcriptomics

Spatially resolved transcriptomic profiling was performed using the 10x Genomics Xenium platform and the predesigned Xenium Prime 5K Mouse Gene Expression Panel. Infected and naive mice were killed, and the small intestine was removed and flushed before isolating and Swiss rolling 5 cm of the terminal ileum. Tissues were fixed in 10% neutral buffered formalin for 18–24 h before paraffin embedding. Sectioning and Xenium imaging were performed in partnership with the University of Pennsylvania Dental School Spatial Transcriptomics Core. Sectioning and sequencing of Stereo-seq spatial transcriptomic data were carried out in collaboration with Complete Genomics.

Analysis of Xenium spatial transcriptomic data. Cell segmentation of Xenium data was performed using Proseg (v2.0.3), a probabilistic cell segmentation framework, and Xenium Ranger (v3.11). For each sample, raw Xenium outputs generated by Xenium Ranger were imported. Cell-level quality metrics, including total transcript counts and detected features, were used for filtering. Cells with values outside ± 2.5 median absolute deviations from the median for either metric were excluded. The remaining high-quality cells were retained for downstream SCTransform normalization, dimensionality reduction and clustering.

To linearize the spatial organization of the intestinal Swiss roll, Xenium spatial coordinates were exported and processed using Intest-Line to manually define points along the outer and inner boundaries of the rolled tissue⁸⁶. These boundary coordinates were used to construct a closed polygon using the sf (1.0-21) package in R. Cells located within the polygon were subsetting from the data, and a continuous spatial trajectory (centerline) through the tissue was calculated. Cumulative Euclidean distances were then computed along this line to establish a normalized positional coordinate representing the relative proximal-to-distal location of each cell along the unrolled axis. Each cell within the polygon was assigned a value corresponding to the nearest point on the interpolated centerline. This derived spatial

variable was added to the Seurat object and used for downstream visualization and analysis, effectively ‘unrolling’ the Swiss roll into a one-dimensional axis.

Analysis of STOmics Stereo-seq spatial transcriptomic data. Raw sequencing data were processed using the Stereo-seq Analysis Workflow (SAW, v8.2.2). SAW aligns reads to the mouse genome (GRCm39 release 115) using STAR. Bowtie2 and Kraken2 (PlusPF 2025 database) were used for host and microbial sequence characterization, respectively. Raw outputs were first converted to single-cell resolution using the ‘bin2cell’ function of the SAW pipeline to integrate gene expression files with single-stranded DNA mask images. Spatial data were then converted to a Seurat-compatible format using addImag2annData (v0.0.1). Secondary analysis was performed in R (v4.5.1) using the Seurat framework. Count data were collapsed to gene symbols using org.Mm.eg.db (v3.22.0). Quality control was performed by filtering cells based on transcript sums (150–2,000), unique feature counts (40–450) and mitochondrial content (<8%). Data were normalized and batch-corrected using SCTransform. Automated cell annotation was performed using the GutPath data through robust cell type decomposition with the spacexr (v2.2.1) package. For microbial localization, *Yersinia* genus-level reads were extracted from the SAW output, and spatial coordinates were aligned with the existing Seurat image coordinates for visualization. Final visualizations, including spatial dimensionality reduction plots and microbial overlays, were generated using ggplot2 (v4.0.1).

Use of artificial intelligence. Artificial intelligence (Gemini 3.5 Flash and NVIDIA Nemotron 3 Super) was used as a companion tool during code and script development. It was not used generatively to perform analysis. All code suggestions generated by artificial intelligence models were independently verified and executed by the authors.

Immunohistochemical staining for phosphorylated STAT1 and STAT3

Formalin-fixed, paraffin-embedded sections of the small intestine were prepared and stained by the Comparative Pathology Core at the University of Pennsylvania School of Veterinary Medicine. For immunohistochemistry (IHC) and in situ hybridization (ISH), 5-µm-thick paraffin sections were mounted on ProbeOn slides (Thermo Fisher Scientific, cat. no. 15-188-51). Chromogenic IHC was performed as described elsewhere⁸⁷ using a Leica BOND RXm automated platform combined with the Bond Polymer Refine Detection Kit (Leica, cat. no. DS9800). Briefly, after dewaxing and rehydration, sections were pretreated with the epitope retrieval BOND ER2 high pH buffer (Leica, cat. no. AR9640) for 20 min at 98 °C. Endogenous peroxidase was inactivated with 3% H₂O₂ for 10 min at room temperature. Nonspecific tissue–antibody interactions were blocked by incubating the sections for 30 min at room temperature with Leica PowerVision IHC/ISH Super Blocking Solution (PV6122). The same blocking solution also served as a diluent for the primary antibodies. Primary antibodies (anti-phospho-STAT1 and anti-phospho-STAT3; Cell Signaling Technology, cat. nos. 9167 and 9145, respectively) were applied to the sections and incubated for 45 min at room temperature. A biotin-free, polymeric detection system consisting of horseradish peroxidase-conjugated anti-rabbit IgG (Leica, cat. no. DS9800) was then applied for 25 min at room temperature. Immunoreactivity was revealed using the diaminobenzidine chromogen reaction. Tissue sections were counterstained with hematoxylin, dehydrated through an ethanol series, cleared in xylene and permanently mounted using a resinous mounting medium (Thermo Scientific ClearVue Coverslipper).

Metabolite measurements

To measure the effect of *Nippostrongylus* infection on fatty acids and bile acids, four mice were each infected with 200 L3 larvae. At 5 dpi,

luminal contents (including fecal pellets) were collected from the colon and ileum for both assays. Epithelial scrapings of the ileum were collected by gliding forceps along its epithelial surface. Adult worms were counted in the duodenum and jejunum of each mouse to confirm infection and quantify the worm burden.

Fatty acid measurement. Samples were spiked with 10 µg each of uniformly ¹³C-labeled palmitic acid and stearic acid as internal standards. Calibration curves were constructed simultaneously using a mixture of 14 fatty acids, spiked with the same amount of labeled acids and extracted using the same protocol as that used for samples from the small intestine. The highest calibration contained 7 µg of each fatty acid, and serial 1:1 dilutions were performed. Methyl-*tert*-butyl ether extraction and data analysis were performed as described previously^{88,89}.

Bile acid measurement. Bile acids were quantified using a Waters Acquity ultra-performance liquid chromatography (UPLC) system with a Cortecs UPLC C18+ 1.6-µm 2.1 × 50 mm column and a QDa single quadrupole mass detector. Briefly, fecal samples were suspended in methanol (5 µl per mg of stool), vortexed for 1 min and centrifuged twice at 13,000g for 5 min. The supernatant was transferred to a new tube, sealed and stored at 4 °C until analysis. The flow rate was 0.8 ml min⁻¹; the injection volume was 4 µl; the column temperature was 30 °C; the sample temperature was 4 °C; and the run time was 4 min per sample. Eluent A was 0.1% formic acid in water; eluent B was 0.1% formic acid in acetonitrile; the weak needle wash was 0.1% formic acid in water; the strong needle wash was 0.1% formic acid in acetonitrile; and the seal wash was 10% acetonitrile in water. The gradient was as follows: initial flow 70% eluent A; linear gradient to 100% eluent B over 2.5 min; hold at 100% eluent B for 0.6 min; and linear gradient to 70% eluent A over 0.9 min. The mass detection channels were +357.35 for chenodeoxycholic acid and deoxycholic acid; +359.25 for lithocholic acid; –407.5 for cholic, α-muricholic, β-muricholic, γ-muricholic and ω-muricholic acids; –432.5 for glycolithocholic acid; –448.5 for glycochenodeoxycholic and glycodeoxycholic acids; –464.5 for glycocholic acid; –482.5 for tauroolithocholic acid; –498.5 for taurochenodeoxycholic and taurodeoxycholic acids; and –514.4 for taurocholic acid. Samples were quantified against standard curves of at least five points, each run in triplicate (chemicals from Santa Cruz Biotechnology and Steraloids). Standard curves were run at the beginning and end of each metabolomic run. Blanks and standards were run every eight samples.

Reporting summary

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

Data availability

CITE-seq data are available as processed, fully annotated Seurat objects and as downsampled data loaded into web-accessible CELLxGENE instances at <https://gutpath.org>. In addition, all raw FASTQ files for CITE-seq experiments are available in the Sequence Read Archive (SRA) (accession no. [PRJNA1378118](https://www.ncbi.nlm.nih.gov/sra/PRJNA1378118)). 10x Genomics Xenium 5K mouse and Complete Genomics Stereo-seq spatial transcriptomic data are available in the Gene Expression Omnibus (GEO) (accession no. [GSE313636](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE313636) and [GSE335102](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE335102), respectively).

Code availability

All code used in the analysis and visualization of CITE-seq and spatial transcriptomic data is available via GitHub at <https://github.com/hartandrew/GutPath>.

References

- Dumaine, J. E. et al. The enteric pathogen *Cryptosporidium parvum* exports proteins into the cytosol of the infected host cell. *Elife* **10**, e70451 (2021).

79. Strong, D. W., Thackray, L. B., Smith, T. J. & Virgin, H. W. Protruding domain of capsid protein is necessary and sufficient to determine murine norovirus replication and pathogenesis in vivo. *J. Virol.* **86**, 2950–2958 (2012).
80. Farkas, A. M. et al. Induction of Th17 cells by segmented filamentous bacteria in the murine intestine. *J. Immunol. Methods* **421**, 104–111 (2015).
81. Mulè, M. P., Martins, A. J. & Tsang, J. S. Normalizing and denoising protein expression data from droplet-based single cell profiling. *Nat. Commun.* **13**, 2099 (2022).
82. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
83. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
84. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
85. Jin, S. et al. Inference and analysis of cell–cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).
86. Yuzeir, A. et al. IntestLine: a shiny-based application to map the rolled intestinal tissue onto a line. *Bioinformatics* **39**, btad140 (2023).
87. Radaelli, E. et al. Mitochondrial defects caused by PARL deficiency lead to arrested spermatogenesis and ferroptosis. *Elife* **12**, e84710 (2023).
88. Matyash, V., Liebisch, G., Kurzchalia, T. V., Shevchenko, A. & Schwudke, D. Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *J. Lipid Res.* **49**, 1137–1146 (2008).
89. Wang, D., Xu, P. & Mesaros, C. Analytical considerations for reducing the matrix effect for the sphingolipidome quantification in whole blood. *Bioanalysis* **13**, 1037–1049 (2021).

Acknowledgements

The Mucosal Immunology Studies Team (MIST; <https://www.mucosal.org/>) is a long-running research consortium funded by the National Institute of Allergy and Infectious Diseases (NIAID). Since 2011, MIST investigators have focused on defining the complex and broad-ranging immunological interactions that occur at mucosal surfaces. The long-term goals of the consortium are to gain insight into how the mucosal immune system responds to microbial factors at barrier surfaces, maintains homeostasis, and communicates systematically in health and disease. By achieving these goals, we will be better positioned to facilitate the future development of immunotherapies and vaccines that protect mucosal surfaces from infection and immune-mediated pathology or disease. IHC samples were prepared and stained by the University of Pennsylvania Penn Vet Comparative Pathology Core Facility (RRID:SCR_022438), which is partially supported by the Abramson Cancer Center Support Grant (P30 CA016520). The scanner used for whole-slide imaging and the image analysis software were supported by an NIH Shared Instrumentation Grant (S10 OD023465-01A1). The Leica BOND RXm instrument used for the IHC was acquired through the Penn Vet IIZD Core pilot grant opportunity. Stereo-seq OMNI spatial transcriptomic experiments were performed in collaboration with D. Taussig and

Complete Genomics. Fatty acid profiling by mass spectrometry was performed in collaboration with the Biomolecular Mass Spectrometry Core (BMSC) at the University of Pennsylvania, which is partially funded by the National Institute of Environmental Health Sciences (P30 ES013508). Bile acid processing was assisted by the Microbial Culture and Metabolomics Core of the PennCHOP Microbiome Program and the Center for Molecular Studies in Digestive and Liver Diseases (NIH P30DK050306).

Author contributions

Planning and project management: A.H., M.M., C.H., B.E.H., I.S.C., Y.Y., K.C., J.M., H.Y., X.H., D.C., I.I.I., I.E.B., D.A., C.A.H., D.P.B. Infections: M.M., C.H., B.E.H., I.S.C., S.B., Y.Y., H.Y., X.H. Tissue and single-cell suspension collection: A.H., M.M., C.H., B.E.H., I.S.C. 10x CITE-seq sequencing: D.C. Tissue collection for spatially resolved transcriptomics: S.B., B.A.W. Single-cell and spatial transcriptomic analysis: A.H., R.X. Computational resource management: R.X., A.H., D.P.B. Study conceptualization: the MIST Consortium, B.S., D.A., C.A.H., D.P.B. Intellectual contribution: A.H., K.C., J.M., I.I.I., B.S., S.S., I.E.B., D.A., C.A.H., D.P.B. Writing of the manuscript: A.H., C.A.H., D.P.B.

Funding

This research was supported by the National Institutes of Health (DK126871, AI151599, AI095466, AI095608, AR070116, AI172027 and DK132244, all to D.A.; K99AI180354 to H.Y.; AI163671, AI163671-05S1 and AI148249 to B.S. and C.A.H.; DK093668 and DK050306 to K.C.; AI186538 to S.B.; AI191440 to I.E.B. and S.S.); the Allen Discovery Center program, a Paul G. Allen Frontiers Group advised program of the Allen Family Philanthropies; the Crohn's and Colitis Foundation Research Fellowship Award (no. 937437 to H.Y. and no. 1455492 to X.H.); and Cure for IBD, Weill Cornell Medicine Jill Roberts Institute, the Sanders Family, the Rosanne H. Silbermann Foundation, and the Parker Institute for Cancer Immunotherapy at Weill Cornell Medicine (all to D.A.). Y.Y. was supported by a Cancer Research Institute Irvington Postdoctoral Fellowship (CRI5231). S.B. was supported by the Philadelphia Program for Mentored Research Training in Kidney, Urologic, and Hematologic Diseases (TL1DK143326).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41590-026-02665-6>.

Correspondence and requests for materials should be addressed to Daniel P. Beiting.

Peer review information *Nature Immunology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: S. Houston, in collaboration with the *Nature Immunology* team.

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

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 (<i>n</i>) 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 <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☒	☐ 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. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ 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 <i>d</i> , Pearson's <i>r</i>), 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	Data were collected using the following software: Xenium Ranger (v3.11),
Data analysis	All code information is included in the github repo linked (https://github.com/hartandrew/GutPath) to this publication. Data analysis was performed utilizing R and Python with the following software tools: Cell Ranger (v8.0.0) , R (v4.5.0) , Seurat (v5.3.0), (dsb v2.0.0), (Monocle3 v1.3.1), scDblFinder (v1.22.0), SingleR (v2.11.1), gProfiler (v0.2.3), escape (v2.4.0), UCell (v2.6.2), scCellfie (v0.4.7), Python (v3.10.18), scanpy (v1.11.3), tradeSeq (v1.13.12) , RcisTarget (v1.28.1), CellChat (v2.1.2), Proseq (v2.0.3), Xenium Ranger (v3.11), sf (1.0-21) , Stereo-seq Analysis Workflow (SAW, v8.2.2), Bowtie2, Kraken2 (PlusPF 2025 database), addlm2annData (v0.0.1), org.Mm.eg.db (v3.22.0), spacexr (v2.2.1) , ggplot2 (v4.0.1), CellxGene VIP v1.3.0. The entire collective R package environment is provided below for additional context abind (≤1.4-8), admisc (≤0.39), affy (≤1.88.0), affyio (≤1.80.0), akima (≤0.6-3.6), alabaster.base (≤1.10.0), alabaster.matrix (≤1.10.0), alabaster.ranges (≤1.10.0), alabaster.schemas (≤1.10.0), alabaster.se (≤1.10.0), annotate (≤1.88.0), AnnotationDbi (≤1.72.0), AnnotationFilter (≤1.34.0), AnnotationHub (≤4.0.0), ape (≤5.8-1), apeg (≤1.32.0), aplot (≤0.2.9), ArchR (≤1.0.3), aricode (≤1.0.3), arrow (≤21.0.0), askpass (≤1.2.1), assertthat (≤0.2.1), assertthat (≤1.4.0), AUCell (≤1.32.0), Azimuth (≤0.5.0), babelgene (≤22.9), babelwhale (≤1.2.0), backports (≤1.5.0), Banksy (≤1.6.0), base (≤4.5.1), base64enc (≤0.1-3), basilisk (≤1.23.0), basilisk.utils (≤1.23.1), batchelor (≤1.26.0), bayesm (≤3.1-7), bayesplot (≤1.15.0), bbmle (≤1.0.25.1), bbsBayes2 (≤1.1.3.3), bdsmatrix (≤1.3-7), beachmat (≤2.26.0), beeswarm (≤0.4.0), BH (≤1.87.0-1), bigD (≤0.3.1), biglm (≤0.9-3), Biobase (≤2.70.0), BiocBaseUtils (≤1.12.0), BiocFileCache (≤3.0.0), BiocGenerics (≤0.56.0), BiocIO (≤1.20.0), biocmake (≤1.2.0), BiocManager (≤1.30.27), BiocNeighbors (≤2.4.0), BiocParallel (≤1.44.0), BiocSingular (≤1.26.1), BiocVersion (≤3.22.0), biomaRt (≤2.66.0), Biostrings (≤2.78.0), biovizBase (≤1.58.0), bit (≤4.6.0), bit64 (≤4.6.0-1), bitops (≤1.0-9), blob (≤1.2.4), bluster (≤1.20.0), boot (≤1.3-32), BPCells (≤0.3.1), brew (≤1.0-10), bridgesampling (≤1.2-1), brio (≤1.1.5), brms (≤2.23.0), Broddingnag (≤1.2-9), broom (≤1.0.10), BSgenome (≤1.78.0), BSgenome.Hsapiens.UCSC.hg38 (≤1.4.5), BSgenome.Mmusculus.UCSC.mm10 (≤1.4.3), bslib (≤0.9.0), cachem (≤1.1.0), Cairo (≤1.7-0), callr (≤3.8.0), car (≤3.1-3), carData (≤3.0-5), caret (≤7.0-1), carrier (≤0.3.0.4), caTools (≤1.18.3), CellChat (≤2.2.0), cellDex

(≤1.20.0), cellranger (≤1.1.0), checkmate (≤2.3.3), ChIPseeker (≤1.46.1), chromoMap (≤4.1.1), chromote (≤0.5.1), chromVAR (≤1.32.0), chromVARmotifs (≤0.2.0), cigarillo (≤1.0.0), circlize (≤0.4.16), class (≤7.3-23), classInt (≤0.4-11), cli (≤3.6.6), clipr (≤0.8.0), clock (≤0.7.4), Clonotrace (≤1.1.0), clue (≤0.3-66), ClustAssess (≤1.1.0), cluster (≤2.1.8.1), clusterMole (≤1.1.1), clusterProfiler (≤4.18.2), ClusterR (≤1.3.5), cmdstanr (≤0.9.0.9000), coda (≤0.19-4.1), codetools (≤0.2-20), colorspace (≤2.1-2), commonmark (≤2.0.0), compiler (≤4.5.1), ComplexHeatmap (≤2.26.0), ComplexUpset (≤1.3.3), compositions (≤2.0-9), CompQuadForm (≤1.4.4), condiments (≤1.18.0), conflicted (≤1.2.0), coro (≤1.1.0), corpcor (≤1.6.10), corrplot (≤0.95), cowplot (≤1.2.0), cpp11 (≤0.5.2), crayon (≤1.5.3), credentials (≤2.0.3), crossmatch (≤1.3-1), crosstalk (≤1.2.2), cubature (≤2.1.4-1), curl (≤7.1.0), dada2 (≤1.38.0), data.table (≤1.17.8), data.tree (≤1.2.0), datasets (≤4.5.1), DBI (≤1.2.3), dbplyr (≤2.5.1), dbscan (≤1.2.3), DelayedArray (≤0.36.0), DelayedMatrixStats (≤1.32.0), deldir (≤2.0-4), dendextend (≤1.19.1), DEoptimR (≤1.1-4), Deriv (≤4.2.0), desc (≤1.4.3), DESeq2 (≤1.50.2), devtools (≤2.4.6), dffdx (≤0.2-0), diagram (≤1.6.5), dichromat (≤2.0-0.1), diffobj (≤0.3.6), digest (≤0.6.39), dir.expiry (≤1.18.0), DirichletMultinomial (≤1.52.0), distinct (≤1.22.0), distributional (≤0.5.0), doBy (≤4.7.0), docopt (≤0.7.2), doMC (≤1.3.8), doParallel (≤1.0.17), doRNG (≤1.8.6.3), DOSE (≤4.4.0), doSNOW (≤1.0.20), dotCall64 (≤1.2), DoubletFinder (≤2.0.6), downlit (≤0.4.5), dplyr (≤1.1.4), dqrng (≤0.4.1), DropletUtils (≤1.30.0), dsb (≤2.0.1), DT (≤0.34.0), dtplyr (≤1.3.2), duckdb (≤1.4.4), dyndimred (≤1.0.4), dynfeature (≤1.0.1), dynguidelines (≤1.0.1), dynmethods (≤1.0.5.9000), dyno (≤0.1.2), dynparam (≤1.0.2), dynplot (≤1.1.2), dynutils (≤1.0.12), dynwrap (≤1.3.0), e1071 (≤1.7-16), Ecdat (≤0.4.7), Ecume (≤0.9.2), edgeR (≤4.8.0), ellipse (≤0.5.0), ellipsis (≤0.3.2), ellmer (≤0.4.0), emdbook (≤1.3.14), emmeans (≤2.0.0), enrichplot (≤1.30.3), enrichR (≤3.4), EnsDb.Hsapiens.v86 (≤2.99.0), EnsDb.Mmusculus.v79 (≤2.99.0), EnsDb.Rnorvegicus.v79 (≤2.99.0), ensembleDb (≤2.34.0), entropy (≤1.3.2), EpiDISH (≤2.24.0), escape (≤2.6.1), estimability (≤1.5.1), evaluate (≤1.0.5), evd (≤2.3-7.1), evmix (≤2.12), ExperimentHub (≤3.0.0), extrafont (≤0.2), extrafontdb (≤1.1), factoextra (≤1.0.7), FactoMineR (≤2.12), fansi (≤1.0.7), farver (≤2.1.2), FastCAR (≤0.1.0), fastcluster (≤1.3.0), fastDummies (≤1.7.5), fastGQHquad (≤1.0.1), fastmap (≤1.2.0), fastmatch (≤1.1-6), fgsea (≤1.36.0), fields (≤17.1), filelock (≤1.0.3), fitdistrplus (≤1.2-4), flashClust (≤1.01-2), fmsheer (≤0.5.0.9006), FNN (≤1.1.4.1), fontawesome (≤0.5.3), fontBitstreamVera (≤0.1.1), fontLiberation (≤0.1.0), fontquiver (≤0.2.1), forcats (≤1.0.1), foreach (≤1.5.2), foreign (≤0.8-90), formatR (≤1.14), Formula (≤1.2-5), fs (≤1.6.6), furr (≤0.3.1), futile.logger (≤1.4.3), futile.options (≤1.0.1), future (≤1.68.0), future.apply (≤1.20.0), GA (≤3.2.5), gamm4 (≤0.2-7), gargle (≤1.6.0), gdtools (≤0.4.4), genefilter (≤1.92.0), generics (≤0.1.4), GenomelnfoDb (≤1.46.0), GenomelnfoDbData (≤1.2.15), GenomicAlignments (≤1.46.0), GenomicFeatures (≤1.62.0), GenomicRanges (≤1.62.0), geofacet (≤0.2.4), geogrid (≤0.1.2), gert (≤2.2.0), GetoptLong (≤1.1.0), ggalluvial (≤0.12.5), GGally (≤2.4.0), ggbeeswarm (≤0.7.3), ggbiplot (≤0.6.2), ggbreak (≤0.1.6), ggcorplot (≤0.1.4.1), ggendo (≤0.2.0), ggdist (≤3.3.3), ggforce (≤0.5.0), ggformula (≤1.0.0), ggfortify (≤0.4.19), ggfun (≤0.2.0), ggiraph (≤0.9.2), ggnetwork (≤0.5.14), ggnewscale (≤0.5.2), ggplot2 (≤4.0.1), ggplotify (≤0.1.3), ggprism (≤1.0.7), ggpubr (≤0.6.2), ggraph (≤2.2.2), ggrastr (≤1.0.2), ggrepel (≤0.9.6), ggridges (≤0.5.7), ggsci (≤4.1.0), ggsignif (≤0.6.4), ggstats (≤0.13.0), ggtaggle (≤0.0.9), ggtext (≤0.1.2), ggthemes (≤5.2.0), ggtree (≤4.0.1), ggVennDiagram (≤1.5.7), gh (≤1.5.0), gitcreds (≤0.1.2), glmGamPoi (≤1.22.0), GlobalOptions (≤1.1.3), globals (≤0.18.0), glue (≤1.8.0), GMedian (≤1.2.7), gmp (≤0.7-5), GO.db (≤3.22.0), goftest (≤1.2-3), googledrive (≤2.1.2), googlesheets4 (≤1.1.2), GOSemSim (≤2.36.0), gower (≤1.0.2), GPArotation (≤2025.3-1), gplots (≤3.3.0), gprofiler2 (≤0.2.4), graph (≤1.88.0), graphics (≤4.5.1), graphlayouts (≤1.2.2), grDevices (≤4.5.1), grid (≤4.5.1), gridBase (≤0.4-7), gridExtra (≤2.3), gridGraphics (≤0.5-1), gridtext (≤0.1.5), grr (≤0.9.5), GSEABase (≤1.72.0), gsl (≤2.1-9), gson (≤0.1.0), GSVA (≤2.4.2), gt (≤1.1.0), gtable (≤0.3.6), gtools (≤3.9.5), gypsum (≤1.6.0), h5mread (≤1.2.1), hardhat (≤1.4.2), harmony (≤1.2.4), hash (≤2.2.6.3), haven (≤2.5.5), HCATonsilData (≤1.8.0), HDF5Array (≤1.38.0), hdf5r (≤1.3.12), HDInterval (≤0.2.4), here (≤1.0.2), Herper (≤1.20.0), hexbin (≤1.28.5), highr (≤0.11), Hmisc (≤5.2-4), hms (≤1.1.4), homologue (≤1.4.68.19.3.27), hrbrthemes (≤0.9.2), htmlTable (≤2.4.3), htmltools (≤0.5.9), htmlwidgets (≤1.6.4), httpuv (≤1.6.16), httr (≤1.4.7), httr2 (≤1.2.1), hwriter (≤1.3.2.1), ica (≤1.0-3), ids (≤1.0.1), igraph (≤2.2.1), immApex (≤1.4.0), iNEXT (≤3.0.2), ini (≤0.3.1), INLA (≤25.06.13), inlabru (≤2.13.0), INLAspacetime (≤0.1.12.009), INLAtools (≤0.0.5), inline (≤0.3.21), interp (≤1.1-6), ipred (≤0.9-15), IRanges (≤2.44.0), irlba (≤2.3.5.1), isoband (≤0.3.0), iterators (≤1.0.14), janitor (≤2.2.1), JASPAR2020 (≤0.99.10), jpeg (≤0.1-11), jquerylib (≤0.1.4), jsonlite (≤2.0.0), jsonvalidate (≤1.5.0), juicyjuice (≤0.1.0), KEGGREST (≤1.50.0), kernlab (≤0.9-33), KernSmooth (≤2.23-26), keyring (≤1.4.1), knitr (≤1.5), labeling (≤0.4.3), labelled (≤2.16.0), lambda.r (≤1.2.4), later (≤1.4.8), lattice (≤0.22-7), latticeExtra (≤0.6-31), lava (≤1.8.2), lazyeval (≤0.2.2), leaps (≤3.2), leiden (≤0.4.3.1), leidenAlg (≤1.1.5), leidenbase (≤0.1.35), liana (≤0.1.14), lifecycle (≤1.0.5), limma (≤3.66.0), listenv (≤0.10.0), litedown (≤0.8), lmds (≤0.1.0), lme4 (≤1.1-37), lmtest (≤0.9-40), lobstr (≤1.2.1), locfdr (≤1.1-8), locfit (≤1.5-9.12), logger (≤0.4.1), loo (≤2.9.0), LoomExperiment (≤1.28.0), lsa (≤0.73.3), lubridate (≤1.9.4), magick (≤2.9.0), magrittr (≤2.0.5), mapproj (≤1.2.12), maps (≤3.4.3), markdown (≤2), MASS (≤7.3-65), MAST (≤1.36.0), mathjaxr (≤2.0-0), Matrix (≤1.7-4), MatrixGenerics (≤1.22.0), MatrixModels (≤0.5-4), matrixStats (≤1.5.0), mclogit (≤0.9.15), mclust (≤6.1.2), mcmc (≤0.9-8), MCMCglmm (≤2.36), MCMCpack (≤1.7-1), memisc (≤0.99.31.8.3), memoise (≤2.0.1), memuse (≤4.2-3), metadat (≤1.4-0), metafor (≤4.8-0), metapod (≤1.18.0), methods (≤4.5.1), mgcv (≤1.9-4), microbenchmark (≤1.5.0), miloR (≤2.6.0), mime (≤0.13), miniUI (≤0.1.2), minqa (≤1.2.8), mirai (≤2.6.1), mixtools (≤2.0.0.1), mlogit (≤1.1-3), mnormt (≤2.1.1), ModelMetrics (≤1.2.2.2), modelr (≤0.1.11), monocle3 (≤1.4.26), mosaic (≤1.9.2), mosaicCore (≤0.9.5), mosaicData (≤0.20.4), motifmatchr (≤1.32.0), msgdbr (≤26.1.0), multcompView (≤0.1-10), multicross (≤2.1.0), MuSiC (≤1.0.0), mvtnorm (≤1.3-3), nabor (≤0.5.0), nanonext (≤1.8.0), NbClust (≤3.0.1), nbpMatching (≤1.5.6), network (≤1.19.0), nleqslv (≤3.3.5), nlme (≤3.1-168), nloptr (≤2.2.1), NLP (≤0.3-2), NMF (≤0.28), nnet (≤7.3-20), nnls (≤1.6), nortest (≤1.0-4), numDeriv (≤2016.8-1.1), OmnipathR (≤3.18.2), openssl (≤2.3.4), openxlsx (≤4.2.8.1), ordinal (≤2025.12-29), org.Hs.eg.db (≤3.22.0), org.Mm.eg.db (≤3.22.0), otel (≤0.2.0), pak (≤0.10.0), paletteer (≤1.6.0), pals (≤1.1), parallel (≤4.5.1), parallelly (≤1.45.1), patchwork (≤1.3.2), pbapply (≤1.7-4), pbkrtest (≤0.5.5), pbmc3k.SeuratData (≤3.1.4), pbmcapply (≤1.5.1), pbmcRef.SeuratData (≤1.0.0), permute (≤0.9-8), pheatmap (≤1.0.13), pillar (≤1.11.1), pixmap (≤0.4-14), pkgbuild (≤1.4.8), pkgconfig (≤2.0.3), pkgdown (≤2.2.0), pkgload (≤1.4.1), plogr (≤0.2.0), plotly (≤4.11.0), plotrix (≤3.8-13), plyr (≤1.8.9), png (≤0.1-8), Polychrome (≤1.5.4), polyclip (≤1.10-7), polynomial (≤1.4-1), posterior (≤1.6.1), pracma (≤2.4.6), praise (≤1.0.0), preprocessCore (≤1.72.0), presto (≤1.0.0), prettyunits (≤1.2.0), princurve (≤2.1.6), prismatic (≤1.1.2), pROC (≤1.19.0.1), processx (≤3.9.0), prodlim (≤2025.04.28), profvis (≤0.4.0), progress (≤1.2.3), progressr (≤0.18.0), projpred (≤2.10.0), promises (≤1.5.0), ProtGenerics (≤1.42.0), proxy (≤0.4-27), proxyC (≤0.5.2), ps (≤1.9.3), pscl (≤1.5.9), psych (≤2.5.6), purrr (≤1.2.0), pwalgn (≤1.6.0), qclMatrix (≤0.9.9), qs2 (≤0.1.7), quadprog (≤1.5-8), qualpalr (≤2.0.0), quantreg (≤6.1), QuickJSR (≤1.9.0), qvalue (≤2.42.0), R.methodsS3 (≤1.8.2), R.oo (≤1.27.1), R.utils (≤2.13.0), R6 (≤2.6.1), ragg (≤1.5.0), ragnar (≤0.3.0), ranger (≤0.18.0), RANN (≤2.6.2), rappdirs (≤0.3.3), rbitutils (≤2.4), RcisTarget (≤1.29.0), rcmdcheck (≤1.4.0), RColorBrewer (≤1.1-3), Rcpp (≤1.1.1-1.1), RcppAnnoy (≤0.0.22), RcppArmadillo (≤15.2.2-1), RcppEigen (≤0.3.4.0.2), RcppHNSW (≤0.6.0), RcppHungarian (≤0.3), RcppML (≤0.3.7), RcppNumerical (≤0.6-0), RcppParallel (≤5.1.11-1), RcppProgress (≤0.4.2), RcppRoll (≤0.3.1), RcppTOML (≤0.2.3), RCurl (≤1.98-1.17), Rdpack (≤2.6.4), reactable (≤0.4.4), reactR (≤0.6.1), readr (≤2.1.6), readxl (≤1.4.5), recipes (≤1.3.1), reformulas (≤0.4.2), registry (≤0.5-1), rematch (≤2.0.0), rematch2 (≤2.1.2), remotes (≤2.5.0), repress (≤2.1.1), reshape (≤0.8.10), reshape2 (≤1.4.5), ResidualMatrix (≤1.20.0), restfulr (≤0.0.16), reticulate (≤1.44.1), Rfast (≤2.1.5.2), rgl (≤1.3.31), rhdf5 (≤2.54.0), rhdf5filters (≤1.22.0), Rhdf5lib (≤1.32.0), RhpcBLASctl (≤0.23-42), Rhtslib (≤3.6.0), Rigraphlib (≤1.2.0), rjson (≤0.2.23), rlang (≤1.2.0), rlist (≤0.4.6.2), rmarkdown (≤2.3), rnatleearth (≤1.2.0.9000), rngtools (≤1.5.2), robustbase (≤0.99-6), ROCR (≤1.0-11), roxygen2 (≤7.3.3), rpart (≤4.1.24), rprojroot (≤2.1.1), rrvgo (≤1.22.0), rsample (≤1.3.1), Rsampletools (≤2.26.0), RSpectra (≤0.16-2), RSQLite (≤2.4.5), rstan (≤2.32.7), rstantools (≤2.6.0), rstatis (≤0.7.3), rstudioapi (≤0.17.1), rsvd (≤1.0.5), rtracklayer (≤1.70.0), Rtsne (≤0.17), RTtf2pt1 (≤1.3.14), rversions (≤3.0.0), rvest (≤1.0.5), s2 (≤1.1.9), S4Arrays (≤1.10.0), S4Vectors (≤0.48.0), S7 (≤0.2.1), sass (≤0.4.0.1), sbtools (≤1.4.1), ScaledMatrix (≤1.18.0), scales (≤1.4.0), scatter (≤1.38.0), scattermore (≤1.2), scatterpie (≤0.2.6), scatterplot3d (≤0.3-44), sccore (≤1.0.6), scCustomize (≤3.2.2), scDbfFinder (≤1.24.0), sceasy (≤0.0.7), scico (≤1.5.0), SCPA (≤1.6.2), scan (≤1.38.0), scrapper (≤1.4.0), scRepertoire (≤2.5.8), sctransform (≤0.4.2), scuttle (≤1.20.0), segmented (≤2.1-4), selectr (≤0.5-0), Seqinfo (≤1.0.0), seqLogo (≤1.76.0), sessioninfo (≤1.2.3), Seurat (≤5.3.1), SeuratData (≤0.2.2.9002), SeuratDisk (≤0.0.0.9021), SeuratExtend (≤1.2.7), SeuratExtendData (≤0.3.0), SeuratObject (≤2.0.0), SeuratWrappers (≤0.4.0), sf (≤1.0-23), shape (≤1.4.6.1), shiny (≤1.11.1), shinyBS (≤0.61.1), shinychat (≤0.3.0), shinydashboard (≤0.7.3), shinyjs (≤2.1.0), shinyLP (≤1.1.3), shinyWidgets (≤0.9.1), ShortRead (≤1.68.0), Signac (≤1.16.0), SignatuR (≤0.3.0),

SingleCellExperiment ($\leq 1.32.0$), SingleR ($\leq 2.12.0$), singscore ($\leq 1.30.0$), sitmo ($\leq 2.0.2$), slam ($\leq 0.1.55$), slider ($\leq 0.3.3$), slingshot ($\leq 2.7.0$), sm ($\leq 2.2.6.0$), sn ($\leq 2.1.1$), sna (≤ 2.8), snakecase ($\leq 0.11.1$), snow ($\leq 0.4.4$), SnowballC ($\leq 0.7.1$), SoupX ($\leq 1.6.2$), sourcetools ($\leq 0.1.7-1$), sp ($\leq 2.2.0$), spacexr ($\leq 2.2.1$), spam ($\leq 2.11-1$), SparseArray ($\leq 1.10.3$), SparseM ($\leq 1.84-2$), sparseMatrixStats ($\leq 1.22.0$), sparsesvd ($\leq 0.2-3$), sparsevctrs ($\leq 0.3.6$), spatial ($\leq 7.3-18$), SpatialExperiment ($\leq 1.20.0$), spatstat.data ($\leq 3.1-9$), spatstat.explore ($\leq 3.6-0$), spatstat.geom ($\leq 3.6-1$), spatstat.random ($\leq 3.4-3$), spatstat.sparse ($\leq 3.1-0$), spatstat.univar ($\leq 3.1-5$), spatstat.utils ($\leq 3.2-0$), spData ($\leq 2.3.4$), spdep ($\leq 1.4-1$), speckle ($\leq 1.10.0$), speedglm ($\leq 0.3-5$), splancs ($\leq 2.01-45$), splines ($\leq 4.5.1$), SPLIT ($\leq 0.2.3$), SQUAREM (≤ 2021.1), StanHeaders ($\leq 2.32.10$), statmod ($\leq 1.5.1$), statnet.common ($\leq 4.12.0$), stats ($\leq 4.5.1$), stats4 ($\leq 4.5.1$), stringfish ($\leq 0.18.0$), stringi ($\leq 1.8.7$), stringr ($\leq 1.6.0$), SummarizedExperiment ($\leq 1.40.0$), survival ($\leq 3.8-3$), sva ($\leq 3.58.0$), svglite ($\leq 2.2.2$), sys ($\leq 3.4.3$), systemfonts ($\leq 1.3.1$), tcltk ($\leq 4.5.1$), tensor ($\leq 1.5.1$), tensorA ($\leq 0.36.2.1$), terra ($\leq 1.8-86$), testthat ($\leq 3.3.1$), textshaping ($\leq 1.0.4$), TFBSTools ($\leq 1.48.0$), TFMPvalue ($\leq 0.0.9$), tibble ($\leq 3.3.0$), tidygraph ($\leq 1.3.1$), tidyr ($\leq 1.3.1$), tidyselect ($\leq 1.2.1$), tidyterra ($\leq 0.7.2$), tidytree ($\leq 0.4.6$), tidyverse ($\leq 2.0.0$), timechange ($\leq 0.7.2$), timeDate (≤ 4052.112), tinytex (≤ 0.58), tislingshot ($\leq 2.0.1$), tm ($\leq 0.7-18$), TOAST ($\leq 1.22.0$), tonsilref.SeuratData ($\leq 2.0.0$), tools ($\leq 4.5.1$), tradeSeq ($\leq 1.24.0$), TrajectoryUtils ($\leq 1.18.0$), transport ($\leq 0.15-4$), treeio ($\leq 1.34.0$), treemap ($\leq 2.4-4$), tweenr ($\leq 2.0.3$), TxDb.Hsapiens.UCSC.hg19.knownGene ($\leq 3.22.1$), TxDb.Hsapiens.UCSC.hg38.knownGene ($\leq 3.22.0$), tximport ($\leq 1.38.1$), tzdb ($\leq 0.5.0$), UCell ($\leq 2.14.0$), ucminf ($\leq 1.2.3$), UCSC.utils ($\leq 1.6.0$), umap ($\leq 0.2.10.0$), units ($\leq 1.0-0$), UpSetR ($\leq 1.4.0$), urlchecker ($\leq 1.0.1$), usethis ($\leq 3.2.1$), utf8 ($\leq 1.2.6$), utils ($\leq 4.5.1$), uuid ($\leq 1.2-1$), uwot ($\leq 0.2.4$), V8 ($\leq 8.0.1$), VariantAnnotation ($\leq 1.56.0$), vctrs ($\leq 0.6.5$), vegan ($\leq 2.8-0$), venn (≤ 1.12), VennDiagram ($\leq 1.8.2$), VGAM ($\leq 1.1-14$), vioplot ($\leq 0.5.1$), vipor ($\leq 0.4.7$), viridis ($\leq 0.6.5$), viridisLite ($\leq 0.4.2$), visNetwork ($\leq 2.1.4$), vroom ($\leq 1.6.7$), vsn ($\leq 3.78.1$), waldo ($\leq 0.6.2$), warp ($\leq 0.2.2$), webshot2 ($\leq 0.1.2$), websocket ($\leq 1.4.4$), whisker ($\leq 0.4.1$), withr ($\leq 3.0.3$), wk ($\leq 0.9.4$), wordcloud (≤ 2.6), WriteXLS ($\leq 6.8.0$), xfun (≤ 0.54), xgboost ($\leq 1.7.11.1$), XML ($\leq 3.99-0.20$), xml2 ($\leq 1.5.0$), xopen ($\leq 1.0.1$), xtable ($\leq 1.8-4$), XVector ($\leq 0.50.0$), yaml ($\leq 2.3.11$), yulab.utils ($\leq 0.2.2$), zellkonverter ($\leq 1.20.1$), zigg ($\leq 0.0.2$), zip ($\leq 3.0.0$), zoo ($\leq 1.8-14$)

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](#)

CITE-seq data are available as processed, fully annotated Seurat objects, and as downsampled data loaded into web-accessible CellxGene instances at GutPath.org. In addition, all raw fastq files for CITE-seq experiments are available on the Sequence Read Archive (SRA accession, PRJNA1378118). Xenium 5K mouse spatial transcriptomic data are available on the Gene Expression Omnibus (GEO accession, GSE313636).

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

NA

Reporting on race, ethnicity, or other socially relevant groupings

NA

Population characteristics

NA

Recruitment

NA

Ethics oversight

NA

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](https://www.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

No sample size calculation was performed. Predetermined sample sizes were used. 2 independent biological replicates were sequenced. at minimum Partial technical replicates were captured through the paired isolation of IEC and LP compartments within each mouse and further

technical replication was captured within each sample through individual cell-level data. We further corroborated single cell results with alternative molecular methods and confirmed concordance across biological replicates.

Data exclusions No data was excluded from this study.

Replication Technical replication of data was made possible through individual cells within a given CITEseq sample and across isolation techniques (IEC and LP) which capture overlapping tissue/cell populations. 2x Biological replicates were collected for CITEseq at minimum to ensure reproducibility across each infection. Xenium data were similarly produced for 2 biological replicates and utilized to confirm CITEseq findings through and independent experimental technique (Figure S8). Separate experiments (Mass spec, IHC) were performed on independent experiments and samples to corroborate transcriptional findings. Mass spec of fatty acids/bile acids was replicated across 4 individual mice within one experiment.

Randomization No randomization would have applied to this study.

Blinding No blinding was applied to this study

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
- ☐ ☒ Antibodies
- ☒ ☐ Eukaryotic cell lines
- ☒ ☐ Palaeontology and archaeology
- ☐ ☒ Animals and other organisms
- ☒ ☐ Clinical data
- ☒ ☐ Dual use research of concern
- ☒ ☐ Plants

Methods

- n/a Involved in the study
- ☒ ☐ ChIP-seq
- ☒ ☐ Flow cytometry
- ☒ ☐ MRI-based neuroimaging

Antibodies

Antibodies used

All antibodies used come from TotalSeq™-B Mouse Universal Cocktail, Cat. no. 199902. Antibodies in the CITE-seq cocktail include: anti-mouse CD4 (RM4-5), anti-mouse CD8a (53-6.7), anti-mouse CD366 (Tim-3) (RMT3-23), anti-mouse CD279 (PD-1) (RMP1-30), anti-mouse CD117 (c-kit) (2B8), anti-mouse Ly-6C (HK1.4), anti-mouse/human CD11b (M1/70), anti-mouse Ly-6G (1A8), anti-human/mouse CD49f (GoH3), anti-mouse/human CD44 (IM7), anti-mouse CD54 (YN1/1.7.4), anti-mouse CD90.2 (Thy1.2) (30-H12), anti-mouse CD73 (TY/11.8), anti-mouse CD49d (R1-2), anti-mouse CD200 (OX2) (OX-90), Mouse IgG1, κ isotype Ctrl (MOPC-21), Mouse IgG2a, κ isotype Ctrl (MOPC-173), Mouse IgG2b, κ isotype Ctrl (MPC-11), anti-mouse CD19 (6D5), Rat IgG2b, κ isotype Ctrl (RTK4530), anti-mouse CD45 (30-F11), anti-mouse CD25 (PC61), anti-mouse/human CD45R/B220 (RA3-6B2), anti-mouse CD115 (CSF-1R) (AFS98), anti-mouse CD11c (N418), anti-mouse CD5 (53-7.3), anti-mouse CD62L (MEL-14), anti-mouse F4/80 (BM8), anti-mouse FcεRIα (MAR-1), anti-mouse I-A/I-E (M5/114.15.2), anti-mouse NK-1.1 (PK136), anti-mouse TCR β chain (H57-597), anti-mouse Ly-6A/E (Sca-1) (D7), anti-mouse CD45.2 (104), anti-human/mouse/rat CD278 (ICOS) (C398.4A), anti-mouse CD3 (17A2), anti-mouse CD274 (B7-H1, PD-L1) (MIH6), anti-mouse/rat/human CD27 (LG.3A10), anti-mouse CD20 (SA275A11), anti-mouse CD357 (GITR) (DTA-1), anti-mouse CD69 (H1.2F3), anti-mouse CD127 (IL-7Rα) (A7R34), anti-mouse CD86 (GL-1), anti-mouse CD103 (2E7), anti-mouse CD150 (SLAM) (TC15-12F12.2), anti-mouse TCR Vy2 (UC3-10A6), anti-mouse CD24 (M1/69), anti-human/mouse integrin β7 (FIB504), anti-mouse CD106 (429 (MVCAM.A)), anti-mouse CD122 (IL-2Rβ) (5H4), anti-mouse CD183 (CXCR3) (CXCR3-173), anti-mouse CD8b (Ly-3) (YTS156.7.7), anti-mouse TCR Vβ8.1, 8.2 (KJ16-133.18), Rat IgG1, κ isotype Ctrl (RTK2071), Rat IgG2a, κ isotype Ctrl (RTK2758), Armenian Hamster IgG Isotype Ctrl (HTK888), anti-mouse/human KLRG1 (MAFA) (2F1/KLRG1), anti-mouse CD223 (LAG-3) (C9B7W), anti-mouse CD300LG (Nepmucin) (ZAQ5), anti-mouse CD163 (S15049I), anti-mouse CD49b (HMA2), anti-mouse CD172a (SIRPα) (P84), anti-mouse CD48 (HM48-1), anti-mouse CD170 (Siglec-F) (S17007L), anti-mouse CD169 (Siglec-1) (3D6.112), anti-mouse CD41 (MWReg30), anti-mouse CD304 (Neuropilin-1) (3E12), anti-mouse CD36 (HM36), anti-mouse CD38 (90), anti-mouse CD55 (DAF) (RIKO-3), anti-mouse CD63 (NVG-2), anti-mouse CD68 (FA-11), anti-mouse CD79b (Igβ) (HM79-12), anti-mouse CX3CR1 (SA011F11), anti-mouse MERTK (Mer) (2B10C42), anti-mouse CD301b (MGL2) (URA-1), anti-mouse Tim-4 (RMT4-54), anti-mouse/rat XCR1 (ZET), anti-mouse IgD (11-26c.2a), anti-mouse CD11a (M17/4), anti-mouse CD138 (Syndecan-1) (281-2), anti-mouse CD317 (BST2, PDCA-1) (927), anti-mouse CD9 (MZ3), anti-mouse CD34 (SA376A4), anti-mouse CD371 (CLEC12A) (5D3/CLEC12A), anti-mouse CD22 (OX-97), anti-mouse Ly49H (3D10), anti-mouse Ly49D (4E5), anti-mouse Ly-49A (YE1/48.10.6), anti-mouse CD185 (CXCR5) (L138D7), anti-mouse CD49a (HMA1), anti-mouse CD1d (CD1.1, Ly-38) (1B1), anti-mouse CD199 (CCR9) (CW-1.2), anti-mouse CD40 (3/23), anti-mouse CD31 (390), anti-mouse CD107a (LAMP-1) (1D4B), anti-mouse CD182 (CXCR2) (SA044G4), anti-mouse/rat CD61 (2C9.G2 (HMβ3-1)), anti-mouse Ly108 (330-AJ), anti-mouse CD98 (4F2) (RL388), anti-mouse CD210 (IL-10 R) (1B1.3a), anti-mouse CD88 (C5aR) (20/70)

Validation

No separate experimental validation of the antibodies was performed. In the manuscript (Figure S3) we demonstrate expected patterns of CITEseq antibody binding such as CD4 antibody expression measured on CD4 T cells and CD19 protein expression detected on B cells.

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	All animals in this study were purchased from Jackson labs and were female C57BL/6J between 4-6 weeks of age upon arrival to the University of Pennsylvania Hill Pavilion animal Vivarium where they were kept for experimental studies. Vivarium conditions include a 12 hour light/dark cycle, ambient temperatures of 68°F - 77°F and humidity ranges from 35% - 55% humidity in accordance with institutional guidelines. All mice between 6-9 weeks of age at organ harvest.
Wild animals	No wild animals were used in this study
Reporting on sex	This study examines intestinal single cell transcriptional profiles for only female mice and thus we cannot confirm that all conclusions contained within are also attributable to male mice.
Field-collected samples	No field-collected samples were used in this study.
Ethics oversight	All animals were cared for under the guidance and approval of the University of Pennsylvania IACUC (#805432, #804523, and #807356).

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

Plants

Seed stocks	NA
Novel plant genotypes	NA
Authentication	NA