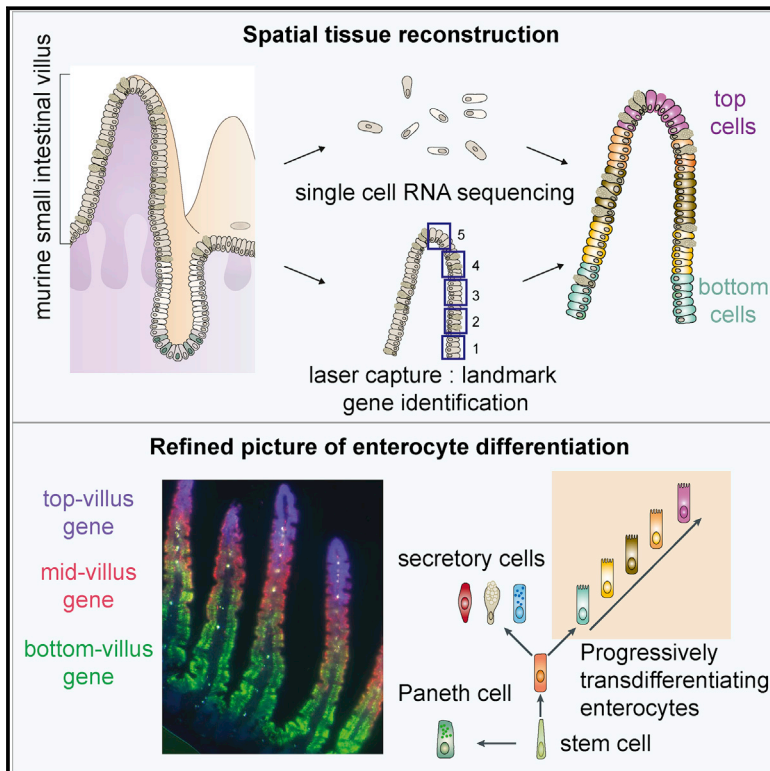# Cell

Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation along the Intestinal Villus Axis

Graphical Abstract

Authors
Andreas E. Moor, Yotam Harnik,
Shani Ben-Moshe, ..., Raya Eilam,
Keren Bahar Halpern, Shalev Itzkovitz

Correspondence
andreas.moor@weizmann.ac.il (A.E.M.),
shalev.itzkovitz@weizmann.ac.il (S.I.)

In Brief

A broadly applicable single-cell spatial transcriptomics approach reveals broad regional and functional heterogeneity of small intestinal enterocytes.

Highlights

- Laser capture microdissection reveals a large panel of enterocyte landmark genes

- These genes are used to spatially localize single RNA-sequenced enterocytes

- Enterocyte function is broadly zonated along the villus axis

- Enterocytes traverse a series of cell states during their migration along the villus

CellPress

Resource

Cell

Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation along the Intestinal Villus Axis

Andreas E. Moor,[1,*] Yotam Harnik,[1] Shani Ben-Moshe,[1] Efi E. Massasa,[1] Milena Rozenberg,[1] Raya Eilam,[2]
Keren Bahar Halpern,[1] and Shalev Itzkovitz[1,3,*]
[1]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel
[2]Department of Veterinary Resources, Weizmann Institute of Science, Rehovot, Israel
[3]Lead Contact
*Correspondence: andreas.moor@weizmann.ac.il (A.E.M.), shalev.itzkovitz@weizmann.ac.il (S.I.)
https://doi.org/10.1016/j.cell.2018.08.063

SUMMARY

The intestinal epithelium is a highly structured tissue composed of repeating crypt-villus units. Enterocytes perform the diverse tasks of absorbing a wide range of nutrients while protecting the body from the harsh bacterium-rich environment. It is unknown whether these tasks are spatially zonated along the villus axis. Here, we extracted a large panel of landmark genes characterized by transcriptomics of laser capture microdissected villus segments and utilized it for single-cell spatial reconstruction, uncovering broad zonation of enterocyte function along the villus. We found that enterocytes at villus bottoms express an anti-bacterial gene program in a microbiome-dependent manner. They next shift to sequential expression of carbohydrates, peptides, and fat absorption machineries in distinct villus compartments. Finally, they induce a Cd73 immune-modulatory program at the villus tips. Our approach can be used to uncover zonation patterns in other organs when prior knowledge of landmark genes is lacking.

INTRODUCTION

The intestinal tract is responsible for nutrient digestion and absorption, secretion of mucus and hormones, interactions with commensal microbiota, and protection of the organism from pathogenic microbes (Crosnier et al., 2006; van der Flier and Clevers, 2009). This wide array of tasks requires the presence of different cell types that are specialized for their respective functions. Enterocytes, which represent the majority of cells in the epithelial layer, constantly migrate along the villus walls until they are shed off from their tips 3–5 days after their emergence from crypts. The positions of enterocytes along the villus axis correlate with their age (Moor and Itzkovitz, 2017), exposure to morphogen gradients (Crosnier et al., 2006), and hypoxia (Zheng et al., 2015), but the positional effects on enterocyte function are largely unknown. Previous work investigated transcriptomic changes along the small intestinal crypt-villus axis with bulk samples and DNA microar-ray-based expression profiles in mouse (Mariadason et al., 2005; Stegmann et al., 2006) and human tissue (George et al., 2008). This body of work revealed some broad compositional differences of the crypt and the villus, but its low spatial resolution (comparing bulk crypts with bulk villi), uncontrolled mixes of different cell types, and the low sensitivity of microarray-based transcriptomics precluded the detection of spatial expression changes and heterogeneity of enterocytes along the villus.

Single-cell RNA sequencing (scRNA-seq) has revolutionized our ability to characterize individual cells in depth (Kolodziejczyk et al., 2015); it was recently utilized in the intestine to identify cell types (Grün et al., 2015) and sub-populations of intestinal stem cells (Yan et al., 2017a), tuft cells (Haber et al., 2017; Herring et al., 2018), and enteroendocrine cells (Glass et al., 2017; Grün et al., 2015; Haber et al., 2017; Yan et al., 2017b). However, spatial heterogeneity and specialization along the villus axis of enterocytes, the largest cell compartment, has not been addressed. Relating such heterogeneity to tissue coordinates is challenging because the spatial origin of individual cells is lost when the tissue is dissociated for scRNA-seq. We and others have developed approaches to spatially reconstruct scRNA-seq data by making use of known expression profiles of landmark genes characterized by RNA *in situ* hybridization (Achim et al., 2015; Halpern et al., 2017; Karaiskos et al., 2017; Satija et al., 2015; Scialdone et al., 2016; Zeisel et al., 2018). This approach is infeasible, however, when no prior knowledge exists regarding zonated landmark genes. Here we established a comprehensive panel of landmark genes, characterized by RNA-seq of laser capture microdissected epithelial samples originating from differential villus zones (Figure 1). We used these to reconstruct the spatial tissue coordinates of enterocytes in scRNA-seq data and uncovered vast heterogeneity and spatial sub-specialization. Our work exposes spatially restricted enterocyte cell states and demonstrates that enterocytes are not terminally differentiated cells but, rather, continuously transdifferentiate as they migrate along the villus axis.

RESULTS

Landmark Genes Enable Spatial Reconstruction of Single Enterocytes along the Villus Axis

Optimized protocols for RNA-seq of laser capture microdissected tissue (LCM-RNA-seq) facilitate obtaining expression
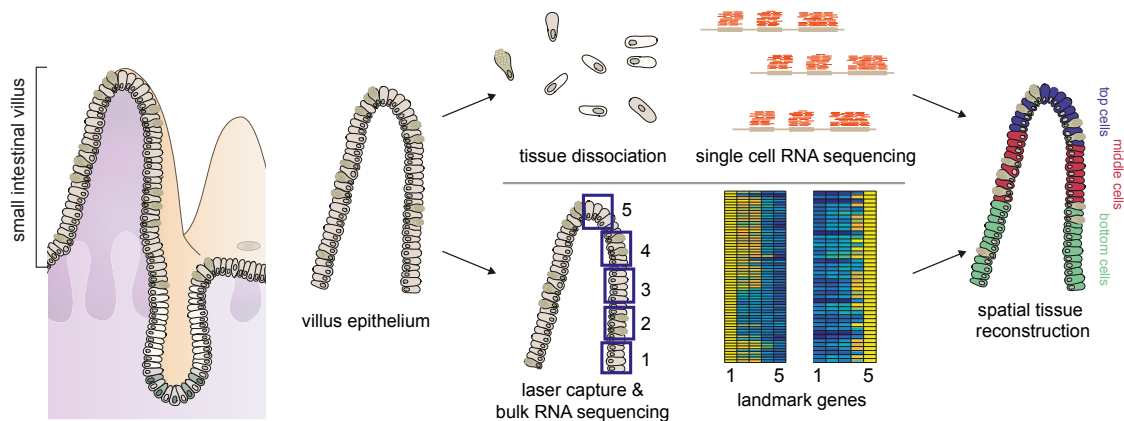
Cell

Figure 1. Schematic of the Experimental Approach

The villus epithelium is dissociated into single cells; these cells are profiled by scRNA-seq. In parallel, spatial landmark genes are retrieved by bulk RNA-seq of villus quintiles obtained using laser capture microdissection (LCM). The original position of the sequenced single cells is then inferred based on their expression levels of the landmark genes.

signatures from exceedingly small tissue regions (Chen et al., 2017; Moor et al., 2017; Nichterwitz et al., 2016; Peng et al., 2016). In the intestine, we have recently achieved coverage of ~2,000 genes in samples equivalent to 10 cells (Moor et al., 2017). To extract a panel of enterocyte landmark genes, we used LCM to isolate epithelial cells from five equally spaced compartments between the bottom and tops of villi in the mouse jejunum (Figure 2A). RNA-seq of these isolated villus quintiles revealed genes with decreasing (Figure 2B) and increasing (Figure 2C) expression gradients. We defined a set of 62 villus bottom landmark genes and 43 villus top landmark genes to be used for spatial reconstruction of scRNA-seq data (Figure S1).

We used our LCM-RNA-seq reference to identify an scRNA-seq dataset (Yan et al., 2017b) that included enterocytes that spanned the entire villus axis (Figure S1). Mature and progenitor enterocytes were clearly demarcated by the expression of Alpi and Mki67, respectively (Figure S1C). We assigned each sequenced mature enterocyte a unit-less spatial coordinate x that was based on the ratio between the summed expression of the top and bottom landmark genes. For each cell, x correlated with its position along the villus axis (STAR Methods; Figures S1H–S1J). By computing the x values of the five laser-captured areas, we were able to assign each cell to one of 6 zones from the bottom to the top of the villus (Figure 2D; Figure S1D). We averaged, for every gene, the expression of single cells in each of these zones to obtain a comprehensive spatial map of gene expression along the intestinal villus (Figures 2E and 2F).

Our spatial reconstruction included more than 9,832 enterocyte-expressed genes, 8,126 of which (83%) were significantly zonated (STAR Methods; q < 0.05). Thus, differentiated enterocytes exhibit ubiquitous spatial heterogeneity, with only a small minority of genes invariably expressed from the bottom to the top of the villi. We used single-molecule fluorescence in situ hybridization (smFISH; Halpern et al., 2017) to validate our predicted zonated expression profiles for 15 enterocyte genes,

demonstrating the accuracy of reconstruction (Figure S2F). Reconstruction errors were relatively low for panels of 50 landmark genes but became more substantial for reduced panels of 20 landmark genes (Figures S2A and S2B). Our approach faithfully reconstructed the zonated expression patterns even for genes that exhibited high local variability in gene expression, such as Reg1 and Reg3g (Figures S2C–S2E). The reconstructed spatial gene expression map that we derived by combining LCM and scRNA-seq was more accurate than only using LCM for spatial transcriptomics (Figure S3).

Clustering of Zonated Genes Reveals Distinct Spatial Enrichment of Enterocyte Functions

To study the design principles of the uncovered villus zonation patterns, we used k-means to cluster the genes into five distinct groups, ordered from villus bottom to top according to their average zonation profiles (Figure 3). Gene set enrichment analysis (Subramanian et al., 2005) revealed enriched gene ontology (GO) terms for each cluster (Figure 3). Cluster 1 (29% of the genes) contained genes that decreased progressively from crypts to villus tips. These included a global decline of translation, transcription, and RNA splicing genes. Thus, enterocyte biosynthetic capacity seems to be gradually decreasing as enterocytes migrate along the villus axis. Mitochondrial GO terms were enriched in cluster 2 (20% of genes; Figure 3; Figure S4A). This decrease in mitochondrial content may be an adaptation to the decreasing gradient of oxygen concentration, previously demonstrated along the villus axis (Zheng et al., 2015). Cluster 2 also contained glutathione transferase activity, which contains Gstm3 (Figure S4B and S4C), as well as acute phase response genes such as Reg genes, which were highly expressed at the villus bottom but not expressed in the adjacent crypt (Figure 4; Figures S2D and S2E). Cluster 3 (20% of genes) consisted of intestinal transport annotations, which peaked at the mid-villus zones. Genes in cluster 4 (12% of genes) increased in expression up to the mid-villus zones and included many brush border components. Cluster 5 (19% of genes) included lipoprotein
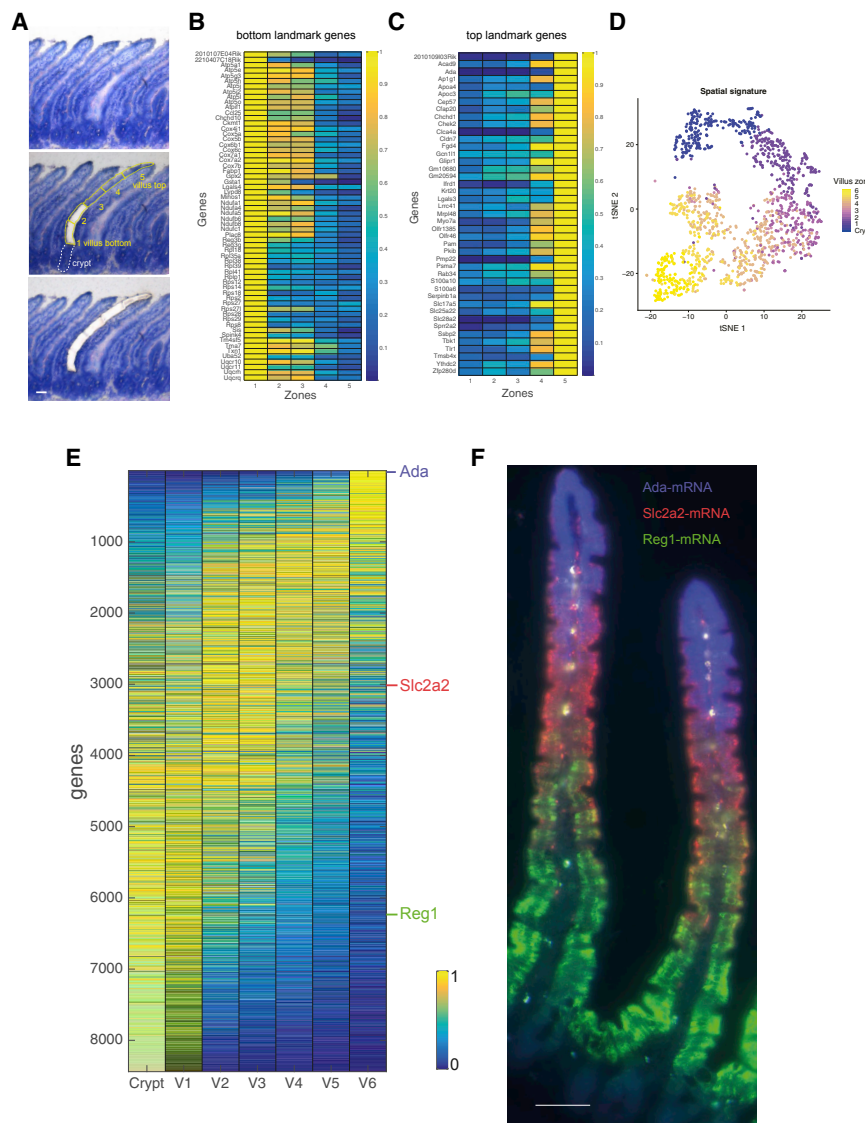
Figure 2. Spatial Reconstruction of Villus Enterocytes

(A) LCM of villus epithelium quintiles. Scale bar, 50 μm.

(B) LCM-RNA-seq expression of the villus bottom landmark genes.

(C) LCM-RNA-seq expression of the villus top landmark genes.

The profiles in (B) and (C) are normalized to the maximum expression level for each gene.

(D) t-distributed stochastic neighbor embedding (tSNE) plot of the enterocyte and progenitor populations in a sorted Lgr5-negative dataset. Each cell is colored according to its inferred villus zone. Depicted analyses are based on raw data from NCBI GEO datasets GSM2644349 and GSM2644350.

(E) Spatially reconstructed gene expression zonation profiles. Profiles are normalized to their maximum and sorted according to their center of mass. Shown are genes with a maximal zonation value above 10^{-5}.

(F) smFISH staining of Ada mRNA (blue), which is expressed in villus tip enterocytes; Slc2a2 mRNA (red), which is expressed in villus middle enterocytes; and Reg1 mRNA (green), which is expressed in villus bottom enterocytes. Scale bar, 50 μm.

See also Figures S1–S3 and Table S5.

(Okumura et al., 2016) (Figure 4A). Reg3b and Reg3g are the two most significantly downregulated genes when comparing RNA-seq of conventional with germ-free mice (Peck et al., 2017; Figures 4D and 4E). Their expression is strongly decreased upon microbiome ablation with antibiotic intervention (Chevalier et al., 2015; Figure 4F). We used smFISH to demonstrate a sharp decrease in Reg3g at the bottom of the villus in germ-free mice compared with controls (Figures 4B and 4C). These findings reveal an antimicrobial zone above the crypt that might function as a gatekeeper for the crypt stem cell niche to minimize its exposure to pathogenic microbes.

biosynthesis and cell adhesion processes as well as the long non-coding RNA (lncRNA) markers of paraspeckles (Neat1; Figures S4F and S4G) and speckles (Malat1; Figures S4H and S4I), all monotonically increasing toward the villus tips.

Villus Bottom Cells Express a Microbiome-Dependent Antimicrobial Program

Genes of the Reg family belong to the calcium-dependent lectin genes and encode small secretory proteins (Vaishnava et al., 2011; Burger-van Paassen et al., 2012; Earle et al., 2015). Our spatial reconstruction uncovered a restricted zone at the bottom of the villus in which enterocytes strongly expressed Reg family members as well as other peptides involved in microbiota-host interactions, such as the enterocyte inflammasome components Nlrp6 (Elinav et al., 2011), Il18 (Nowarski et al., 2015), Ccl25 (Bowman et al., 2002), and antibacterial Lypd8

Zonation of Nutrient Transporters

Enterocytes absorb a wide range of nutrients, including carbohydrates, amino acids, and lipids. We found that the transporters for these key nutrient families exhibited distinct zonation profiles (Figures 5A–5E; within-nutrient family profile distance of 0.43 versus between-nutrient distance of 1.23, Kruskal-Wallis p = 1.56e−9; Figure 5F). Although amino acid and carbohydrate transporters were enriched at the middle of the villus (Figure 5A), Slc15a1, which encodes the main peptide transporter Pept1, was shifted in expression toward the upper villus zones, and the cholesterol transporter Npc1l1 and the lipoprotein biosynthesis machinery, necessary for the assembly of chylomicrons,
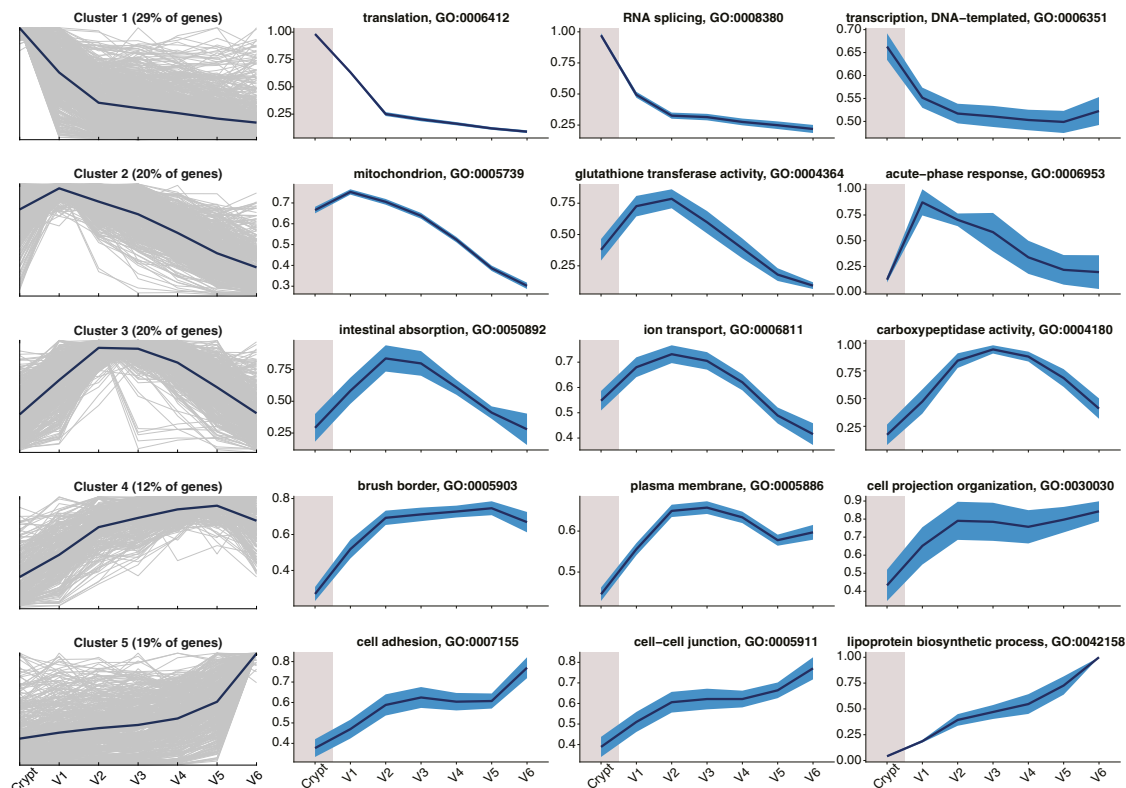
Cell

Figure 3. Functional Sub-specialization of Villus Enterocytes

Column 1: 5 clusters of genes with similar zonation profiles along the villus length. Blue lines represent cluster mean; gray lines depict individual genes. Columns 2 to 4: representative gene ontology (GO) terms enriched in the identified gene clusters. Dark blue lines show GO term mean; light blue areas denote SEM. See also Figure S4 and Tables S1 and S2.

peaked in expression at the villus tips (Figure 5A). The zonated expression of lipoprotein genes at the villus tops can explain previous findings of higher chylomicron density at the villus tips shortly after lipid gavage (Seyer et al., 2013). Thus, enterocytes seem to be sub-specialized in preferential nutrient absorption according to their position along the villus axis.

Villus Tip Cells Exhibit a Distinct Expression Program

Our spatial reconstruction revealed a sharp increase in the expression of distinct signaling pathways and transcription factor sets at the villus tips. These genes included Egfr, Klf4, and the AP-1 transcription factors Fos and Junb (Figures 6A and 6B). Egfr signaling has been implicated in tight junction organization (Tran et al., 2012). Its increased expression at the villus tips might initiate reorganization of cell adhesion (Figure 3; Figures S5A and S5B), which is necessary for subsequent cellular shedding.

Villus tip cells further expressed a signature of purine catabolism genes, including Enpp3, Nt5e, Slc28a2, and Ada (Figures 6C–6E; Figures S5C–S5F). Enpp3 and Nt5e, which are encoding ecto-nucleotidases that convert ATP to AMP and AMP to adenosine (Robson et al., 2006), respectively, were expressed in a sequential manner at the villus tips. Enpp3 increased steeply from villus zones 4 to 6, whereas Nt5e was only expressed in zone 6 at the very tips of the villi. (Figure 6E). The tip-enriched gene Slc28a2 encodes an Na$^+$-coupled high-affinity adenosine transporter (Huber-Ruano et al., 2010) that could shuttle the generated adenosine into the cytosol. There, adenosine can be converted to inosine by adenosine deaminase (Ada), which we also found to be confined in expression to this zone (Figure 6D). We validated the tip-enriched expression of Nt5e protein (also known as Cd73) and observed that most of this ecto-nucleotidase was localized to the luminal side of villus tip enterocytes (Figure 6F). Bacterially dependent luminal ATP is a danger signal that activates intestinal immune cells (Trautmann, 2009). Because adenosine and inosine, the products of the revealed villus tip signaling program, exert potent anti-inflammatory functions in the intestine (Mabley et al., 2003), this villus tip expression program may be important for preventing excessive immune reaction to the microbiome. Indeed, Cd73 knockout mice were shown to suffer from autoimmunity (Blume et al., 2012) and exhibit unresolved inflammation in a colitis model (Bynoe et al., 2012).

Enterocytes are short-lived cells that only operate for a few days as they migrate along the villus, a period that approaches the typical half-lives of many proteins (Schwanhäusser et al., 2011). Thus, decreasing expression gradients of mRNA toward the villus tips should not necessary entail a decline in the protein
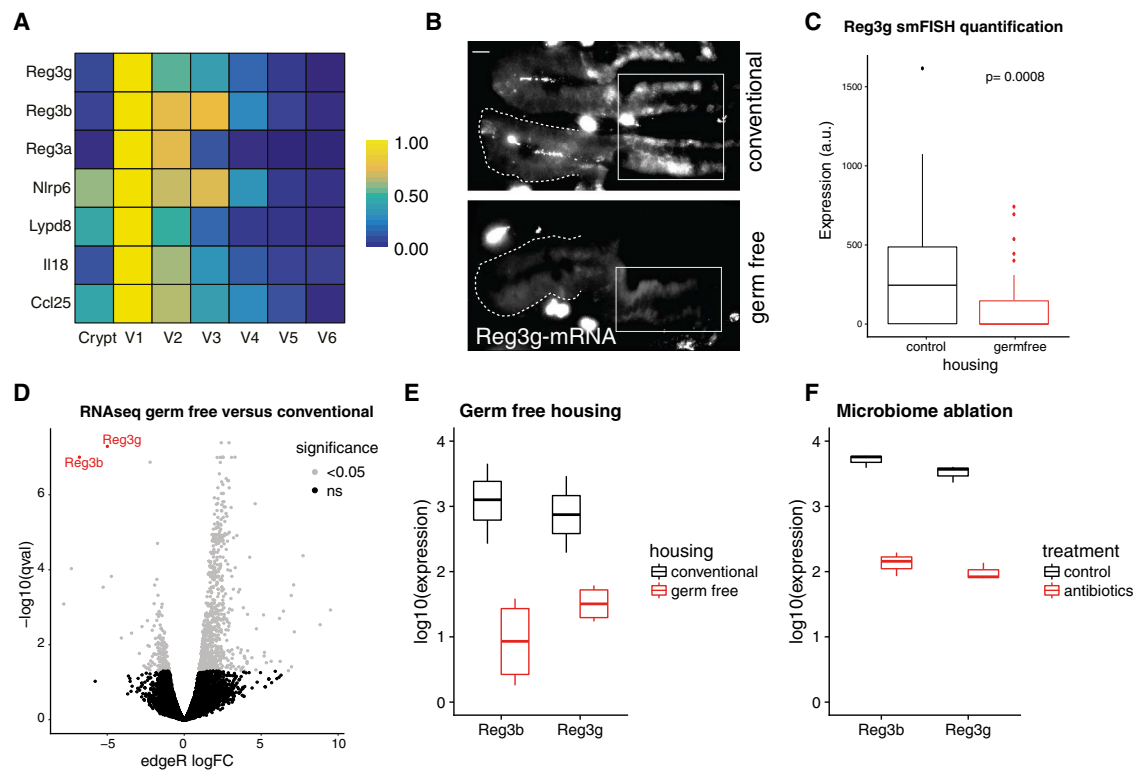
Figure 4. Villus Bottom Cells Express a Microbiome-Dependent Antimicrobial Program

(A) Heatmap of zonation profiles of Reg gene family members as well as other peptides involved in microbiota-host interactions.

(B) smFISH of Reg3g expression in the crypt and adjacent bottom villus area in mice that were either kept conventionally or under germ-free conditions. The dashed lines denote the crypts, and the solid boxes indicate the villus bottom zone where Reg3g is expressed. Scale bar, 10 μm.

(C) Quantification of Reg3g smFISH staining. Cytoplasmic Reg3g intensity of the lowest 10 villus bottom cells was quantified, and the corresponding crypt background staining was subtracted. Data are derived from 2 conventional mice (48 quantified cell stripes) and 2 germ-free mice (42 quantified cell stripes). The p value was calculated with the Wilcoxon rank-sum test.

(D) Volcano plot of differential mRNA expression levels of littermate mice that were either conventionalized for 2 weeks or remained in a germ-free environment. Reg3b and Reg3g exhibit the most significant downregulation in this comparison. Analysis is based on raw data from NCBI GEO: GSE81125 (Peck et al., 2017).

(E) RNA-seq expression levels of Reg3b and Reg3g of littermate mice that were either conventionalized for 2 weeks or remained in a germ-free environment. Analysis is based on raw data from NCBI GEO: GSE81125 (Peck et al., 2017).

(F) RNA-seq expression levels of Reg3b and Reg3g of mice that were treated with vehicle or with antibiotics for 30 days to ablate the microbiome. Analysis is based on raw data from NCBI GEO: GSE74157 (Chevalier et al., 2015). Horizontal lines are medians; boxes are 25–75 percentiles, and vertical lines denote values within 1.5 × interquartile range of the respective quartile.

content. To assess the enterocyte proteome at distinct villus heights, we sorted three populations of enterocytes based on the expression of Nt5e and performed mass spectrometry for proteomics (Figure 6G; Figure S6A). We used RNA-seq to validate that these populations distinguish enterocytes at increasing coordinates along the villus axis (Figures S6B and S6C). We found a similar functional compartmentalization of enterocytes as observed with mRNA (Figure 6G; Table S4). Antimicrobial peptides declined in concentration; transporters for amino acids, carbohydrates, and peptides exhibited distinct domains; and the purine metabolism proteins sharply increased in protein expression, in line with the mRNA profiles. We further validated the zonation of proteins using immunohistochemistry and immunofluorescence for the lower villus genes Tfrc and Reg3b, the mid-villus Slc5a1, and the top villus Cdh1 and Nt5e (Figure S7). Interestingly apolipoproteins exhibited a decrease from the Nt5e-medium to the Nt5e-high population (Figure 6G). Because

apolipoproteins are secreted with lipids in chylomicrons, the combination of high mRNA levels (Figures 5D and 5E) and lower intra-cellular protein levels is consistent with a picture of increased lipid secretion at the villus tips (Seyer et al., 2013).

Continuous Transitions of Enterocytes along the Villus Axis

To demonstrate that enterocytes continuously traverse the zonated cell states as they migrate along the villus axis, we used lineage tracing to track the clonal progenies of single Lgr5+ stem cells. We induced rare reporter recombination in mice that had an Lgr5-cre-ERT2 knockin allele (Barker et al., 2007) and a Confetti reporter cassette (Tóth et al., 2017; Snippert et al., 2010) with low doses of tamoxifen and sacrificed the mice after 10 days. The epithelium exhibited rare clones along the villi (Figure 7A). We observed a continuous transition within these clonal cells in the expression of the bottom villus Reg1
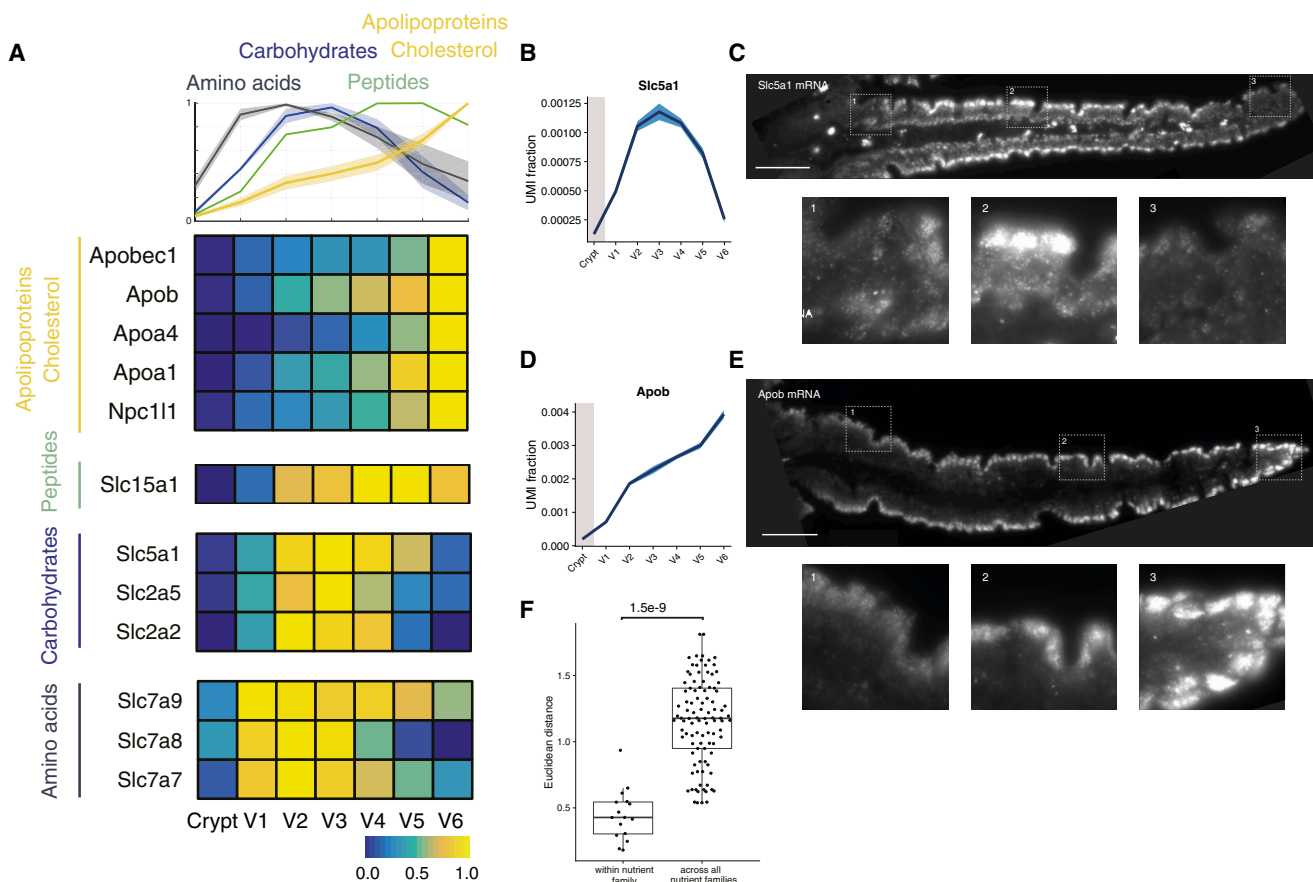
Figure 5. Nutrient Transporters Exhibit Distinct Zonation Profiles

(A) Heatmaps and averaged zonation profiles of selected genes that are important for the absorption of distinct nutrient classes. Black line, amino acid transporters; blue line, carbohydrate transporters; green line, the peptide transporter Slc15a1; yellow line, fat and cholesterol absorption. Profiles are normalized to their maximal value across zones.

(B) scRNA-seq-inferred zonation profile of the sodium-glucose cotransporter Slc5a1. Dark blue line, mean expression; light blue area, SEM.

(C) smFISH of Slc5a1 mRNA expression in whole villus (overview) and in the bottom (1), middle (2), and top (3) parts of the villus (insets).

(D) scRNA-seq-inferred zonation profile of the apolipoprotein B mRNA Apob. Dark blue line, mean expression; light blue area, SEM.

(E) smFISH of Apob mRNA expression in the whole villus (overview) and in the bottom (1), middle (2), and top (3) parts of the villus (insets).

(F) Boxplot of within-nutrient family profile Euclidean distances (n = 16) versus between-nutrient Euclidean distances (n = 100). The p value was obtained with the Kruskal-Wallis test. Distances were computed on the max-normalized zonation profiles.

Scale bars, 50 μm.

and the mid-villus Slc2a2 and between Slc2a2 and the villus tip gene Ada (Figures 7A and 7B). This gradual transition was also apparent when performing pseudo-time analysis of the data using Monocle (Trapnell et al., 2014). Thus, enterocytes continuously traverse an ordered transition of cells states as they migrate along the villus (Figures 7C and 7E).

DISCUSSION

Spatial expression maps are instrumental for understanding the design principles of complex tissues and tumors (Crosetto et al., 2015; Lee et al., 2014; Lein et al., 2017; Moffitt et al., 2016; Moncada et al., 2018; Regev et al., 2017; Shah et al., 2016). Our approach for combining LCM sequencing (LCM-seq) with scRNA-seq is complementary to other spatial transcriptomics

approaches, such as cryo-sectioning (Combs and Eisen, 2013; Junker et al., 2014; Diag et al., 2018) and slide-based spatial transcriptomics (Ståhl et al., 2016; Berglund et al., 2018). Cryo-sectioning-based sequencing physically maps gene expression by mechanically sectioning a tissue along defined axes and has been instrumental in revealing zonation patterns in the *Drosophila* embryo (Combs and Eisen, 2013), the zebrafish embryo (Junker et al., 2014), and the *C. elegans* germline (Diag et al., 2018). This method works efficiently in structures that can be readily aligned along orthogonal axes. Slide-based spatial transcriptomics (Ståhl et al., 2016; Berglund et al., 2018) assays a pre-defined grid with lower spatial resolution (100 μm). The advantages of LCM include the ability to interrogate structures that are harder to align or that are of smaller scales and, thus, not amenable to cryo-sectioning, and the high spatial resolution
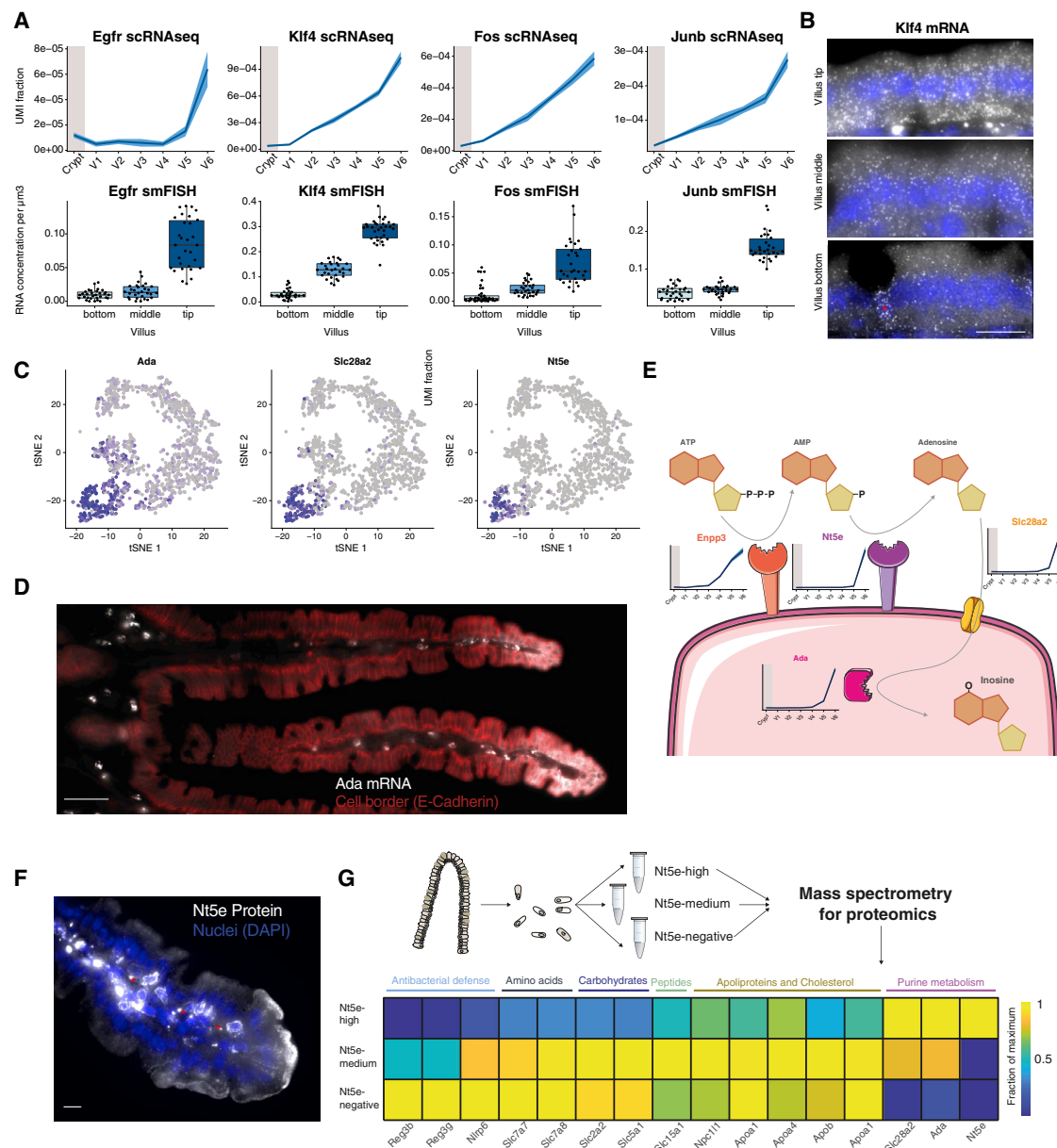
Figure 6. Spatial Reconstruction Reveals a Villus Tip Expression Program

(A) Signaling and transcriptional programs of the villus tip zone. Top row: scRNA-seq-inferred expression profiles. Dark blue line, mean expression; light blue area, SEM. Bottom row: quantification of smFISH expression in the bottom, middle, and top parts of the villus.

(B) smFISH of Klf4 mRNA expression in the bottom, middle, and top parts of the villus. The asterisk in the villus bottom field of view marks a goblet cell (known to express Klf4). Scale bar, 10 μm.

(C) tSNE plots of the expression patterns of Ada, Slc28a2, and Nt5e, three of the identified purine catabolism villus tip marker genes.

(D) smFISH of Ada mRNA (white); cell borders (E-Cadherin protein) are depicted in red. Scale bar, 50 μm.

(E) Model of functional interaction of villus tip genes in purine catabolism. Luminal ATP is hydrolyzed to AMP by Enpp3 and subsequently converted to adenosine by Nt5e. Part of this generated adenosine can be absorbed by the high-affinity adenosine transporter Slc28a2. Intracellular Ada converts adenosine to inosine.

(F) Immunofluorescence staining of Nt5e protein at the villus tip, demonstrating apical localization of the protein. Asterisks mark extracellular Nt5e proteins on intraepithelial lymphocytes. Scale bar, 10 μm.

(G) Protein abundances of cell populations that were sorted according to their Nt5e protein expression levels.

See also Figures S5, S6, and S7 and Table S4.
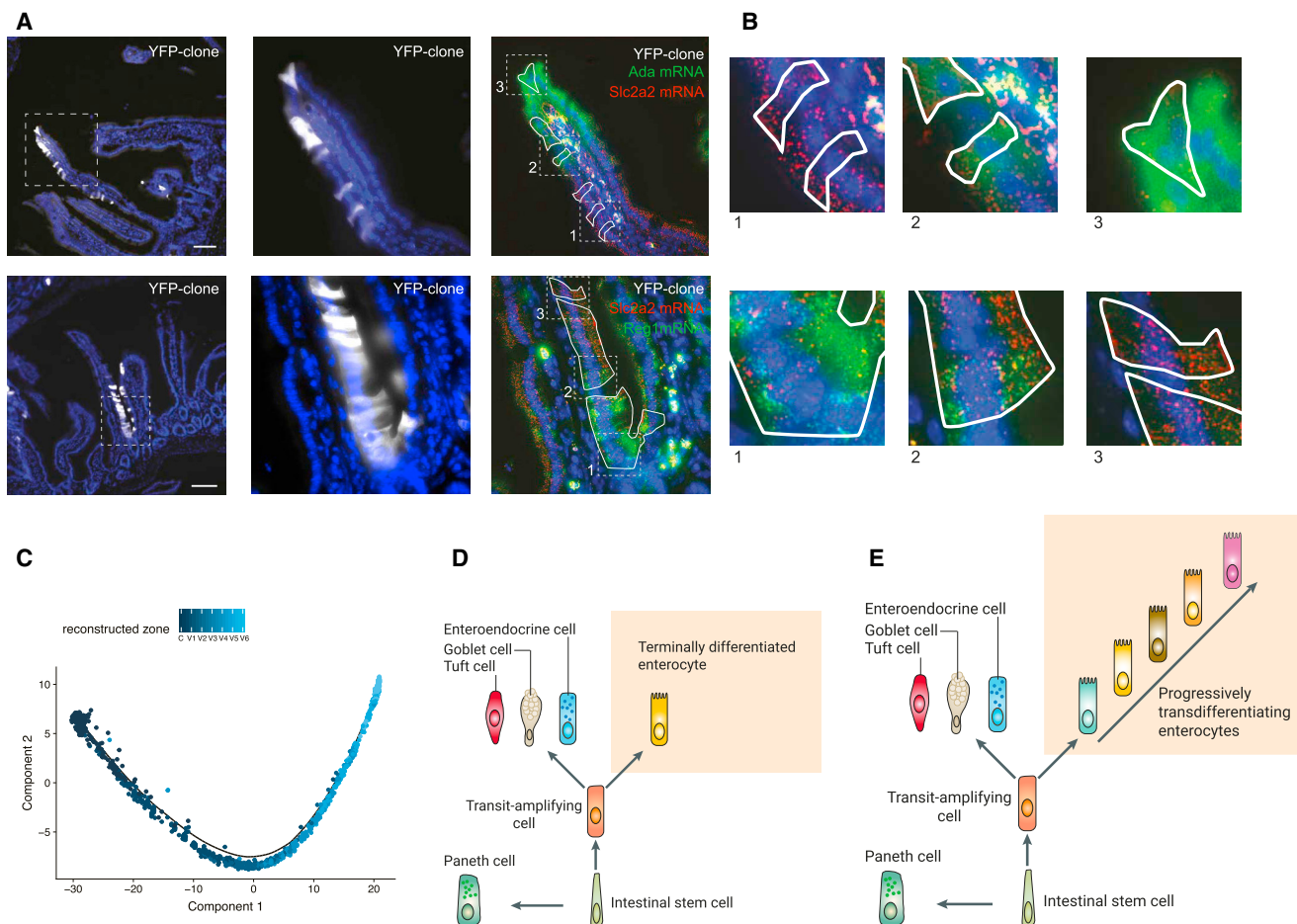
Cell

Figure 7. Enterocytes Transdifferentiate as They Migrate along the Villus Axis

(A) Lineage tracing that tracks the clonal progenies of single Lgr5+ stem cells, demonstrating continuous transitions of enterocytes between the zonated cell states. Right column: smFISH staining of Ada and Slc2a2 mRNA (top) and Slc2a2 and Reg1 mRNA (bottom). The white line represents the outline of the YPF clone.

(B) Insets of (A) that show cells of the same yellow fluorescent protein (YFP) clone that predominantly express the transcripts of the lower villus zone (1), co-express both the lower and higher villus zone transcripts (2), and predominately express the transcripts of the higher villus zone (3).

(C) Visualization of mature enterocytes along the pseudo-time trajectory in reduced dimensional space. Black line, identified trajectory; cells are colored by our reconstructed zones of origin (Figure 2D). Pseudo-time analysis was performed with Monocle (STAR Methods; Trapnell et al., 2014).

(D) The classic view of intestinal epithelial lineages posits a single type of enterocyte.

(E) Our work exposes a series of enterocyte cell states, continuously traversed by enterocytes as they migrate along the villus axis.

Scale bars, 50 μm.

in interrogating non-grid features. Examples of tissues where similar LCM-based spatial reconstruction could be applied include the liver lobules (Halpern et al., 2017), hair follicles, gonadal stem cell niches, the tumor microenvironments (e.g., zonation as a function of distance from invasive front or blood vessels; Heindl et al., 2015), and tissue zonation as a function of distance from localized sites of injury (e.g., sites of inflammation, scar or necrosis; Aragona et al., 2017). Importantly, our approach would only work when the source of spatial variability has a morphological correlate; e.g., tissue structure or apparent cell damage. If the signaling source of spatial variability consists of a local expression pattern of a subset of genes in a seemingly uniform field of cells, then approaches such as grid-based reconstruction would be more suitable.

Landmark gene-based spatial reconstruction of single sequenced cells is another powerful approach for spatial reconstruction but requires *a priori* knowledge of the existence of zonated landmark genes. When such a set is identified, its spatial expression patterns can be precisely characterized using *in situ* approaches (Achim et al., 2015; Halpern et al., 2017; Codeluppi et al., 2018; Karaiskos et al., 2017; Moffitt et al., 2016; Satija et al., 2015; Shah et al., 2016; Zeisel et al., 2018; Zhu et al., 2018). LCM-RNA-seq is an alternative for extracting a large set of landmark genes in an unbiased manner, particularly useful when no prior knowledge of spatial expression patterns exists (Chen et al., 2017; Moor et al., 2017; Nichterwitz et al., 2016; Peng et al., 2016). Although our LCM-RNA-seq measurements provided a coarse spatial expression map (Figure S1A), the

Cell

combination of LCM-RNA-seq with scRNA-seq has important advantages. It enables reconstructing maps of pure cell types, rather than mixtures. This is evident by the seemingly zonated expression profiles in the LCM-seq data of genes that are not expressed in enterocytes (Figures S3A and S3B). It also has higher sensitivity, provided by the massive numbers of sequenced single cells. This was evident by our detection of zonated patterns of lowly expressed genes such as Egfr (Figure 6A), which were only apparent in the scRNA-seq reconstruction and by the more accurate reconstruction of other genes such as Creb3l3, Apob, and Pigr (Figures S3C and S3D). It also provides higher spatial resolution. Although our study coarse-grained the villus into six zones, the single-cell spatial reconstruction assigns every cell a continuous coordinate (Figure S1E) that could be used to explore patterns with still higher spatial frequencies, given enough sampled cells per zone.

Our study uncovered an unexpectedly broad spatial heterogeneity within small intestinal enterocytes; the large majority of genes were significantly zonated, and almost no gene exhibited constant expression levels along the villus axis. The secluded stem cell niche in the intestinal crypt seems to be protected by a layer of gatekeeper enterocytes at the bottoms of the villi that express inflammasome components and secrete antibacterial Reg proteins. These enterocytes may complement secretory Paneth cells in the protection of the crypt-resident stem cells. The absorption machinery of specific nutrients is compartmentalized in distinct villus zones, potentially leading to more efficient nutrient uptake. Villus tip cells appear to orchestrate an immune-modulatory program that might have important implications for host-microbe interactions in health and disease. Thus, enterocytes are not terminally differentiated cells but, rather, continuously transdifferentiate as they migrate along the villus axis (Figure 7E).

We identified a substantial decline in the levels of both mRNA and protein for transporters of amino acids and carbohydrates, an increase in the mRNA levels of apolipoproteins, but a reduction in enterocyte protein content at the villus tips (Figures 5A, 5E, and 6G). This indicates that mid-villus enterocytes preferentially engage in the absorption of amino acids and carbohydrates, whereas villus tip cells are engaged in increased secretion of chylomicrons. Preferential lipid secretion at the villus tips could be a mechanism for optimizing absorption in this relatively hypoxic region of the tissue, where cellular ATP production could be limited. Unlike carbohydrates and amino acids, lipids can freely diffuse and do not require active transport, alleviating the need for ATP-consuming transporters. The zonation of the nutrient absorption machineries we identified could give rise to spatial gradients in the concentrations of nutrients along the luminal sides of the villi. Such spatial heterogeneity could, in turn, result in zonated microbial niches, where distinct microbes would preferentially colonize zones that contain their preferred nutrients. Differential abundance of bacterial taxa has been demonstrated near or distant from the mucosal surface (Nava et al., 2011). It will be interesting to apply LCM to characterize the mucosal microbial profile at a high spatial resolution along the villus axis. Moreover, LCM could be used to explore the zonation of cells in the *lamina propria* (Honda and Littman, 2016; Yissachar et al., 2017); e.g., lymphocytes, myofibroblasts,

and neurons. Obtaining such a transcriptional spatial map of the microbial, epithelial, and mesenchymal components in the gut would reveal the molecular details of their cross-talk.

Our flow cytometry approach to isolate bulk, spatially stratified enterocyte populations (Figure S6A) could enable deeper characterization of the epigenome, metabolome, mutation signatures, mRNA modifications, and other cellular properties along the villus spatial axis. It would be interesting to use our method to explore the zonation profiles of enterocytes in diverse intestinal pathologies. More generally, the use of LCM-RNA-seq to extract a large set of landmark genes in an unbiased manner, is a generic alternative to FISH-based single-cell spatial reconstructions (Achim et al., 2015; Halpern et al., 2017; Karaiskos et al., 2017; Satija et al., 2015), particularly useful when no prior knowledge of zonation exists. This could be used to reconstruct expression cell atlases of other tissues and tumors (Han et al., 2018; The Tabula Muris Consortium et al., 2017; Regev et al., 2017).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Animal experiments
- METHOD DETAILS
  - Single molecule FISH
  - Immunofluorescence
  - Immunohistochemistry
  - LCM
  - Lineage tracing
  - RNA-seq
  - Flow cytometry
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - smFISH quantification
  - Bulk RNA-seq analysis
  - scRNAseq analysis
  - Zonation reconstruction algorithm
  - Pseudotime analysis
  - Clustering and gene ontology enrichment
  - Mass spectrometry for proteomics
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and five tables and can be found with this article online at https://doi.org/10.1016/j.cell.2018.08.063.

Cell

AUTHOR CONTRIBUTIONS

DECLARATION OF INTERESTS

REFERENCES

Achim, K., Pettit, J.-B., Saraiva, L.R., Gavriouchkina, D., Larsson, T., Arendt, D., and Marioni, J.C. (2015). High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. Nat. Biotechnol. 33, 503–509.

Aragona, M., Dekoninck, S., Rulands, S., Lenglez, S., Mascré, G., Simons, B.D., and Blanpain, C. (2017). Defining stem cell dynamics and migration during wound healing in mouse skin epidermis. Nat. Commun. 8, 14684.

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

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B Methodol. 57, 289–300.

Berglund, E., Maaskola, J., Schultz, N., Friedrich, S., Marklund, M., Bergenstråhle, J., Tarish, F., Tanoglidi, A., Vickovic, S., Larsson, L., et al. (2018). Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. Nat. Commun. 9, 2419.

Blume, C., Felix, A., Shushakova, N., Gueler, F., Falk, C.S., Haller, H., and Schrader, J. (2012). Autoimmunity in CD73/Ecto-5′-nucleotidase deficient mice induces renal injury. PLoS ONE 7, e37100.

Bowman, E.P., Kuklin, N.A., Youngman, K.R., Lazarus, N.H., Kunkel, E.J., Pan, J., Greenberg, H.B., and Butcher, E.C. (2002). The intestinal chemokine thymus-expressed chemokine (CCL25) attracts IgA antibody-secreting cells. J. Exp. Med. 195, 269–275.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 34, 525–527.

Burger-van Paassen, N., Loonen, L.M.P., Witte-Bouma, J., Korteland-van Male, A.M., de Bruijn, A.C.J.M., van der Sluis, M., Lu, P., Van Goudoever, J.B., Wells, J.M., Dekker, J., et al. (2012). Mucin Muc2 deficiency and weaning influences the expression of the innate defense genes Reg3β, Reg3γ and angiogenin-4. PLoS ONE 7, e38798.

Bynoe, M.S., Waickman, A.T., Mahamed, D.A., Mueller, C., Mills, J.H., and Czopik, A. (2012). CD73 is critical for the resolution of murine colonic inflammation. J. Biomed. Biotechnol. 2012, 260983.

Chen, J., Suo, S., Tam, P.P., Han, J.J., Peng, G., and Jing, N. (2017). Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-seq. Nat. Protoc. 12, 566–580.

Chevalier, C., Stojanović, O., Colin, D.J., Suarez-Zamorano, N., Tarallo, V., Veyrat-Durebex, C., Rigo, D., Fabbiano, S., Stevanović, A., Hagemann, S., et al. (2015). Gut Microbiota Orchestrates Energy Homeostasis during Cold. Cell 163, 1360–1374.

Codeluppi, S., Borm, L.E., Zeisel, A., Manno, G.L., van Lunteren, J.A., Svensson, C.I., and Linnarsson, S. (2018). Spatial organization of the somatosensory cortex revealed by cyclic smFISH. bioRxiv. https://doi.org/10.1101/276097.

Combs, P.A., and Eisen, M.B. (2013). Sequencing mRNA from cryo-sliced Drosophila embryos to determine genome-wide spatial patterns of gene expression. PLoS ONE 8, e71820.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. 26, 1367–1372.

Crosetto, N., Bienko, M., and van Oudenaarden, A. (2015). Spatially resolved transcriptomics and beyond. Nat. Rev. Genet. 16, 57–66.

Crosnier, C., Stamataki, D., and Lewis, J. (2006). Organizing cell renewal in the intestine: stem cells, signals and combinatorial control. Nat. Rev. Genet. 7, 349–359.

Diag, A., Schilling, M., Klironomos, F., Ayoub, S., and Rajewsky, N. (2018). Regulation of spatial and temporal gene expression in an animal germline. bioRxiv. https://doi.org/10.1101/348425.

Earle, K.A., Billings, G., Sigal, M., Lichtman, J.S., Hansson, G.C., Elias, J.E., Amieva, M.R., Huang, K.C., and Sonnenburg, J.L. (2015). Quantitative Imaging of Gut Microbiota Spatial Organization. Cell Host Microbe 18, 478–488.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10, 48.

Elinav, E., Strowig, T., Kau, A.L., Henao-Mejia, J., Thaiss, C.A., Booth, C.J., Peaper, D.R., Bertin, J., Eisenbarth, S.C., Gordon, J.I., and Flavell, R.A. (2011). NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. Cell 145, 745–757.

George, M.D., Wehkamp, J., Kays, R.J., Leutenegger, C.M., Sabir, S., Grishina, I., Dandekar, S., and Bevins, C.L. (2008). In vivo gene expression profiling of human intestinal epithelial cells: analysis by laser microdissection of formalin fixed tissues. BMC Genomics 9, 209.

Glass, L.L., Calero-Nieto, F.J., Jawaid, W., Larraufie, P., Kay, R.G., Göttgens, B., Reimann, F., and Gribble, F.M. (2017). Single-cell RNA-sequencing reveals a distinct population of proglucagon-expressing cells specific to the mouse upper small intestine. Mol. Metab. 6, 1296–1303.

Grün, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., Clevers, H., and van Oudenaarden, A. (2015). Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature 525, 251–255.

Haber, A.L., Biton, M., Rogel, N., Herbst, R.H., Shekhar, K., Smillie, C., Burgin, G., Delorey, T.M., Howitt, M.R., Katz, Y., et al. (2017). A single-cell survey of the small intestinal epithelium. Nature 551, 333–339.

Halpern, K.B., Shenhav, R., Matcovitch-Natan, O., Tóth, B., Lemze, D., Golan, M., Massasa, E.E., Baydatch, S., Landen, S., Moor, A.E., et al. (2017). Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. Nature 542, 352–356.

Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., Saadatpour, A., Zhou, Z., Chen, H., Ye, F., et al. (2018). Mapping the Mouse Cell Atlas by Microwell-Seq. Cell 172, 1091–1107.e17.

Hecht, G., Bar-Nathan, C., Milite, G., Alon, I., Moshe, Y., Greenfeld, L., Dotsenko, N., Suez, J., Levy, M., Thaiss, C.A., et al. (2014). A simple cage-autonomous method for the maintenance of the barrier status of germ-free mice during experimentation. Lab. Anim. 48, 292–297.

Cell

Heindl, A., Nawaz, S., and Yuan, Y. (2015). Mapping spatial heterogeneity in the tumor microenvironment: a new era for digital pathology. Lab. Invest. *95*, 377–384.

Herring, C.A., Banerjee, A., McKinley, E.T., Simmons, A.J., Ping, J., Roland, J.T., Franklin, J.L., Liu, Q., Gerdes, M.J., Coffey, R.J., et al. (2018). Unsupervised Trajectory Analysis of Single-Cell RNA-Seq and Imaging Data Reveals Alternative Tuft Cell Origins in the Gut. Cell Syst. *6*, 37–51.e9.

Honda, K., and Littman, D.R. (2016). The microbiota in adaptive immune homeostasis and disease. Nature *535*, 75–84.

Huber-Ruano, I., Pinilla-Macua, I., Torres, G., Casado, F.J., and Pastor-Anglada, M. (2010). Link between high-affinity adenosine concentrative nucleoside transporter-2 (CNT2) and energy metabolism in intestinal and liver parenchymal cells. J. Cell. Physiol. *225*, 620–630.

Itzkovitz, S., Lyubimova, A., Blat, I.C., Maynard, M., van Es, J., Lees, J., Jacks, T., Clevers, H., and van Oudenaarden, A. (2011). Single-molecule transcript counting of stem-cell markers in the mouse intestine. Nat. Cell Biol. *14*, 106–114.

Junker, J.P., Noël, E.S., Guryev, V., Peterson, K.A., Shah, G., Huisken, J., McMahon, A.P., Berezikov, E., Bakkers, J., and van Oudenaarden, A. (2014). Genome-wide RNA Tomography in the zebrafish embryo. Cell *159*, 662–675.

Karaiskos, N., Wahle, P., Alles, J., Boltengagen, A., Ayoub, S., Kipar, C., Kocks, C., Rajewsky, N., and Zinzen, R.P. (2017). The *Drosophila* embryo at single-cell transcriptome resolution. Science *358*, 194–199.

Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C., and Teichmann, S.A. (2015). The technology and biology of single-cell RNA sequencing. Mol. Cell *58*, 610–620.

Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Yang, J.L., Ferrante, T.C., Terry, R., Jeanty, S.S.F., Li, C., Amamoto, R., et al. (2014). Highly multiplexed subcellular RNA sequencing in situ. Science *343*, 1360–1363.

Lein, E., Borm, L.E., and Linnarsson, S. (2017). The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. Science *358*, 64–69.

Lyubimova, A., Itzkovitz, S., Junker, J.P., Fan, Z.P., Wu, X., and van Oudenaarden, A. (2013). Single-molecule mRNA detection and counting in mammalian tissue. Nat. Protoc. *8*, 1743–1758.

Mabley, J.G., Pacher, P., Liaudet, L., Soriano, F.G., Haskó, G., Marton, A., Szabó, C., and Salzman, A.L. (2003). Inosine reduces inflammation and improves survival in a murine model of colitis. Am. J. Physiol. Gastrointest. Liver Physiol. *284*, G138–G144.

Mariadason, J.M., Nicholas, C., L'Italien, K.E., Zhuang, M., Smartt, H.J.M., Heerdt, B.G., Yang, W., Corner, G.A., Wilson, A.J., Klampfer, L., et al. (2005). Gene expression profiling of intestinal epithelial cell maturation along the crypt-villus axis. Gastroenterology *128*, 1081–1088.

Moffitt, J.R., Hao, J., Wang, G., Chen, K.H., Babcock, H.P., and Zhuang, X. (2016). High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. Proc. Natl. Acad. Sci. USA *113*, 11046–11051.

Moncada, R., Chiodin, M., Devlin, J.C., Baron, M., Hajdu, C.H., Simeone, D., and Yanai, I. (2018). Building a tumor atlas: integrating single-cell RNA-Seq data with spatial transcriptomics in pancreatic ductal adenocarcinoma. bioRxiv. https://doi.org/10.1101/254375.

Moor, A.E., and Itzkovitz, S. (2017). Spatial transcriptomics: paving the way for tissue-level systems biology. Curr. Opin. Biotechnol. *46*, 126–133.

Moor, A.E., Golan, M., Massasa, E.E., Lemze, D., Weizman, T., Shenhav, R., Baydatch, S., Mizrahi, O., Winkler, R., Golani, O., et al. (2017). Global mRNA polarization regulates translation efficiency in the intestinal epithelium. Science *357*, 1299–1303.

Nava, G.M., Friedrichsen, H.J., and Stappenbeck, T.S. (2011). Spatial organization of intestinal microbiota in the mouse ascending colon. ISME J. *5*, 627–638.

Nichterwitz, S., Chen, G., Aguila Benitez, J., Yilmaz, M., Storvall, H., Cao, M., Sandberg, R., Deng, Q., and Hedlund, E. (2016). Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling. Nat. Commun. *7*, 12139.

Nowarski, R., Jackson, R., Gagliani, N., de Zoete, M.R., Palm, N.W., Bailis, W., Low, J.S., Harman, C.C.D., Graham, M., Elinav, E., and Flavell, R.A. (2015). Epithelial IL-18 Equilibrium Controls Barrier Function in Colitis. Cell *163*, 1444–1456.

Okumura, R., Kurakawa, T., Nakano, T., Kayama, H., Kinoshita, M., Motooka, D., Gotoh, K., Kimura, T., Kamiyama, N., Kusu, T., et al. (2016). Lypd8 promotes the segregation of flagellated microbiota and colonic epithelia. Nature *532*, 117–121.

Peck, B.C.E., Mah, A.T., Pitman, W.A., Ding, S., Lund, P.K., and Sethupathy, P. (2017). Functional Transcriptomics in Diverse Intestinal Epithelial Cell Types Reveals Robust MicroRNA Sensitivity in Intestinal Stem Cells to Microbial Status. J. Biol. Chem. *292*, 2586–2600.

Peng, G., Suo, S., Chen, J., Chen, W., Liu, C., Yu, F., Wang, R., Chen, S., Sun, N., Cui, G., et al. (2016). Spatial Transcriptome for the Molecular Annotation of Lineage Fates and Cell Identity in Mid-gastrula Mouse Embryo. Dev. Cell *36*, 681–697.

Pimentel, H., Bray, N.L., Puente, S., Melsted, P., and Pachter, L. (2017). Differential analysis of RNA-seq incorporating quantification uncertainty. Nat. Methods *14*, 687–690.

Preibisch, S., Saalfeld, S., and Tomancak, P. (2009). Globally optimal stitching of tiled 3D microscopic image acquisitions. Bioinformatics *25*, 1463–1465.

The Tabula Muris Consortium, Quake, S.R., Wyss-Coray, T., and Darmanis, S. (2017). Single-Cell Transcriptomic characterization of 20 organs and tissues from mouse at single cell resolution creates a Tabula Muris. bioRxiv. https://doi.org/10.1101/237446.

Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., et al.; Human Cell Atlas Meeting Participants (2017). The Human Cell Atlas. eLife *6*, e27041.

Robson, S.C., Sévigny, J., and Zimmermann, H. (2006). The E-NTPDase family of ectonucleotidases: Structure function relationships and pathophysiological significance. Purinergic Signal. *2*, 409–430.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342.

Scialdone, A., Tanaka, Y., Jawaid, W., Moignard, V., Wilson, N.K., Macaulay, I.C., Marioni, J.C., and Göttgens, B. (2016). Resolving early mesoderm diversification through single-cell expression profiling. Nature *535*, 289–293.

Seyer, A., Cantiello, M., Bertrand-Michel, J., Roques, V., Nauze, M., Bézirard, V., Collet, X., Touboul, D., Brunelle, A., and Coméra, C. (2013). Lipidomic and spatio-temporal imaging of fat by mass spectrometry in mice duodenum during lipid digestion. PLoS ONE *8*, e58224.

Shah, S., Lubeck, E., Zhou, W., and Cai, L. (2016). In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus. Neuron *92*, 342–357.

Snippert, H.J., van der Flier, L.G., Sato, T., van Es, J.H., van den Born, M., Kroon-Veenboer, C., Barker, N., Klein, A.M., van Rheenen, J., Simons, B.D., and Clevers, H. (2010). Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. Cell *143*, 134–144.

Ståhl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science *353*, 78–82.

Stegmann, A., Hansen, M., Wang, Y., Larsen, J.B., Lund, L.R., Ritié, L., Nicholson, J.K., Quistorff, B., Simon-Assmann, P., Troelsen, J.T., and Olsen, J.

Cell

(2006). Metabolome, transcriptome, and bioinformatic cis-element analyses point to HNF-4 as a central regulator of gene expression during enterocyte differentiation. Physiol. Genomics 27, 141–155.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102, 15545–15550.

Tóth, B., Ben-Moshe, S., Gavish, A., Barkai, N., and Itzkovitz, S. (2017). Early commitment and robust differentiation in colonic crypts. Mol. Syst. Biol. 13, 902.

Tran, Q.T., Kennedy, L.H., Leon Carrion, S., Bodreddigari, S., Goodwin, S.B., Sutter, C.H., and Sutter, T.R. (2012). EGFR regulation of epidermal barrier function. Physiol. Genomics 44, 455–469.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. 32, 381–386.

Trautmann, A. (2009). Extracellular ATP in the immune system: more than just a "danger signal". Sci. Signal. 2, pe6–pe6.

Vaishnava, S., Yamamoto, M., Severson, K.M., Ruhn, K.A., Yu, X., Koren, O., Ley, R., Wakeland, E.K., and Hooper, L.V. (2011). The antibacterial lectin RegIIIgamma promotes the spatial segregation of microbiota and host in the intestine. Science 334, 255–258.

van der Flier, L.G., and Clevers, H. (2009). Stem cells, self-renewal, and differentiation in the intestinal epithelium. Annu. Rev. Physiol. 71, 241–260.

Yan, K.S., Janda, C.Y., Chang, J., Zheng, G.X.Y., Larkin, K.A., Luca, V.C., Chia, L.A., Mah, A.T., Han, A., Terry, J.M., et al. (2017a). Non-equivalence of Wnt and R-spondin ligands during Lgr5+ intestinal stem-cell self-renewal. Nature 545, 238–242.

Yan, K.S., Gevaert, O., Zheng, G.X.Y., Anchang, B., Probert, C.S., Larkin, K.A., Davies, P.S., Cheng, Z.F., Kaddis, J.S., Han, A., et al. (2017b). Intestinal Enteroendocrine Lineage Cells Possess Homeostatic and Injury-Inducible Stem Cell Activity. Cell Stem Cell 21, 78–90.e6.

Yissachar, N., Zhou, Y., Ung, L., Lai, N.Y., Mohan, J.F., Ehrlicher, A., Weitz, D.A., Kasper, D.L., Chiu, I.M., Mathis, D., and Benoist, C. (2017). An Intestinal Organ Culture System Uncovers a Role for the Nervous System in Microbe-Immune Crosstalk. Cell 168, 1135–1148.e12.

Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Häring, M., Braun, E., Borm, L.E., La Manno, G., et al. (2018). Molecular architecture of the mouse nervous system. Cell 174, 999–1014.e22.

Zheng, L., Kelly, C.J., and Colgan, S.P. (2015). Physiologic hypoxia and oxygen homeostasis in the healthy intestine. A Review in the Theme: Cellular Responses to Hypoxia. Am. J. Physiol. Cell Physiol. 309, C350–C360.

Zhu, Q., Shah, S., Dries, R., Cai, L., and Yuan, G.-C. (2018). Decomposing spatially dependent and cell type specific contributions to cellular heterogeneity. bioRxiv. https://doi.org/10.1101/275156.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
TruStain fcX (anti-mouse CD16/32) Antibody	Biolegend	Clone 93; Cat#101320; RRID: AB_1574975
APC anti-mouse CD73 antibody	Biolegend	Clone TY/11.8; Cat# 127210; RRID:AB_11218786
FITC Mouse Anti- E-Cadherin	BD Biosciences	Clone 36/E-Cadherin; Cat# 612131; RRID:AB_2076677
Alexa Fluor 647 Rat Anti-Mouse CD73	BD Biosciences	Clone TY/23; Cat# 561543; RRID: AB_10896329
FITC anti-mouse CD71 antibody	Biolegend	Clone RI7217; Cat# 113806; RRID:AB_313567
Sglt1 antibody	Abcam	Polyclonal; Cat# ab14686, RRID:AB_301411
Mouse Reg3B mAb	R&D systems	Clone 518630; Cat# MAB5110; RRID:AB_2178585
Cy3-AffiniPure Donkey Anti-Rat IgG (H+L)	Jackson ImmunoResearch Labs	Polyclonal; Cat# 712-165-153; RRID:AB_2340667
Rat IgG VisUCyte HRP Polymer Antibody	R&D systems	Polyclonal; Cat# VC005-025
Biological Samples		
Normal Horse Serum Blocking Solution	Vector laboratories	Cat# S-2000
Chemicals, Peptides, and Recombinant Proteins		
Paraformaldehyde solution 4% in PBS	Santa Cruz	Cat# sc-281692
O.C.T. Compound Cryostat Embedding Medium	Scigen	Cat# 4586
DAPI	Sigma-Aldrich	Cat# D9542
Nuclease free Water	Sigma-Aldrich	Cat# W4502
Critical Commercial Assays		
Bond Polymer Refine Detection	Leica biosystems	Cat# DS9800
Membrane Slide Nf 1.0 Pen	Carl Zeiss Microscopy	Cat# 415190-9081-000
Histogene LCM Frozen Section Staining Kit	ThermoFisher Scientific	Cat# KIT0401
AdhesiveCap 200 clear	Carl Zeiss Microscopy	Cat# 415190-9191-000
SMART-Seq v4 Ultra Low Input RNA Kit	Clontech	Cat# 634888
Nextera XT DNA Library Prep Kit	Illumina	Cat# FC-131-1024
NextSeq 500 Kits v2 (75 cycles)	Illumina	Cat# FC-404-2005
Deposited Data		
LCM raw sequencing data	this paper	GEO: GSE109413
LCM TPM values (table_A_LCM_TPM_values.tsv)	this paper	https://doi.org/10.5281/zenodo.1320734
scRNaseq UMI counts of used cells (table_B_scRNaseq_UMI_counts.tsv)	this paper	https://doi.org/10.5281/zenodo.1320734
scRNaseq tsne coordinates of used cells (table_C_scRNaseq_tsne_coordinates_zones.tsv)	this paper	https://doi.org/10.5281/zenodo.1320734
Zonation reconstruction (table_D_zonation_reconstruction.tsv)	this paper	https://doi.org/10.5281/zenodo.1320734
Lgr5-eGFPneg_1	Yan et al., 2017b	GEO: GSM2644349
Lgr5-eGFPneg_2	Yan et al., 2017b	GEO: GSM2644350
Germfree sequencing data	Peck et al., 2017	GEO: GSE81125
Antibiotic microbiome ablation sequencing data	Chevalier et al., 2015	GEO: GSE74157
Mouse cell atlas	Han et al., 2018	http://bis.zju.edu.cn/MCA/

Cell

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Organisms/Strains		
C57BL/6 inbred mice	Envigo	Strain: C57BL/6JOlaHsd
Lgr5-EGFP-IRES-CreERT2	Barker et al., 2007 Jackson Laboratory	Strain: 008875
R26R-Confetti	Snippert et al., 2010, Jackson Laboratory	Strain: 013731
Oligonucleotides		
smFISH probes, see Table S3	This paper	NA
Software and Algorithms		
Zonation reconstruction algorithm	this paper	https://github.com/aemoor/Code_spatial_reconstruction_enterocytes/
Bcl2fastq 2.17	Illumina	https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html RRID:SCR_015058
Kallisto 0.43.0	Bray et al., 2016	https://github.com/pachterlab/kallisto
Sleuth 0.28.1	Pimentel et al., 2017	https://github.com/pachterlab/sleuth
Seurat 2.1.0	Satija et al., 2015	https://github.com/satijalab/seurat
Monocle 2.8	Trapnell et al., 2014	https://bioconductor.org/packages/release/bioc/html/monocle.html
MaxQuant 1.6.0.16	Cox and Mann, 2008	http://www.coxdocs.org/doku.php?id=maxquant:start

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Shalev Itzkovitz (shalev.itzkovitz@weizmann.ac.il).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animal experiments

All mouse experiments were conducted in accordance with institutional guidelines and approved by the Institutional Animal Care and Use Committee of the Weizmann Institute of Science, Rehovot. Experiments were performed with 8-12 week old male C57BL/6 mice that were obtained from the Harlan laboratories or the WIS animal breeding center, Lgr5-CreERT2 mice (Barker et al., 2007) and R26R-Confetti mice (Snippert et al., 2010) were obtained from Jackson laboratory. Mice were housed in individually ventilated cages, were fed regular chow *ad libitum* and were exposed to phase-reversed circadian cycles. Germ-free C57BL/6 mice were housed in sterile isolators (Hecht et al., 2014).

METHOD DETAILS

Single molecule FISH

Mice were sacrificed and the proximal Jejunum was flushed with cold PBS, laterally cut, spread on dry whatman filter paper with the villi facing upward and cut into rectangles with a length of 1.5cm. Flat tissue on whatman paper was fixed in 4% paraformaldehyde (PFA, Santa Cruz Biotechnology, sc-281692) in PBS for 3 hours and subsequently agitated in 30% sucrose, 4% PFA in PBS overnight at 4°C. Fixed tissues were embedded in OCT (Scigen, 4586). We found that flat embedding of Jejunum pieces was important for preserving the intact morphology of full-length villi. 8um thick sections of fixed proximal Jejunum were sectioned onto poly L-lysine coated coverslips and used for smFISH staining. Probe libraries were designed using the Stellaris FISH Probe Designer Software (Biosearch Technologies, Petaluma, CA), see Table S3. The intestinal sections were hybridized with smFISH probe sets according to a previously published protocol (Itzkovitz et al., 2011). DAPI (Sigma-Aldrich, D9542) and a FITC-conjugated antibody against E-Cadherin (BD Biosciences, 612131) were used as nuclear and cell-membrane counterstains, respectively. SmFISH imaging was performed on a Nikon-Ti-E inverted fluorescence microscope with 60x or 100 × oil-immersion objectives and a Photometrics Pixis 1024 CCD camera using MetaMorph software as previously reported (Itzkovitz et al., 2011).

Probe libraries for messenger RNAs of interest were coupled to Cy5 and Alexa594, full-length villi were identified by the presence of co-stained villus tip maker gene expression (Nt5e or Ada coupled to TMR) on the same section. smFISH signal detection requires 60x or 100x magnifications, hence several fields of view were stitched together to create composite images that cover the whole crypt-villus unit. Stitching was performed with the fusion mode linear blending and default settings of the pairwise stitching plugin (Preibisch et al., 2009) in Fiji (Schindelin et al., 2012). Stitched villi were cropped rectangularly and underlaid with black background (which is visible in the stitched composite images in areas that lack data).

Immunofluorescence
8um thick sections of fixed proximal Jejunum were sectioned onto poly L-lysine coated coverslips and fixed with cold methanol for 20 minutes. Sections were briefly washed 3 times with PBST (1xPBS, 1% BSA and 0.1% Tween 20) and were further incubated 10 minutes in PBSTX (1X PBS, 0.25% Triton 100X and 1% BSA) at room temperature for permeabilization. After 3 PBST washes, sections were blocked with PBS supplemented with 0.1% Tween 20 and 5% Normal Horse Serum (Vector laboratories, S-2000) for 1h at room temperature, followed by an overnight incubation at 4°C with Alexa Fluor 647 rat anti mouse CD73 conjugated antibody (BD biosciences, 561543), 1:50 or a FITC-conjugated antibody against E-Cadherin (BD Biosciences, 612131), 1:100. Tfrc staining was performed with overnight incubation at 4°C with a rat anti-mouse Cd71 antibody (Biolegend, 113806), 1:50, followed by a 1h incubation at room temperature with a Cy3 AffiniPure Donkey Anti-Rat IgG (H+L) antibody (JacksonImmunoResearch, 712-165-153), 1:100). Sections were then washed again with PBST 3 times and were incubated with DAPI (1:200 in PBS) for 20 minutes. Imaging was carried out using the same setting as for the smFISH experiments.

Immunohistochemistry
Immunohistochemical staining was performed on 4μm sections using the Leica Bond III system (Leica Biosystems). Tissues were pretreated with epitope-retrieval (ER) solutions (Sglt1: 20 m ER solution 2, Leica Biosystems, AR9640, Reg3b: 20 m ER solution 1 BOND, Leica Biosystems, AR9961) followed by 30 minutes incubation with the following primary antibodies: Sglt1: Abcam, Ab14686, 1:100, Reg3b: RD systems, MAB5110, 1:100. The Leica Refine-HRP kit (Leica Biosystems, DS9800) used for detection of the Sglt1 antibody, Rat IgG VisUCyte HRP polymer antibody (RD systems, VC005-025) was used for detection of Reg3b antibody and counter-staining was performed with Hematoxylin.

LCM
Tissue blocks for microdissection were obtained from three 8 week-old male C57BL/6 mice. The proximal Jejunum was briefly washed in cold PBS and embedded and frozen in OCT without fixation. 8 μm thick sections were cut from the frozen block, mounted on polyethylene- naphthalate membrane-coated glass slides (Zeiss, 415190-9081-000), air-dried for 1 m at room temperature, washed in 70% ethanol (30 s), incubated in water (Sigma-Aldrich, W4502, 30 s), stained with HistoGene Staining Solution (ThermoFisher Scientific, KIT0401, 20 s), washed vigorously in water for a total of 30 s. The stained sections were dehydrated with subsequent 30 s incubations in 70%,95% and 100% ethanol and air-dried for 3 m before microdissection.

Tissue sections were microdissected on a UV laser-based PALM Microbeam (Zeiss). The system makes use of a pulsed UV laser that cuts the tissue at indicated marks with minimal damage to surrounding cells; the cutting was performed with the following parameters: PALM 20X lens, cut energy 48 (1-100), cut focus 65 (1-100). Tissue fragments were catapulted and collected in 0.2ml adhesive cap tubes (Zeiss, 415190-9191-000) with these settings: LPC energy 67 (1-100), LPC focus 67 (1-100). The capturing success was visually confirmed by focusing the PALM on the targeted adhesive cap after the collection session. 8-10 Villi above 500μm length selected for microdissection for each of three mice, their villus epithelium was divided into 5 segments of equal length and isolated. A total of 30'000-45'000 μm^2 of villus epithelium area was collected for each of the five villus zones per mouse.

Lineage tracing
Lineage tracing experiments were performed as previously described (Tóth et al., 2017). Briefly, Lgr5-EGFP-Ires-CreERT2 mice (Barker et al., 2007) were crossed with R26R-Confetti (Snippert et al., 2010) mice to track the fate of the progeny of individual Lgr5+ stem cells. The Cre enzyme was induced in 8- to 12-week-old male mice by a single intraperitoneal injection of 3 mg tamoxifen per mouse and mice were subsequently sacrificed for lineage tracing after 10 days.

RNA-seq
Library preparation for microdissected tissues was performed based on a previously published protocol (Moor et al., 2017) with minor modifications. Specifically, we resuspended microdissected fragments in 9.5 μL H2O and 1 μL of 10x reaction buffer of the SMART-Seq v4 Ultra Low Input RNA Kit (Clontech, 634888) in the adhesive cap of the collection tubes. Tissue lysis was achieved by incubation for 5 m at room temperature; the lysed samples were flash frozen until library preparation. The RNA was amplified with the SMART-Seq v4 kit according to the manufacturer's instructions and by using 15 PCR cycles for the cDNA amplification step. 1ng of the amplified cDNA was converted into sequencing library with the Nextera XT DNA Library kit (Illumina, FC-131-1024). The quality control of the resulting libraries was performed with an Agilent High Sensitivity D1000 ScreenTape System (Agilent, 5067- 5584). Libraries that passed quality control were loaded with a concentration of 1.8pM on 75 cycle high output flow cells (Illumina,

Cell

FC-404-2005) and sequenced on a NextSeq 500 (Illumina) with the following cycle distribution: 8bp index 1, 8bp index 2, 38bp read 1, 38bp read 2.

Flow cytometry

Enterocytes were isolated as previously described (Yan et al., 2017b). The cells were resuspended in cold FACS buffer (2mM EDTA, 0.5% BSA in PBS) in a concentration of 10^7 cells in 1ml. Next, cells were incubated with 20μl TruStain fcX (BioLegend, 101320) to block non-specific binding of immunoglobulin to the Fc receptors and stained with 6μl APC anti-mouse Cd73 antibody (BioLegend, 127210) for 30 m at 4 degrees. Last, after washing the cells (1000rpm, 5min, 4°C), cells were resuspended in FACS buffer (10^7 cells in 1ml) and DAPI was added (0.2μg/ml) to stain dead cells.

The samples were sorted with a SORP-FACSAriaII sorter (BD) using a 100 μm nozzle. The APC-intensity for Nt5e staining of viable enterocytes (gating as previously described (Yan et al., 2017b)) was used to isolate the populations of interest (Figure S6A). The differential populations of two mice were used for RNA-seq to validate the sorting strategy (Figures S6B and S6C) and of further six mice for proteomic analyses (Figure 6G). 10,000 enterocytes from each gate were sorted into FACS buffer for RNA-seq. After sorting they were spun down, resuspended in lysis buffer and frozen in −80°C until processing. 50,000 enteroytes from each population were collected into FACS buffer, and resuspended twice with PBS to wash away serum proteins. Pellets were flash frozen and sent to The Smoler Protein Research Center (Technion, Israel) for proteomic analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

smFISH quantification

We used two different methods to quantify the expression profiles of transcripts along the villus-axis from the smFISH images, depending on the abundance of the transcripts of interest. For low abundance genes, dots were counted using custom MATLAB program (Lyubimova et al., 2013) (MATLAB Release 2016a, The MathWorks, USA). The bottom, top and lateral epithelial borders of each quantified villus were manually segmented based on nuclear and cell-membrane counterstains. The epithelium was automatically further segmented into 20 units from bottom to top of the villus and mRNA density (number of mRNA per unit volume, for low abundance genes) or mRNA signal intensity (mean background-subtracted intensity in segmented unit, for high abundance genes) was computed along the villus-axis. For each transcript, we quantified at least 10 villi from 3 different mice.

Bulk RNA-seq analysis

Illumina output files were demultiplexed with bcl2fastq 2.17 (Illumina) and the resulting fastq files of mRNA-seq experiments were pseudoaligned with Kallisto 0.43.0 (Bray et al., 2016) to a transcriptome index of the GRCm38 release 90 (Ensembl), filtered to transcripts with a source entry of "ensembl_havana." The following flag was used for kallisto: -b 100. Sleuth 0.28.1 (Pimentel et al., 2017) running on R 3.3.2 was utilized to create a TPM table (Transcripts Per Million) for each sample, according to the Kallisto pseudoalignments (https://zenodo.org/record/1320734, table_A_LCM_TPM_values.tsv).

scRNaseq analysis

Two Lgr5-eGFP negative scRNaseq (Chromium, 10x Genomics) datasets were acquired from the NCBI GEO dataset browser (accessions GSM2644349 and GSM2644350 (Yan et al., 2017b)). scRNaseq analysis was performed with Seurat 2.1.0 (Satija et al., 2015) in R 3.3.2. Cells were filtered based on mitochondrial gene content, unique molecular identifier (UMI) counts were log-normalized according to default Seurat settings. Variable genes were identified (FindVariableGenes, parameters: x.low.cutoff = 0.0125, x.high.cutoff = 3, y.cutoff = 0.5) and the following three sources of variation were regressed out: UMI number, biological replicate number and mitochondrial gene content. Principle Component Analysis was performed on the expression levels of the detected variable genes. The first 10 principal components were included for further downstream analyses based on visual inspection of Seurat's PCElbowPlot. To identify enterocyte cells, all cells were clustered based on the principal component analysis with the following granularity parameters: dims.use = 1:10, resolution = 1.3. Mature enterocytes and transient amplifying clusters were identified based on Alpi and Mki67 expression, respectively. A few mis-assigned goblet, tuft and Paneth cells were removed by filtering based on expression of the following marker genes: Muc2 and Hepacam2 (Goblet), Dclk1 (Tuft) and Lyz1 (Paneth). All results in the paper are insensitive to the clustering method and parameters (data not shown). Non-linear dimensional reduction (tSNE) was used to visualize the previously computed clusters. Raw UMI counts of the resulting 1383 enterocytes and transient amplifying cells were exported and utilized for zonation reconstruction algorithm (https://zenodo.org/record/1320734 under the files table_B_scRNaseq_UMI_counts.tsv and table_C_scRNaseq_tsne_coordinates_zones.tsv).

Zonation reconstruction algorithm

To reconstruct the zonation profiles from the scRNaseq data we used the summed expression of the landmark gene (LM) panels to infer the locations of each sequenced enterocyte along the villus spatial axis. Each cell i was assigned a spatial coordinate $0 \leq x_i \leq 1$, which correlated with its location along the villus axis and was computed as the ratio of the summed expression of the top landmark genes (tLM), and the summed expression of the bottom (bLM) and top LM genes to yield

$$x_i = \frac{\sum_{g \in tLM} E_{g,i}}{\left(\sum_{g \in bLM} E_{g,i} + \sum_{g \in tLM} E_{g,i} \right)}$$

(Equation 1)

Where $E_{g,i}$ is the expression of gene g in cell i in units of fraction of total cellular UMIs. To map x_i values to spatial coordinates along the villus axis, we used the same equation to calculate the coordinate x_{LCMi} for each of the five Laser-capture-microdissected villus zones LCM_i. We assigned each cell to one of 6 villus zones, $V_1..V_6$ as follows: cells with $x_i < x_{LCM1}$ were assigned to V_1, cells with $x_{LCM1} \leq x_i < x_{LCM2}$ were assigned to V_2, cells with $x_{LCM2} \leq x_i < x_{LCM3}$ were assigned to V_3, cells with $x_{LCM3} \leq x_i < x_{LCM4}$ were assigned to V_4, cells with $x_{LCM4} \leq x_i < x_{LCM5}$ were assigned to V_5 and cells with $x_{LCM5} \leq x_i$ were assigned to V_6. For each gene and zone we calculated the means and standard errors of the means (SEM) of the expression of all genes over the cells assigned to the respective zone. Crypt gene expression was computed by the mean and SEM over the expression of single cells assigned by Seurat to the two transient amplifying clusters (https://zenodo.org/record/1320734, table_D_zonation_reconstruction.tsv).

To compute zonation significance, we used a non-parametric permutation test. We considered, as a summary statistic, the profile's dynamic range, defined as the difference between the maximum and minimum values of the mean-normalized profile along V_1-V_6 as a summary statistic. For each gene, we compared the dynamic range to those obtained for 1,000 datasets in which the cells' assigned zones were randomly reshuffled. We included genes with maximal zonation larger than $5*10^{-6}$ when computing the fraction of zonated genes. For each gene, we calculated Z-scores for the observed dynamic range compared to the permuted dynamic ranges and used the normal distribution to obtain p values. This was done to increase the p value resolution beyond the 1/1000 limit imposed by the number of permutations (all significantly zonated genes were also significant when computing a numerical p value instead). We used Storey's method to compute q-values (https://zenodo.org/record/1320734, table_D_zonation_reconstruction.tsv).

Pseudotime analysis
Pseudotime analysis was performed with Monocle 2.8 (Trapnell et al., 2014). The filtered single cell expression matrix was imported from its Seurat object with the importCDS function. Size factors were calculated with the estimateSizeFactors function and dispersion was estimated with estimateDispersions. Genes that were used for cell ordering were selected by their expression in at least 5% of all cells. Informative principle components for this filtered dataset were assessed with the plot_pc_variance_explained function and the first five principal components were chosen for downstream analysis. The dimensionality of the data was reduced with reduceDimension according to the DDRTree method and cells were ordered along the pseudo time trajectory with orderCells.

Clustering and gene ontology enrichment
Gene ontology (GO) terms were obtained from Ensembl (GRCm38 release 90). All GO annotations that contained more than three genes with highly expressed zonated enterocyte genes (UMI fraction above 10^{-4}, 2118 genes) were chosen for enrichment analysis. The expression profiles along the villus-axis of these genes were normalized to their maximum expression. The normalized profiles were partitioned into five mutually exclusive clusters with k-Means clustering using MATLAB by using correlation as distance measure (Table S1). The significance of enrichment of the selected GO terms in each of these five clusters was assessed with the hypergeometric test. Storey's method was used to compute q-values (Table S2).

Mass spectrometry for proteomics
All 18 samples (3 populations of 6 mice) were digested with trypsin and analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) on a Q Exactive plus mass spectrometer (Thermo Fischer Scientific). The data were analyzed with MaxQuant 1.6.0.16 (Cox and Mann, 2008) by using the mouse uniprot database as reference. Further analysis was done using the perseus software. The identifications were filtered for proteins identified with FDR < 0.01 with at least 2 MSMS counts and 2 peptides in the project in at least 3 samples in one of the groups (2992 proteins, Table S4). We subtracted the minimal non-zero log2LFQ value from all log2LFQ values and then transformed back to lineage space. We next averaged these background-subtracted LFQ for the 6 mice from each of the three populations. We next filtered out genes that had a maximal zonation smaller than $5*10^{-6}$ of the cellular RNA in any of the zones and normalized the remaining 2,818 proteins to the summed expression (values presented in Figure 6G).

DATA AND SOFTWARE AVAILABILITY

The generated sequencing data have been deposited in the GenBank GEO database (https://www.ncbi.nlm.nih.gov/geo/) under accession code GSE109413. The villus reconstruction algorithm has been desposited in Github (https://github.com/aemoor/Code_spatial_reconstruction_enterocytes/) the corresponding raw and intermediary input files have been deposited in Zenodo (https://zenodo.org/record/1320734. Seurat export files of the single cells that were used in this study (UMI counts, cell barcodes, tSNE coordinates and reconstructed zones) were deposited in the dataset at (https://zenodo.org/record/1320734) under the files table_B_scRNaseq_UMI_counts.tsv and table_C_scRNaseq_tsne_coordinates_zones.tsv.
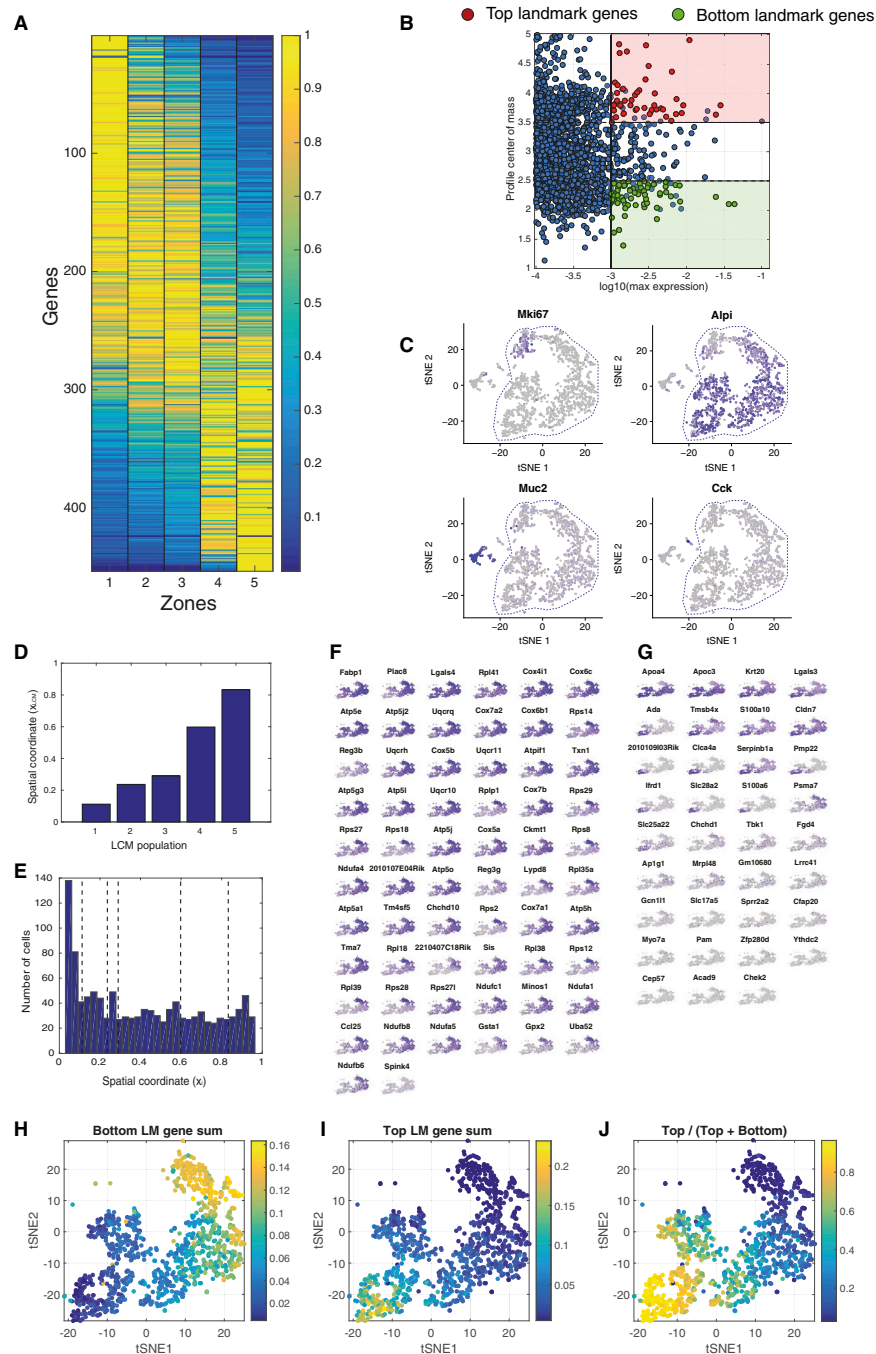
Supplemental Figures

Figure S1. RNA Sequencing of Microdissected Villus Epithelium Fragments and Spatial Reconstruction of Zonation, Related to Figure 2

(A) Gene expression heatmap of bulk RNA sequencing of villus quintiles that were isolated by laser capture microdissection (LCM). Gene expression profiles are normalized to their maximum and sorted according to their center of mass.

(B) Selection of bottom and top landmark genes. Each dot represents a detected gene (TPM fraction above $5*10^{-4}$). Red dots indicate genes that were selected as top landmark genes (Figure 2C). These were selected based on high expression (TPM fraction above 10^{-3}), profile center of mass larger than 3.5 and maximum expression in zone 5. Green dots indicate genes that were selected as bottom landmark genes (Figure 2B). These were selected based on high expression (TPM fraction above 10^{-3}), profile center of mass smaller than 2.5 and maximum expression in zone 1.

(C) tSNE plots of intestinal marker gene expression in single Lgr5-eGFP negative cells (Yan et al., 2017b). Mki67 is expressed in transient amplifying cells in the crypt, Alpi in villus enterocytes, Muc2 in goblet cells and Cck in enteroendocrine cells. Depicted analyses are based on raw data from NCBI GEO datasets GSM2644349 and GSM2644350 (Yan et al., 2017b). Dashed lines mark the cells used in our reconstruction.

(legend continued on next page)

(D) Inferred spatial coordinate (x_{LCMi}, Methods: Zonation reconstruction algorithm) of the five microdissected fragments. These values serve as cutoffs to classify the continuous spatial coordinates of the single sequenced enterocytes into one of 6 villus zones.

(E) Histogram of all single cells and their inferred spatial coordinate (x_i). Vertical dashed lines indicate the spatial coordinate values of the five microdissected fragments (D).

(F and G) tSNE plots of bottom landmark genes (F) and top landmark genes (G) that were in expressed in 10 or more enterocytes. (H-J) tSNE plots showing the mature enterocytes that were subjected to spatial reconstruction.

(H) Color indicates the summed expression of the bottom landmark genes (Figure 2B).

(I) Color indicates the summed expression of the top landmark genes (Figure 2C).

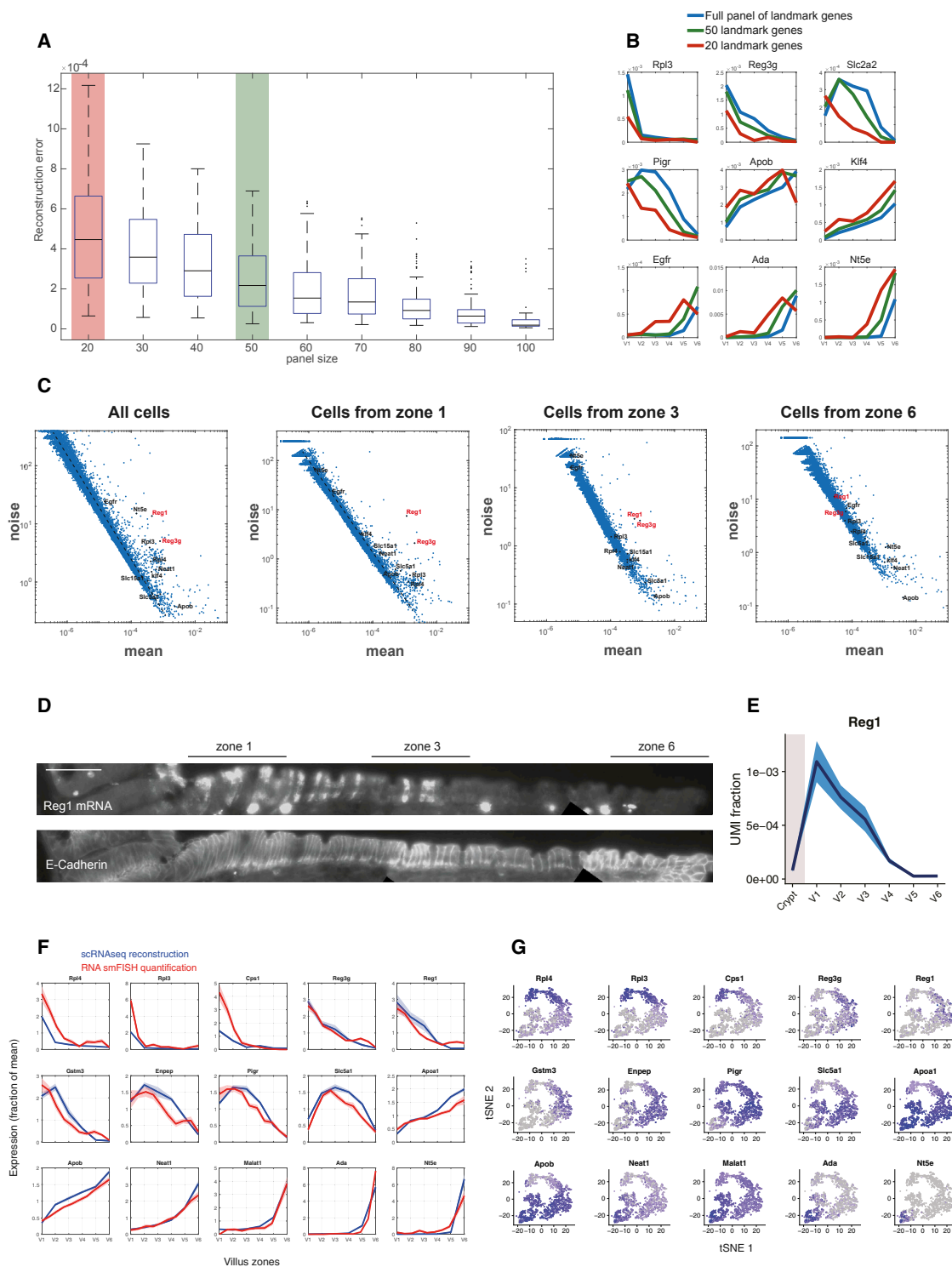(J) Color indicates the ratio of top to (top + bottom) sums.

Figure S2. Effects of Panel Size and Intra-zonal Variability on Zonation Reconstruction, Related to Figure 2

(A) Zonation reconstruction error as a function of landmark gene panel size. For each panel size that ranged between 20 and 100 landmark genes, 100 landmark gene sets were randomly sampled from the complete LCM landmark gene panel, preserving the proportion of low villus and high villus landmark gene sets. Zonation was reconstructed as described in Methods and the mean squared error between the zonation profiles reconstructed with the sub-sampled panel and those reconstructed with the full panel was computed for all genes with maximal zonation expression above 10^{-5} of cellular UMI. Horizontal lines are medians, boxes are 25-75 percentiles. Red and green marks highlight the panels that were used in (B).

(legend continued on next page)

(B) Examples for reconstructed zonation profiles based on the sub-sampled landmark gene panels.

(C) Noise-mean scatterplots for enterocyte genes, computed over all cells (left) and over cells from specific zones (lower villus zone 1, mid-villus zone 3 and villus top zone 6). Zonated genes such as Nt5e, Rpl3, Neat1 and Klf4 are highly variables when considering all cells but become more uniform when stratifying for the cells' villus zone. The zonated genes Reg1 and Reg3g remain highly variable among cells even when stratifying for zones.

(D and E) Reg1 zonation is faithfully reconstructed even though it is locally variable among single cells in each of the villi zones. (D) smFISH of Reg1 mRNA expression in whole villus (top) and E-Cadherin antibody staining that labels cell boundaries (bottom). (E) scRNaseq-inferred zonation profile of Reg1. Dark blue line: mean expression, light blue area: SEM.

(F) Validation of the reconstructed zonation profiles using smFISH. Dark blue line depicts scRNaseq mean expression level, light blue area denotes its SEM (standard error of the mean). Dark red line depicts smFISH mean expression level, light red area denotes its SEM. All profiles are normalized by their means across zones. SmFISH profiles based on measurements from at least 10 villi from 3 different mice.

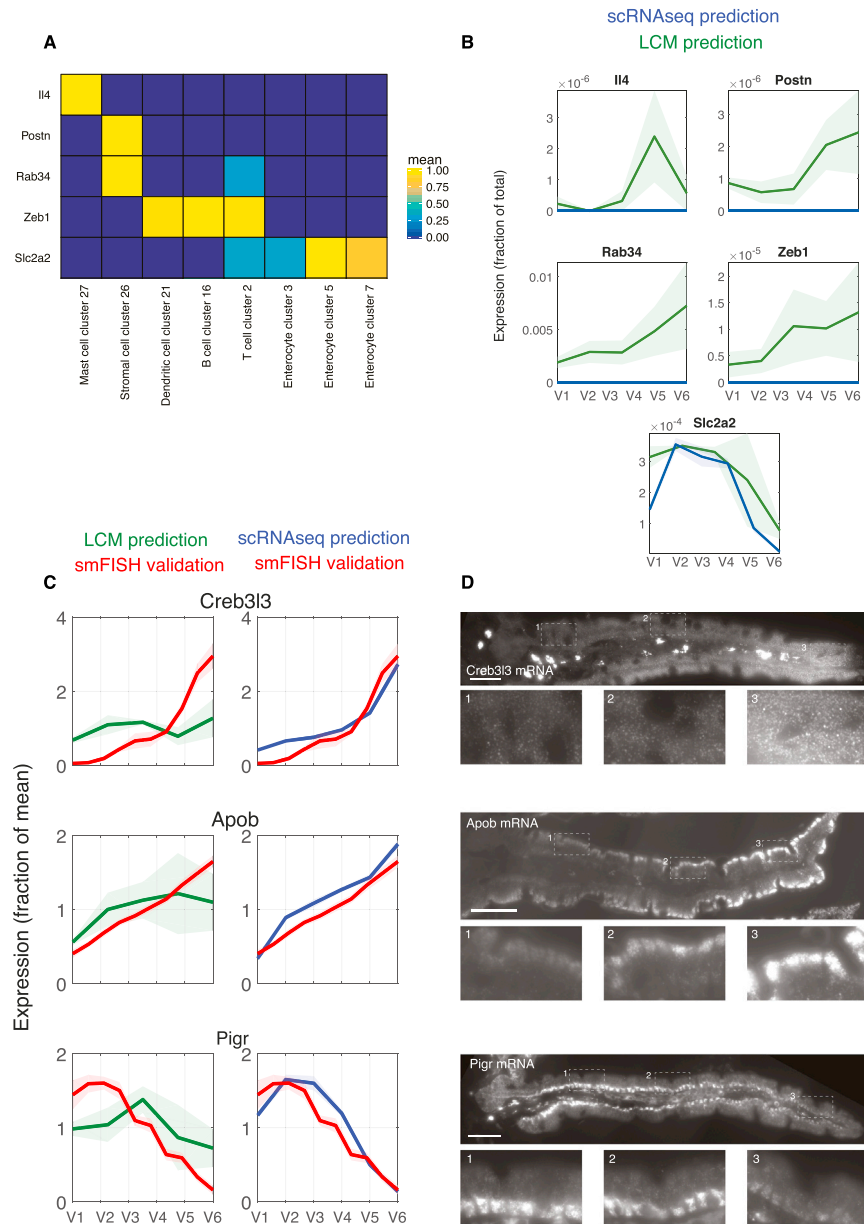(G) tSNE plots of genes that are featured in (F).

A

B

scRNAseq prediction
LCM prediction

C

LCM prediction   scRNAseq prediction
smFISH validation   smFISH validation

D

Figure S3. scRNA-Seq Predictions of Zonation Are More Accurate Than LCM-Based Predictions, Related to Figure 2

(A) Heatmap of non-epithelial gene expression (Il4, Postn, Rab34 and Zeb1) in cell types of relevance. Analysis based on Han et al., 2018 and extracted from http://bis.zju.edu.cn/MCA/. Expression levels are normalized to their maximal value across clusters.

(B) False positive prediction of epithelial zonation by LCM. Dark blue line depicts scRNaseq mean expression level, light blue area denotes its SEM (standard error of the mean). Dark green line depicts LCM mean expression level, light green area denotes its SEM. The epithelial expression of Slc2a2 is shown as positive control where scRNaseq-reconstruction overlaps LCM-reconstruction.

(C) Examples of epithelial genes for which LCM-based zonation (left) was less accurate than the scRNaseq-based reconstruction (right). Left: Dark green line depicts LCM mean expression level, light green area denotes its SEM Right: Dark blue line depicts scRNaseq mean expression level, light blue area denotes its SEM. In both left and right columns dark red lines depict smFISH mean expression levels, light red area denotes its SEM.

(D) smFISH of Creb3l3, Apob and Pigr RNA expression in whole villus (overview) and in the bottom (1), middle (2) and top (3) parts of the villus (inserts) demonstrate the increased accuracy of the scRNaseq-based reconstruction. All scale bars are 50μm.
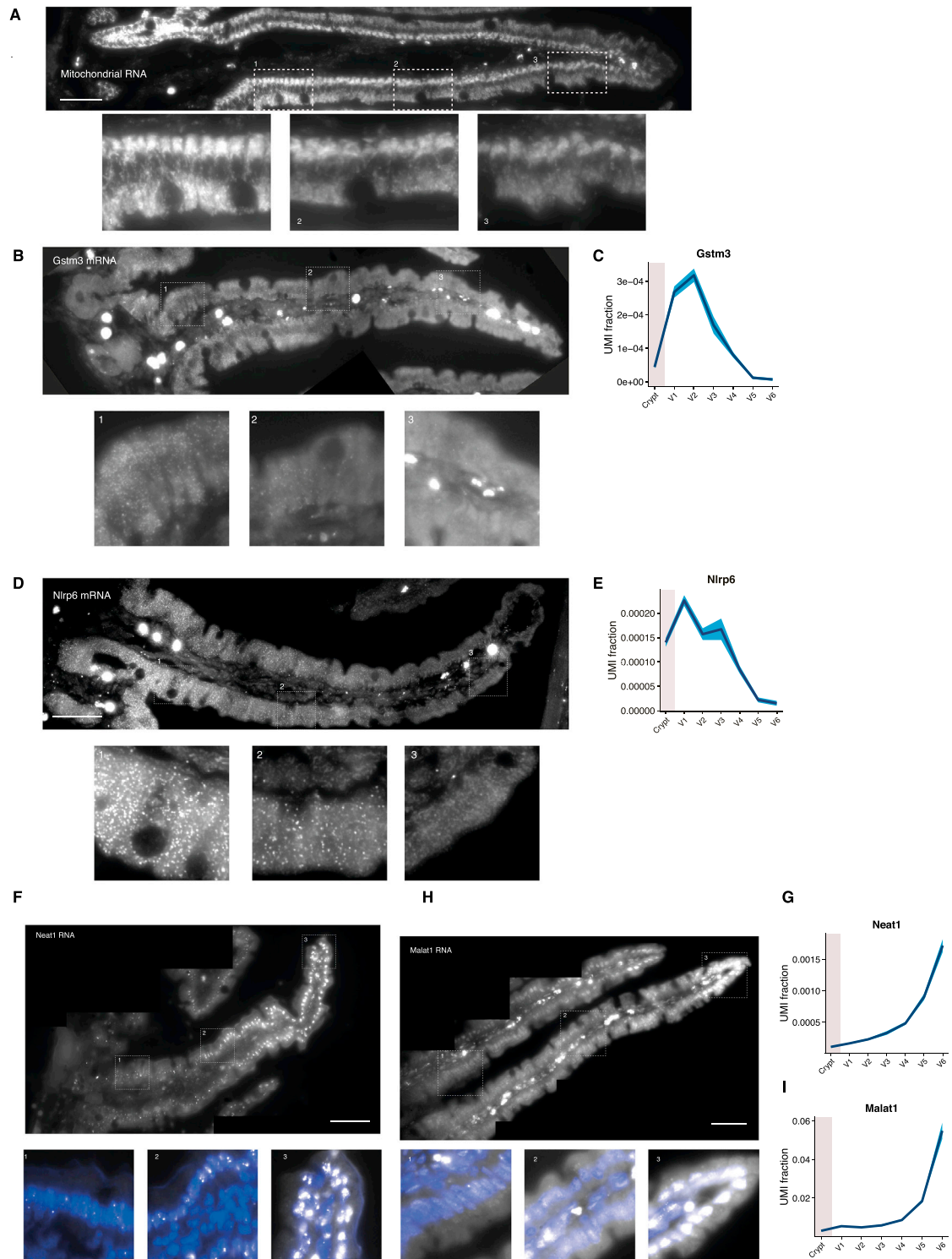
Figure S4. smFISH Validation of Zonated Expression of Representative Clusters, Related to Figure 3

(A) smFISH of mitochondrial light-strand RNA. Enlarged inserts show the gradual decrease in mitochondrial light-strand RNA expression from villus bottom to villus top.

(B) smFISH of Gstm3 mRNA expression in whole villus (overview) and in the bottom (1), middle (2) and top (3) parts of the villus (inserts).

(C) scRNaseq-inferred zonation profile of the phase II xenobiotic metabolism enzyme Gstm3. Dark blue line: mean expression, light blue area: SEM.

(D) smFISH of inflammasome component Nlrp6 mRNA expression in whole villus (overview) and in the bottom (1), middle (2) and top (3) parts of the villus (inserts).

(legend continued on next page)

(E) scRNaseq-inferred zonation profile of Nlrp6. Dark blue line: mean expression, light blue area: SEM.

(F) smFISH of Neat1 RNA expression in whole villus (overview) and in the bottom (1), middle (2) and top (3) parts of the villus (inserts).

(G) scRNaseq-inferred zonation profile of Neat1. Dark blue line: mean expression, light blue area: SEM.

(H) smFISH of Malat1 RNA expression in whole villus (overview) and in the bottom, middle and top part of the villus (inserts). Blue in inserts is DAPI nuclear staining.

(I) scRNaseq-inferred zonation profile of Malat1. Dark blue line: mean expression, light blue area: SEM. All scale bars are 50μm.

A

Cdh1

B

Cdh1 mRNA

C

Nt5e

D
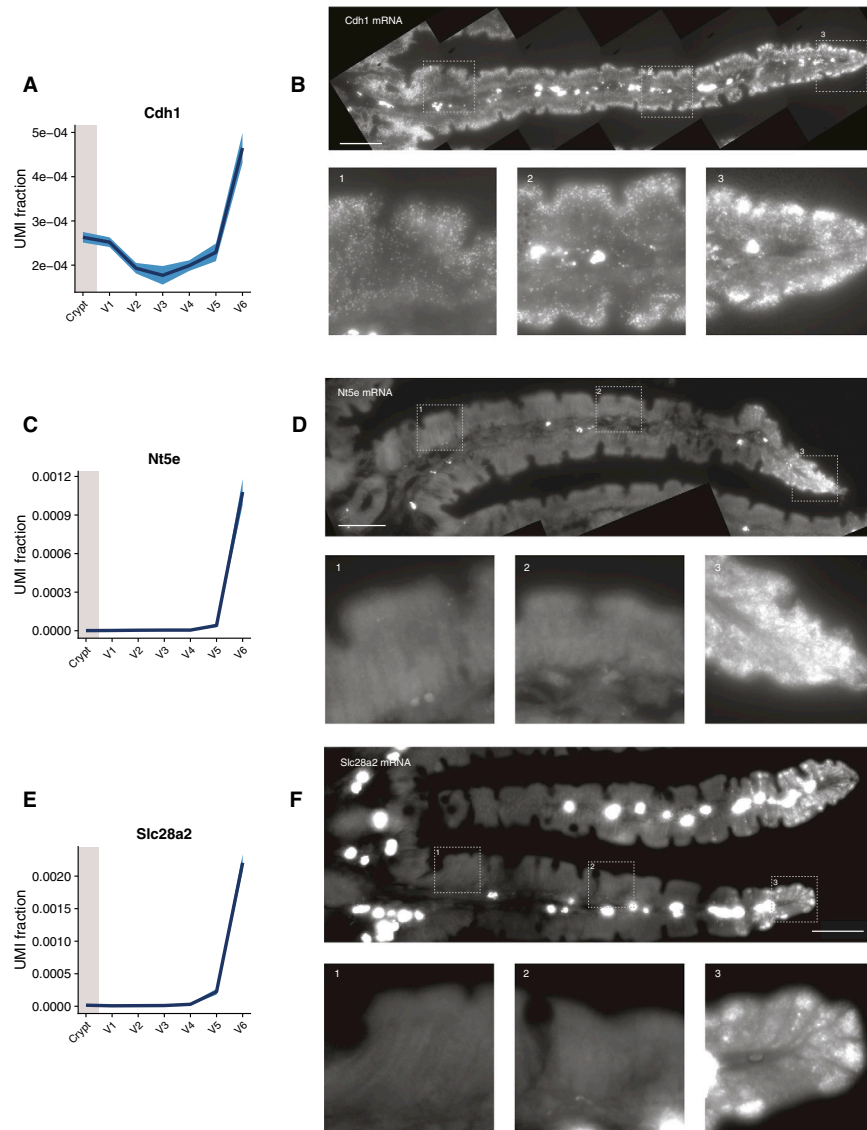
Nt5e mRNA

E

Slc28a2

F

Slc28a2 mRNA

Figure S5. Zonation Profiles of Tip-Enriched mRNAs, Related to Figure 6

(A) scRNaseq-inferred zonation profile of Cdh1 mRNA. Dark blue line: mean expression, light blue area: SEM.

(B) smFISH of Cdh1 mRNA expression in whole villus (overview) and in the bottom (1), middle (2) and top (3) parts of the villus (inserts).

(C) scRNaseq-inferred zonation profile of Nt5e. Dark blue line: mean expression, light blue area: SEM.

(D) smFISH of Nt5e mRNA expression in whole villus (overview) and in the bottom (1), middle (2) and top (3) parts of the villus (inserts).

(E) scRNaseq-inferred zonation profiles of Slc28a2. Dark blue line: mean expression, light blue area: SEM.

(F) smFISH of Slc28a2 mRNA expression in whole villus (overview) and in the bottom (1), middle (2) and top (3) parts of the villus (inserts). All scale bars are 50μm.
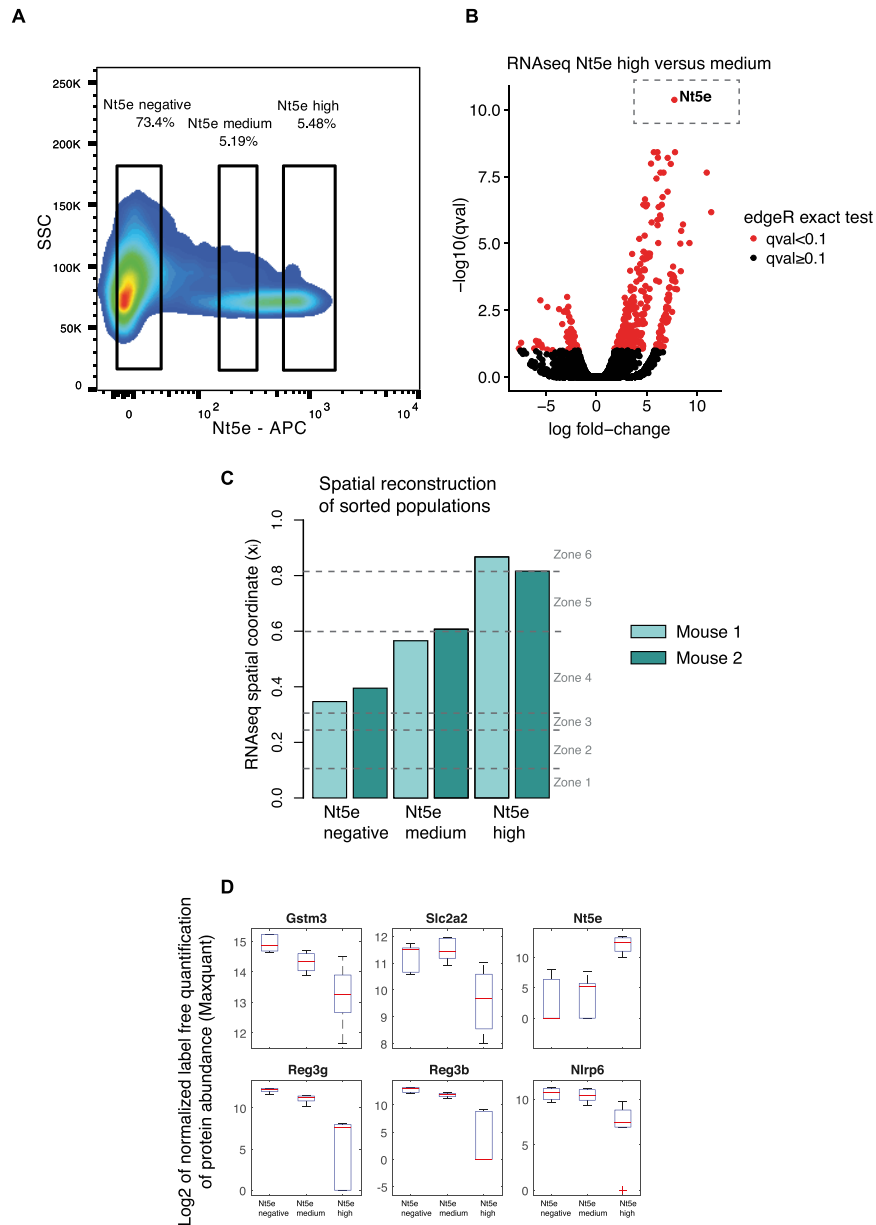
Figure S6. Spatial Sorting According to Nt5e Expression Levels, Related to Figure 6

(A) Gating strategy for isolating cells according to their Nt5e expression level.

(B) Differential gene expression analysis that compares RNaseq data of Nt5e high versus medium cells.

(C) inferred spatial coordinates of the RNaseq data of sorted populations according to the zonation reconstruction algorithm (STAR Methods; Figure S1).

(D) Boxplot of protein abundances (log2 of normalized label free quantification values after subtracting 18 (reported minimum)). Data is based on the mass spectromic analyses of three sorted populations with differential Nt5e abundance of six mice (18 samples).
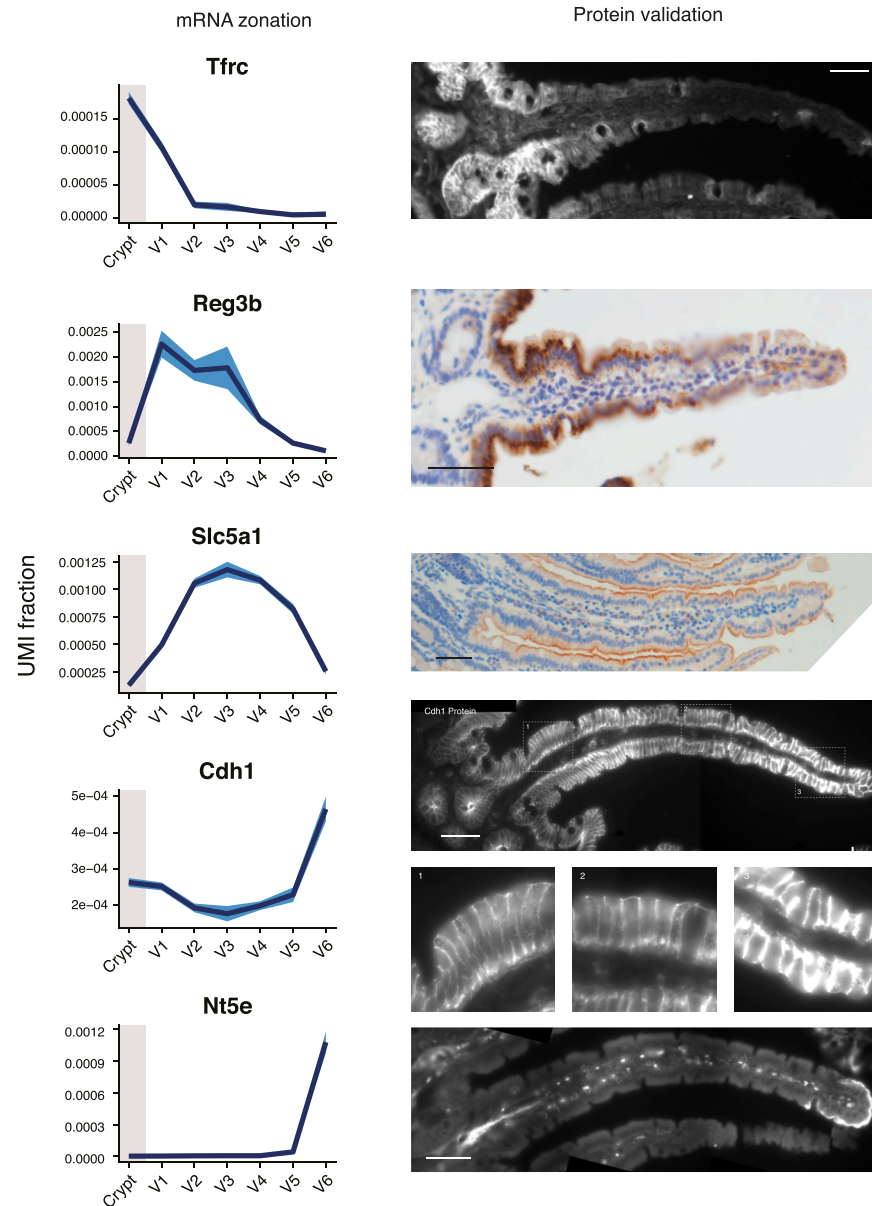
Figure S7. Protein Zonation Is Correlated with mRNA Zonation, Related to Figure 6
Left: scRNaseq-inferred zonation profile of Tfrc, Reg3b, Slc5a1, Cdh1 and Nt5e mRNA. Dark blue line: mean expression, light blue area: SEM. Right: antibody staining of Tfrc, Reg3b, Sglt1 (encoded by Slc5a1), Cdh1 and Nt5e. All scale bars are 50μm. Protein validation for Tfrc, Cdh1 and Nt5e was performed using immunofluorescense, validation for Reg3b and Sglt1 was performed using immunohistochemistry.