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

Graphical Abstract

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

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.

Moor et al., 2018, Cell 175, 1–12
November 1, 2018 © 2018 Elsevier Inc.
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, and the low sensitivity of microarray-based transcriptomics precluded the detection of spatial expression 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 (Kołodziejczyk 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; Halpem 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
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.

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
biosynthesis and cell adhesion processes as well as the long non-coding RNA (lncRNA) markers of paraspeckles (Neat1; Figures S4H and S4I), and speckles (Malat1; Figures S4F and S4G), 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 the crypt that might function as a gatekeeper for the crypt stem cell niche to minimize its exposure to pathogenic microbes.

**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,
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 necessarily entail a decline in the protein translation, GO:0006412
mitochondrion, GO:0005739
intestinal absorption, GO:0050892
brush border, GO:0005903
cell adhesion, GO:0007155
lipoprotein biosynthetic process, GO:0042158

Please cite this article in press as: Moor et al., Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation along the Intestinal Villus Axis, Cell (2018), https://doi.org/10.1016/j.cell.2018.08.063
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 intracellular 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
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.
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; Codelluppi 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
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 Litman, 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; Kariakos 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.

ACKNOWLEDGMENTS
We thank the Lorry Lokey Animal Facility (Weizmann Institute of Science, Rehovot, Israel), Gavish Research Services (Beit Elzaria, Israel), and the Smoler Protein Research Center (Technion, Haifa, Israel) for help with experimental procedures. A.E.M. is supported by the Swiss National Science Foundation (grant 158999) and the EMBO Long-Term Fellowship Program (ALTF 306-2016). S.I. is supported by the Henry Chanoch Krenter Institute for Biomedical Imaging and Genomics, the Leir Charitable Foundations, the Richard

Please cite this article in press as: Moor et al., Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation along the Intestinal Villus Axis, Cell (2018), https://doi.org/10.1016/j.cell.2018.08.063
Jakubskind Laboratory of Systems Biology, the Cymeryn-Jakubskind Prize, the Lord Sieff of Brompton Memorial Fund, the Wolfson Foundation and Wolfson Family Charitable Trust, the I-CORE Program of the Planning and Budgeting Committee and the Israel Science Foundation (grants 1902/12 and 1796/12), Israel Science Foundation grant 1486/16, the EMBO Young Investigator Program, the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement 335122, the Bert L. and N. Kuggie Vallee Foundation, and the Howard Hughes Medical Institute (HHMI) international research scholar award. S.I. is the incumbent of the Philip Harris and Gerald Ronson Career Development Chair.

**AUTHOR CONTRIBUTIONS**

A.E.M. and S.I. conceived the study. A.E.M. designed and performed most of the experiments. A.E.M., Y.H., E.E.M., S.B.-M., and M.R. performed single-molecule FISH experiments. R.E. performed immunofluorescence experiments. A.E.M. performed LCM experiments. A.E.M. and K.B.H. performed RNA-seq experiments. S.I. and A.E.M. performed data analyses. S.I. and A.E.M. wrote the manuscript. S.I. supervised the study. All authors discussed the results and commented on the manuscript.

**DECLARATION OF INTERESTS**

The authors declare no competing interests.

Received: May 28, 2018  
Revised: July 29, 2018  
Accepted: August 29, 2018  
Published: September 27, 2018

**REFERENCES**


# STAR METHODS

## KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antibodies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TruStain fcX (anti-mouse CD16/32) Antibody</td>
<td>Biolegend</td>
<td>Clone 93; Cat#101320; RRID: AB_1574975</td>
</tr>
<tr>
<td>APC anti-mouse CD73 antibody</td>
<td>Biolegend</td>
<td>Clone TY/11.8; Cat# 127210; RRID:AB_11218786</td>
</tr>
<tr>
<td>FITC Mouse Anti- E-Cadherin</td>
<td>BD Biosciences</td>
<td>Clone 36/E-Cadherin; Cat# 612131; RRID:AB_2076677</td>
</tr>
<tr>
<td>Alexa Fluor 647 Rat Anti-Mouse CD73</td>
<td>BD Biosciences</td>
<td>Clone TY/11.8; Cat# 127210; RRID:AB_11218786</td>
</tr>
<tr>
<td>FITC anti-mouse CD71 antibody</td>
<td>Biolegend</td>
<td>Clone R7217; Cat# 113806; RRID:AB_313567</td>
</tr>
<tr>
<td>Sglt1 antibody</td>
<td>Abcam</td>
<td>Polyclonal; Cat# ab14686, RRID:AB_301411</td>
</tr>
<tr>
<td>Mouse Reg3B mAb</td>
<td>R&amp;D systems</td>
<td>Clone 518630; Cat# MAB5110; RRID:AB_2178585</td>
</tr>
<tr>
<td>Cy3-AffiniPure Donkey Anti-Rat IgG (H+L)</td>
<td>Jackson ImmunoResearch Labs</td>
<td>Polyclonal; Cat# 712-165-153; RRID:AB_2340667</td>
</tr>
<tr>
<td>Rat IgG VisUCyte HRP Polymer Antibody</td>
<td>R&amp;D systems</td>
<td>Polyclonal; Cat# VC005-025</td>
</tr>
<tr>
<td><strong>Biological Samples</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal Horse Serum Blocking Solution</td>
<td>Vector laboratories</td>
<td>Cat# S-2000</td>
</tr>
<tr>
<td><strong>Chemicals, Peptides, and Recombinant Proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paraformaldehyde solution 4% in PBS</td>
<td>Santa Cruz</td>
<td>Cat# sc-281692</td>
</tr>
<tr>
<td>O.C.T. Compound Cryostat Embedding Medium</td>
<td>Scigen</td>
<td>Cat# 4586</td>
</tr>
<tr>
<td>DAPI</td>
<td>Sigma-Aldrich</td>
<td>Cat# D9542</td>
</tr>
<tr>
<td>Nuclease free Water</td>
<td>Sigma-Aldrich</td>
<td>Cat# W4502</td>
</tr>
<tr>
<td><strong>Critical Commercial Assays</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond Polymer Refine Detection</td>
<td>Leica biosystems</td>
<td>Cat# DS9800</td>
</tr>
<tr>
<td>Membrane Slide NF 1.0 Pen</td>
<td>Carl Zeiss Microscopy</td>
<td>Cat# 415190-9081-000</td>
</tr>
<tr>
<td>Histogene LCM Frozen Section Staining Kit</td>
<td>ThermoFisher Scientific</td>
<td>Cat# KIT0401</td>
</tr>
<tr>
<td>AdhesiveCap 200 clear</td>
<td>Carl Zeiss Microscopy</td>
<td>Cat# 415190-9191-000</td>
</tr>
<tr>
<td>SMART-Seq v4 Ultra Low Input RNA Kit</td>
<td>Clontech</td>
<td>Cat# 634888</td>
</tr>
<tr>
<td>Nextera XT DNA Library Prep Kit</td>
<td>Illumina</td>
<td>Cat# FC-131-1024</td>
</tr>
<tr>
<td>NextSeq 500 Kits v2 (75 cycles)</td>
<td>Illumina</td>
<td>Cat# FC-404-2005</td>
</tr>
<tr>
<td><strong>Deposited Data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LCM raw sequencing data</td>
<td>this paper</td>
<td>GEO: GSE109413</td>
</tr>
<tr>
<td>LCM TPM values (table_A_LCM_TPM_values.tsv)</td>
<td>this paper</td>
<td><a href="https://doi.org/10.5281/zenodo.1320734">https://doi.org/10.5281/zenodo.1320734</a></td>
</tr>
<tr>
<td>scRNaseq UMI counts of used cells (table_B_scRNaseq_UMI_counts.tsv)</td>
<td>this paper</td>
<td><a href="https://doi.org/10.5281/zenodo.1320734">https://doi.org/10.5281/zenodo.1320734</a></td>
</tr>
<tr>
<td>scRNaseq tsne coordinates of used cells (table_C_scRNaseq_tsne_coordinates_zones.tsv)</td>
<td>this paper</td>
<td><a href="https://doi.org/10.5281/zenodo.1320734">https://doi.org/10.5281/zenodo.1320734</a></td>
</tr>
<tr>
<td>Zonation reconstruction (table_D_zonation_reconstruction.tsv)</td>
<td>this paper</td>
<td><a href="https://doi.org/10.5281/zenodo.1320734">https://doi.org/10.5281/zenodo.1320734</a></td>
</tr>
<tr>
<td>Lgr5-eGFPneg_1</td>
<td>Yan et al., 2017b</td>
<td>GEO: GSM2644349</td>
</tr>
<tr>
<td>Lgr5-eGFPneg_2</td>
<td>Yan et al., 2017b</td>
<td>GEO: GSM2644350</td>
</tr>
<tr>
<td>Germfree sequencing data</td>
<td>Peck et al., 2017</td>
<td>GEO: GSE81125</td>
</tr>
<tr>
<td>Antibiotic microbiome ablation sequencing data</td>
<td>Chevalier et al., 2015</td>
<td>GEO: GSE74157</td>
</tr>
<tr>
<td>Mouse cell atlas</td>
<td>Han et al., 2018</td>
<td><a href="http://bis.zju.edu.cn/MCA/">http://bis.zju.edu.cn/MCA/</a></td>
</tr>
</tbody>
</table>

(Continued on next page)
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 x oil-immersion objectives and a Photometrics Pixis 1024 CCD camera using MetaMorph software as previously reported (Itzkovitz et al., 2011).

Please cite this article in press as: Moor et al., Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation along the Intestinal Villus Axis, Cell (2018), https://doi.org/10.1016/j.cell.2018.08.063
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**

8μm 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 an 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, AR9960) 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. 8μm 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 μm² 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 (Toth 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,
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-FACS AriaII sorter (BD) using a 100 μm nozzle. The APC-intensity for Ntse 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., 2017) running on 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 ≤ xi ≤ 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...
1.6.0.16 (Cox and Mann, 2008) by using the mouse uniprot database as reference. Further analysis was done using the perseusetry (LC-MS/MS) on a Q Exactive plus mass spectrometer (Thermo Fischer Scientific). The data were analyzed with MaxQuant.

All 18 samples (3 populations of 6 mice) were digested with trypsin and analyzed by liquid chromatography-tandem mass spectrometry for proteomics. The identifications were filtered for proteins identified with FDR < 0.01 with at least 2 MSMS counts and 2 peptides in genes with highly expressed zonated enterocyte genes (UMI fraction above 10%)

Gene ontology (GO) terms were obtained from Ensembl (GRCm38 release 90). All GO annotations that contained more than three 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.

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 estimateDispensions. 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.

Where $E_{g,i}$ is the expression of gene $g$ in cell $i$ in units of fraction of total cellular UMIs. To map $x$ values to spatial coordinates along the villus axis, we used the same equation to calculate the coordinate $x_{LCM}$ 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 $x_i < x_{LCM_1}$ were assigned to $V_1$, cells with $x_{LCM_1} < x_i < x_{LCM_2}$ were assigned to $V_2$, cells with $x_{LCM_2} < x_i < x_{LCM_3}$ were assigned to $V_3$, cells with $x_{LCM_3} < x_i < x_{LCM_4}$ were assigned to $V_4$, cells with $x_{LCM_4} < x_i < x_{LCM_5}$ were assigned to $V_5$ and cells with $x_{LCM_5} < x_i$ 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_B_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 $5 \times 10^{-6}$ 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).

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.

![Equation 1](https://zenodo.org/record/1320734, table_D_zonation_reconstruction.tsv)

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 deposited 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_tsnr_coordinates_zones.tsv.
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 $5 \times 10^{-3}$). 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)
Examples for reconstructed zonation profiles based on the sub-sampled landmark gene panels.

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 variable 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.

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.

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.

tSNE plots of genes that are featured in (F).
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.
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 50 μm.
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 spectrometric 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 50 μm. Protein validation for Tfrc, Cdh1 and Nt5e was performed using immunofluorescence, validation for Reg3b and Sglt1 was performed using immunohistochemistry.