



TECHNICAL ADVANCES AND RESOURCES

Single-cell transcriptome analysis reveals differential nutrient absorption functions in human intestine

Yalong Wang^{1*}, Wanlu Song^{2*} , Jilian Wang³, Ting Wang¹, Xiaochen Xiong¹, Zhen Qi¹, Wei Fu³, Xuerui Yang², and Ye-Guang Chen^{1,4} 

The intestine plays an important role in nutrient digestion and absorption, microbe defense, and hormone secretion. Although major cell types have been identified in the mouse intestinal epithelium, cell type-specific markers and functional assignments are largely unavailable for human intestine. Here, our single-cell RNA-seq analyses of 14,537 epithelial cells from human ileum, colon, and rectum reveal different nutrient absorption preferences in the small and large intestine, suggest the existence of Paneth-like cells in the large intestine, and identify potential new marker genes for human transient-amplifying cells and goblet cells. We have validated some of these insights by quantitative PCR, immunofluorescence, and functional analyses. Furthermore, we show both common and differential features of the cellular landscapes between the human and mouse ilea. Therefore, our data provide the basis for detailed characterization of human intestine cell constitution and functions, which would be helpful for a better understanding of human intestine disorders, such as inflammatory bowel disease and intestinal tumorigenesis.

Introduction

The intestine is the organ responsible for nutrient digestion and absorption (Zorn and Wells, 2009), microorganism defense and immune response (Peterson and Artis, 2014; Tremaroli and Bäckhed, 2012), and hormone secretion (Murphy and Bloom, 2006; Sanger and Lee, 2008). Due to the technology advance of large-scale single-cell transcriptome profiling, more precise and comprehensive descriptions of cell types have been obtained from a multitude of organs (Han et al., 2018b; Tabula Muris Consortium, 2018). With single-cell RNA sequencing (RNA-seq) of mouse intestinal organoids, new markers and novel subtypes of enteroendocrine cells were identified (Grün et al., 2015). Single-cell transcriptome survey of epithelial cells from different regions of murine small intestine revealed differential expression of genes in enterocytes, Paneth cells (PCs), and stem cells in the proximal versus distal regions, and new subsets of enteroendocrine cells and tuft cells were also identified (Haber et al., 2017). Single-cell RNA-seq combined with laser capture microdissection of villi uncovered the functionally zonation distribution of enterocytes along the villus axis (Moor et al.,

2018). Transcriptomes of the human fetal digestive tract and adult large intestine were also surveyed at single-cell resolution, revealing features of transcriptome dynamics during development (Gao et al., 2018). Furthermore, single-cell PCR for selected genes in monoclonal tumor xenograft models revealed that the transcriptional heterogeneity of colon cancer cells is associated with multilineage differentiation (Dalerba et al., 2011).

Despite the extensive transcriptomic analyses of the mouse small intestine, a systematic survey of the gene expression profiles of human intestinal epithelial cells at the single-cell level has not been reported. Detailed landscapes of cell heterogeneity and the related functional annotations of different human intestinal segments are still unknown. In this study, we profile the transcriptomes of 14,537 intestinal epithelial cells from the human ileum, colon and rectum. Our analyses uncover the different nutrient absorption preferences in small and large intestine, suggest the existence of Paneth-like cells (PLCs) in the large intestine, and identify potential new marker genes of specific cell types. Furthermore, our data also reveal the

¹The State Key Laboratory of Membrane Biology, Tsinghua-Peking Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing, China; ²The MOE Key Laboratory of Bioinformatics, Center for Synthetic & Systems Biology, School of Life Sciences, Tsinghua University, Beijing, China; ³Department of General Surgery, Peking University Third Hospital, Beijing, China; ⁴Max-Planck Center for Tissue Stem Cell Research and Regenerative Medicine, Guangzhou Regenerative Medicine and Health Guangdong Laboratory, Guangzhou, China.

*Y. Wang and W. Song contributed equally to this paper; Correspondence to Ye-Guang Chen: ygchen@tsinghua.edu.cn; Wei Fu: fuwei@bjmu.edu.cn; Xuerui Yang: yangxuerui@tsinghua.edu.cn.

© 2019 Wang et al. This article is distributed under the terms of an Attribution–Noncommercial–Share Alike–No Mirror Sites license for the first six months after the publication date (see <http://www.rupress.org/terms/>). After six months it is available under a Creative Commons License (Attribution–Noncommercial–Share Alike 4.0 International license, as described at <https://creativecommons.org/licenses/by-nc-sa/4.0/>).

transcriptomic variations of each cell type among the three human intestinal segments as well as variations of the same cell type between human and mouse ilea. The transcriptome data and the related bioinformatic analyses could serve as an unprecedented resource for better understanding the dynamic cell landscape and the lineage-specific functional heterogeneity of the human intestine.

Results

To obtain comprehensive cell landscapes of the human small and large intestines, we profiled single-cell transcriptomes of epithelial cells of the human ileum, colon, and rectum from six donors, with two for each intestinal segment as biological replicates (Fig. S1 A), on a 10X Genomics system. After quality filtering (see Materials and methods), the transcriptome profiles of 14,537 cells were collected (6,167 cells from two human ilea samples, 4,472 cells from two colon samples, and 3,898 cells from two rectum samples). Statistics of the cells and the detected genes were shown in Fig. S1, B–D. For each intestinal segment (ileum, colon, or rectum), cells from the two donors nicely overlapped (Fig. S1, E and F), indicating high fidelity of the data and reproducibility of the cellular landscapes obtained from the two individuals.

Single-cell characterization of human intestinal epithelium

To compare cell types in the human ileum, colon, and rectum, the 14,537 cells were pooled together, and their transcriptome profiles were subjected to unsupervised graph-based clustering (Butler et al., 2018). Based on previously reported cell markers (Fig. S2 A and Table S1) and other intestinal single-cell sequencing results (Grün et al., 2015; Haber et al., 2017), seven known cell types were identified (Fig. 1, A and B): enterocyte cells (ALPI, SLC26A3, TMEM37, and FABP2), goblet cells (ZG16, CLCA1, FFAR4, TFF3, and SPINK4), PCs (LYZ [Lyz1 and Lyz2 in mouse], CA7, SPIB, CA4, and FKBP1A), enteroendocrine cells (CHGA, CHGB, CPE, NEUROD1, and PYY), progenitor cells (SOX9, CDK6, MUC4, FABP5, PLA2G2A, and LCN2), transient-amplifying (TA) cells (KI67, PCNA, TOP2A, CCNA2, and MCM5), and stem cells (LGR5, RGMB, SMOC2, and ASCL2). Using the same cell markers (Fig. S2 B), these cell types were also identified in the ileum, colon, and rectum segments when analyzed separately (Fig. 1, C–H; and Table S2). Tuft cell markers (POU2F3, GFI1B, and TRPM5) were rarely detected in few cells (Fig. S2 C), while the marker *DCLK1* was not detected.

Classification of the cells revealed distinct cell compositions in the small and large intestine epithelia. Enterocytes were highly enriched in the human ileum, taking up ~70% of the total cells, while only 14% cells from the colon and rectum were annotated as enterocytes. Notably, in the colon and rectum, 20% of the total cells were goblet cells, which dropped to only 5% in the ileum (Fig. S2 D).

Stem cells and TA cells are highly proliferative cells and responsible for fast renewal of intestinal epithelium. Indeed, the genes related to Wnt signaling or the cell cycle were highly expressed in stem cells and TA cells (Fig. S3, A and B; and Table S3). Notably, stem cells and TA cells are enriched with *LGR5* and

KI67, respectively. Interestingly, stem cell signature genes of the three segments showed largely different function enrichments (Fig. S3 A). For example, *FABP2* and *FABP6*, involved in fatty acid metabolic process, were enriched in ileal stem cells, but not in large intestine stem cells, which is consistent with the intestinal functions. By contrast, the signature genes of TA cells in the three segments were highly consistent. In accordance with their proliferation potency, TA cells were mainly in S and G2/M phase (Fig. S3 C), like the ones in the mouse intestine (Haber et al., 2017), while differentiated cells were mainly in G1 phase.

Progenitor cells expressed both stem- and proliferation-related genes (e.g., *SOX4*, *SOX9*, *CDK6*, *SOCS2*, and *RGMB*) as well as differentiation-related factors (e.g., *ATOH1*, *DLL1*, *FOXA2*, and *FOXA3* for secretory progenitors and *HES1* and *CDX2* for enterocyte progenitors; Fig. S3, D and E; and Table S4; Clevers and Batlle, 2013), suggesting that they start to gain physiological functions. Furthermore, as shown by the specific genes (Fig. S3 D), progenitor cells exhibited functional difference in different segments. The ones from the ileum were marked by genes related to lipid and protein metabolism, and the ones from the colon and rectum were marked by genes involved in the immune response. Although stem, TA, and progenitor cells were highly proliferative, stem and progenitor cells were mainly in G1 and S phases, while TA cells were mainly in S and G2/M phases (Fig. S3 C; also see Fig. 6 D).

We have also observed specific expression of transcription factors in different types of cells (*CREB3L3*, *MAF*, and *NR1H4* in enterocytes; *ATOH1*, *SPEDF*, and *FOXA3* in goblet cells; *SPIB*, *HES4*, and *PROX1* in PCs and PLCs; *FEV*, *INSM1*, and *NEUROD1* in enteroendocrine cells; *YBX1* and *PHB* in both TA and stem cells; *HMGB2*, *FOXMI1*, and *MYBL2* in TA cells; and *ASCL2* and *ETS2* in stem cells; Fig. S3 F). These cell type-dependent transcription factors may play important roles in the differentiation or maintenance of the different cell types. Indeed, *INSM1* has been shown to be essential for enteroendocrine cell differentiation (Gierl et al., 2006), and *ASCL2* is required for the maintenance of intestinal stem cell identity (Schuijers et al., 2015).

Distinct expression patterns of genes related to nutrient absorption in the small and large intestine

The intestinal tract is the organ for food digestion, nutrient absorption, and processing, such as sugar, lipid, vitamins, and inorganic and organic solutes. Deficiency in nutrient absorption has been associated with multiple diseases (Lin et al., 2015). Extensive bowel resection causes defects of nutrient absorption, leading to short bowel syndrome and even death of these patients (Tappenden, 2014). However, the differential activity of nutrient absorption in different segments of the human intestine is not very clear. Enterocytes are the major cells responsible for nutrient absorption. To better understand the nutrient-absorption processes in the intestine, we looked into the expression profiles of metabolism-related genes in enterocytes from the ileum, colon, and rectum. In general, functional enrichment analyses showed that the genes involved in protein digestion and absorption and mineral and organic substance transport were enriched in all three segments. The genes participating in lipid metabolism and drug metabolic process were

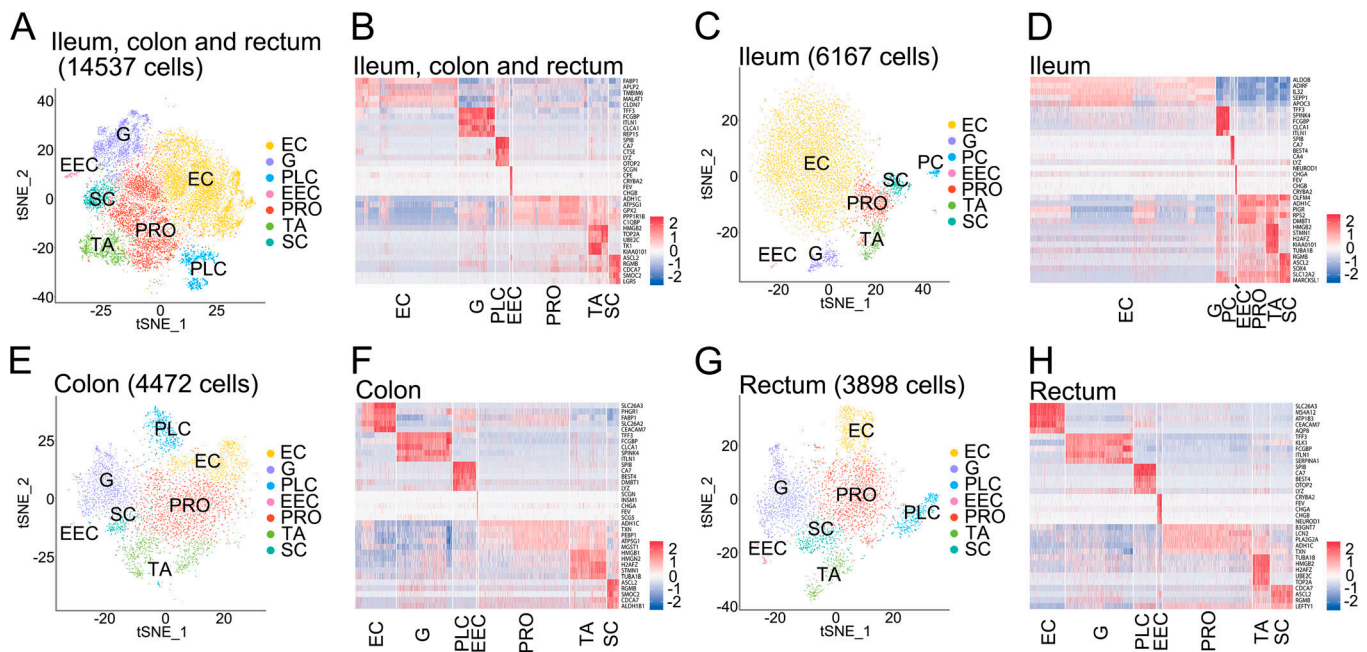


Figure 1. Cell landscapes of human intestines based on single-cell transcriptome profiles. (A, C, E, and G) t-SNE plots of single-cell clusters. The Seurat algorithm was used to visualize the clustering of all 14,537 intestine epithelial cells from six donors (A), 6,167 ileum cells from two donors (C), 4,472 colon cells from two donors (E), and 3,898 rectum cells from two donors (G). **(B, D, F, and H)** Expression heatmaps of cell type-specific genes were obtained by analyzing all cells pooled together (B) or based on cells from one of the three intestine segments: ileum (D), colon (F), and rectum (H). EC, enterocytes; EEC, enteroendocrine cells; G, goblet cells; PRO, progenitor cells; SC, stem cells.

highly expressed in the ileum, and by contrast, the genes related to small molecule transport were enriched in the large intestine (Fig. 2 A and Table S5).

As major nutrient transporters, SLC family genes play critical roles in the transport of a wide range of nutrients and metabolites such as glucose, amino acids, vitamins, inorganic solutes, and ions, and their dysfunction has been associated with numerous diseases (Lin et al., 2015; Zhang et al., 2019). We focused on the expression patterns of these transporters in the three intestine segments. In general, the transporter genes involved in lipid, bile salt, vitamin, and water absorption were enriched in the ileum, and the genes related to metal ion and nucleotide absorption were highly expressed in the large intestine (Fig. 2 B). There were no significant differences in the expression levels of transporters for amino acid, sugar, and inorganic and organic solutes among these segments.

In addition, although the ketone body metabolism gene ACAT1 was found in the large intestine, most lipid assimilation genes were highly expressed in the small intestine, such as APOA1 and APOM for lecithin, sterol, and stearic acid (Fig. 2 C; Zhang et al., 2017). The specific expression of APOA4, APOB, and FABP6 in the ileum was confirmed at the protein level (Fig. S4, A–D). Bile acids, which emulsify fat and fat-soluble nutrients and are essential for their absorption, are secreted into the duodenum through the bile duct (Di Ciaula et al., 2017). The genes involved in bile salt reabsorption were mainly found in the ileum, but not large intestine (Fig. 2 C). Similarly, the genes related to vitamin absorption were enriched in the ileum, such as RBP2, TCN2, CYP4F2, and SLC23A1 for the absorption of vitamin A, B₁₂, K, and C (de Oliveira, 2015; Goncalves et al., 2015; Reboul,

2015). Notably, the large intestine could also transport vitamin B₁₂ and A, as suggested by the expression of CD320, DHRS9, and RBP4 (Arora et al., 2017; Jones et al., 2007; Zhou et al., 2018). The expression of aquaporin (AQP) 1, 3, 7, and 11 was mainly found in the ileum and AQP8 in the large intestine (Fig. 2 C), supporting the note that most water is absorbed in the small intestine and further dehydration occurs in the large intestine (Verkman et al., 2014).

Although there was no significant difference in the mean expression of amino acid, sugar, and inorganic and organic solute transporters among the three segments (Fig. 2 B), some specific transporter genes still displayed distinct expression patterns in the small and large intestine (Fig. 2 C). Consistent with the note that both the small and large intestine are involved in the absorption of essential amino acids, amino acid transporter genes (such as SLC3A2, SLC25A39, and SLC25A13) were found in the ileum, colon, and rectum (Kobayashi et al., 1999; Nicklin et al., 2009; Nilsson et al., 2009). However, SLC7A7 and SLC7A9, the two genes involved in the transport of neutral or cationic amino acids such as leucine or arginine, were only found in the ileum (Suhre et al., 2011; Torrents et al., 1999). SLC38A1, an important cotransporter of glutamine, was confirmed in large intestine (Fig. S4 E). For inorganic solutes, SLC4A7 and SLC34A3 for carbonic acid and phosphoric acid transport, respectively, were mainly expressed in the small intestine, while the gene SLC26A2 for sulfuric acid transport was enriched in the large intestine (Heneghan et al., 2010), which was also confirmed at the protein level (Fig. S4 F). For organic solutes, the Na-dependent dicarboxylate transporter SLC13A2 was enriched in the small intestine, while the choline

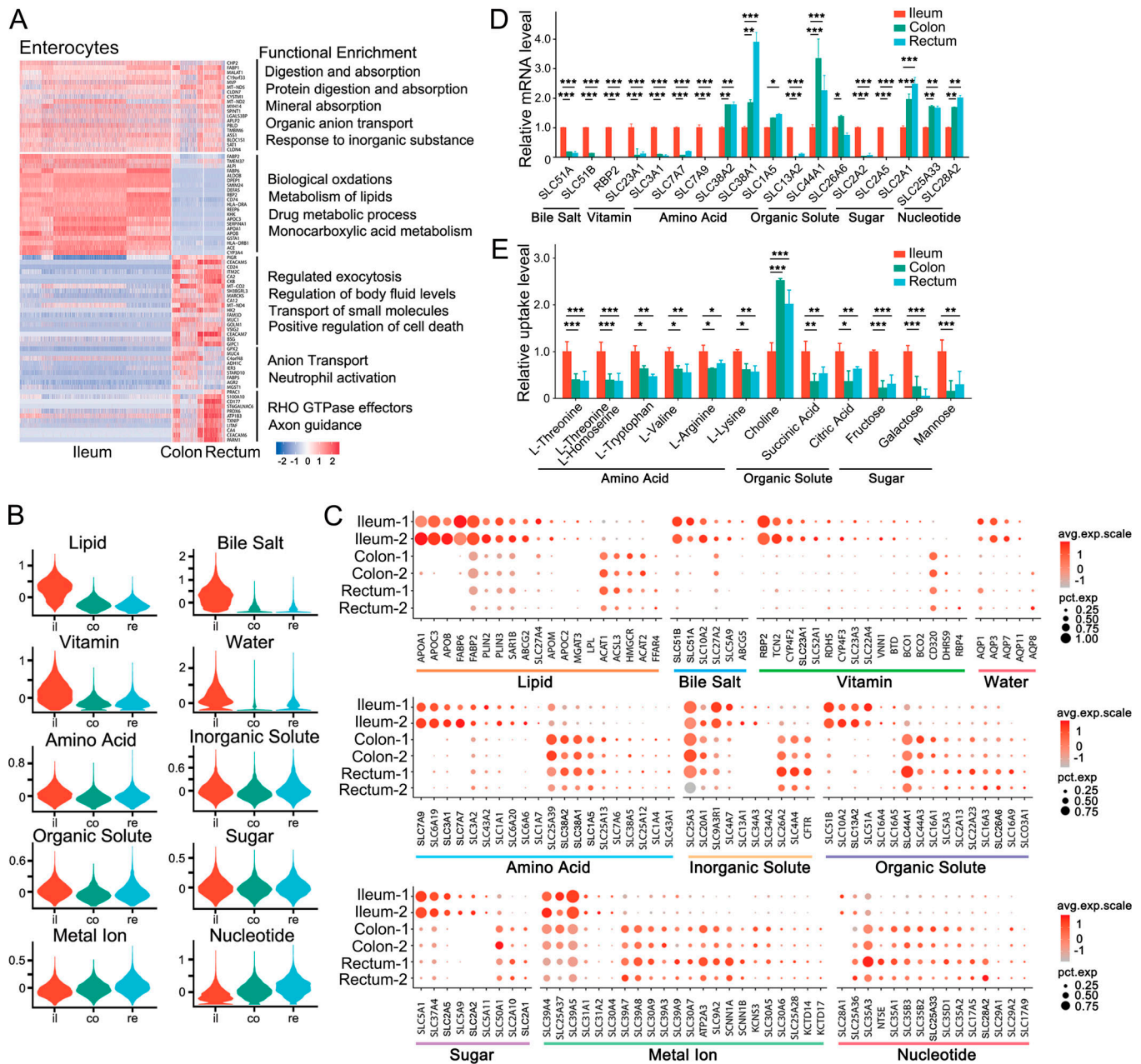


Figure 2. Differential nutrient absorption preferences in the human small and large intestine. (A) Expression heatmap of signature genes in enterocytes and gene functional enrichments. (B) Violin plots showing distributions of mean expression of transporter genes in different segments (il, ileum; co, colon; re, rectum). (C) Expression patterns of specific genes involved in nutrient absorption and transport in different intestine segments. Each dot represents a gene, of which the color saturation indicates the average expression level (scaled by Z-score) in an intestine segment, and the size indicates the percentage of cells expressing the gene. (D) Relative mRNA expression of the gene marked in red in C was confirmed by qPCR in the organoids derived from human ileum, colon, and rectum. (E) Nutrient uptake was measured by LC-MS in the organoids derived from human ileum, colon, and rectum. Data are represented as mean \pm SD of at least three independent experiments. All statistically significant differences were calculated using an ordinary two-way ANOVA followed by Tukey's multiple comparisons test; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

transporters SLC44A1 and SLC44A3 were found to be enriched in large intestine (Figs. 2 C and S4 G; Traiffort et al., 2013).

Both the small and large intestine are involved in sugar absorption, but the small intestine may specially transport monosaccharides such as glucose, fructose, galactose, and xylose based on the enriched expression of SLC5A1, SLC2A5, SLC5A9, and SLC5A11 (Coady et al., 2002; Tazawa et al., 2005; Wood and Trayhurn, 2003), while the large intestine may specially

transport aldoses, including pentoses and hexoses, as suggested by the expression of SLC50A1 (Wright, 2013; Fig. 2 C). The expression of Na, K, and Ca channels (KCNS3, ATP2A3, SCNN1A, and SCNN1B) was consistent with the idea that the large intestine is important for metal ion absorption (Georgiev et al., 2014; Hummler and Beermann, 2000; Kunzelmann and Mall, 2002). SCNN1B expression in the large intestine was also confirmed at the protein level (Fig. S4 H). The Cu transporter SLC31A2 was

enriched in the ileum, while the transporters for bivalent metal ions like Zn and Mn (SLC39A5, SLC39A8, and SLC39A7) were expressed in both the small and large intestine (Bogdan et al., 2016; Choi et al., 2018; Eide, 2004). Most importantly, our data also suggested that the large intestine is the major site for nucleotide or nucleotide sugar absorption (Fig. 2 C), and SLC35A1 expression in the large intestine was confirmed at the protein level (Fig. S4 I; Song, 2013).

Finally, to confirm the functional differences among the three segments, we generated organoids from human ileum, colon, and rectum for various nutrient-uptake experiments. In the organoids, expression patterns of the genes involved in nutrient absorption were confirmed to be consistent with our single-cell transcriptome data (Fig. 2 C). Next, functional assays revealed that six types of amino acids were highly absorbed in the ileum, consistent with the enriched expression of SLC3A1, SLC7A7, and SLC7A9, which are mainly responsible for the absorption of neutral and cationic amino acids, such as arginine and lysine (Suhre et al., 2011; Torrents et al., 1999; Fig. 2 E). The high expression of SLC44A1 was confirmed in the large intestine, which is consistent with more choline absorption in the organoids derived from the large intestine, while SLC13A2, which is responsible for succinic acid and citric acid absorption, was highly expressed in the small intestine, and the uptake experiments also confirmed this (Fig. 2, D and E). SLC2A5 and SLC2A2, which are responsible for galactose, fructose, and mannose absorption, were highly expressed in the small intestine (Fig. 2, C and D). Consistently, sugar uptake analyses revealed that galactose, fructose, and mannose were mainly absorbed in the small intestine (Fig. 2 E), in agreement with an earlier report (Raja et al., 2012).

Differential expression of signaling molecules in the small and large intestine

Differential expression of signaling molecules in enterocytes was also observed in the three segments. For instance, higher expression of some mediators of cell death and TGF- β /BMP signaling was found in the large intestine, especially in the rectum (Fig. 3 A). The Wnt signaling mediators FZD5 and DVL3 were also upregulated in the rectum. The high expression of both the proproliferative and prodeath genes suggests that the epithelium of the large intestine, particularly the rectal epithelium, may undergo more rapid turnover.

Although the enteroendocrine cells in the three segments shared a similar expression profile, such as expression of the hormone secretion-related genes PCSK1N and SCG5, some genes showed a clear segment-specific expression pattern (Fig. 3 B). Analysis of hormone expression in enteroendocrine cells revealed that some of the hormones were highly expressed in the small intestine (e.g., secretin, neurotensin, and cholecystokinin), while some were enriched in the large intestine (e.g., peptidyl glycine α -amidating monooxygenase, peptide tyrosine-tyrosine, and insulin-like peptide 5; Fig. 3 C and Table S6).

Functional enrichment analysis on the immunity-related genes in the three segments suggested that although both the small and large intestine participate in the antimicrobial humoral response, the small intestine may have a strong defense

response to fungi, while the large intestine may be more sensitive to bacterial infection (Fig. 3 D).

Characterization of PLCs in the human large intestine

PCs, located at the bottom of crypts in the small intestine, secrete antimicrobial molecules modulating host-microbe interactions and factors promoting Lgr5⁺ intestine stem cells (Clevers and Bevins, 2013; Zhang and Liu, 2016). Single-cell PCR gene expression analysis has identified a subset of c-Kit⁺ goblet cells in the mouse colon that might have the equivalent function of PCs in supporting Lgr5⁺ stem cells (Rothenberg et al., 2012). Recently, PLCs were reported in the rat ascending colon and human fetal large intestine (Gao et al., 2018; Mantani et al., 2014). To further verify the existence of PLCs in the human large intestine, we examined cells in the colon and rectum using Paneth marker genes (LYZ, CA4, CA7, and SPIB) and found the PLC cluster (Fig. 4 A and Table S7). The PLCs in the large intestine and the PCs in the ileum shared a set of highly expressed genes, which include not only genes for microbiotic defense such as LYZ (Fig. 4, A-D), but also genes encoding the niche factors to sustain Lgr5⁺ stem cells such as EGF, Wnt3, Notch, ephrin A/B, and PDGF ligands (Sato et al., 2011; Fig. 4 E).

Interestingly, PCs in the ileum and PLCs in the large intestine also exhibited marked differences. Functional enrichment of the signature genes showed that PCs and PLCs shared genes involved in lysosome function, neutrophil activation, and Gram-negative bacterium response, while the genes involved in biological oxidation were specific to the ileum and the genes involved in inorganic and sulfur metabolism were enriched in the large intestine (Fig. 4 A). For example, DEFA5, DEFA6, REG1A, and REG3A were found in ileal PCs, but not in large intestinal PLCs, suggesting that the antimicrobial function may be a major difference between these cells. We found that GNPTAB and SOD3 were specially expressed in PLCs in the human large intestine, but not in PCs in the ileum (Fig. 4 F), suggesting that GNPTAB and SOD3 may serve as a potential marker of PLCs.

PCs and PLCs shared some common transcription factors involved in Paneth differentiation and viral defense, such as HES1, HES4, and SPIB (Fig. 4 G). However, some other transcription factors exhibited a segment-specific pattern. For instance, SATB2, a chromatin organizer that functions in chromatin remodeling and gene expression and is involved in carcinogenesis, including colorectal cancer (Naik and Galande, 2019), was enriched in PLCs of the large intestine. RELB, which is involved in NF- κ B signaling, was highly expressed in the ileal PCs. Interestingly, KIT (c-Kit in mouse) was detected in some cells, but not in PLCs (Fig. 4 H). Moreover, another representative gene of mouse c-Kit⁺ goblet cells, CD117, was not detected.

Potential new markers for human TA cells and goblet cells

TA cells are derived from stem cells and generate progenitor cells, which eventually differentiate into mature functional cells (Gehart and Clevers, 2018). However, stem cells, TA cells, and progenitor cells all have proliferation potential, and they are difficult to separate using BrdU or EdU labeling. Therefore, identification of TA cell-specific markers would be critical for further characterization of these cells. Based on the

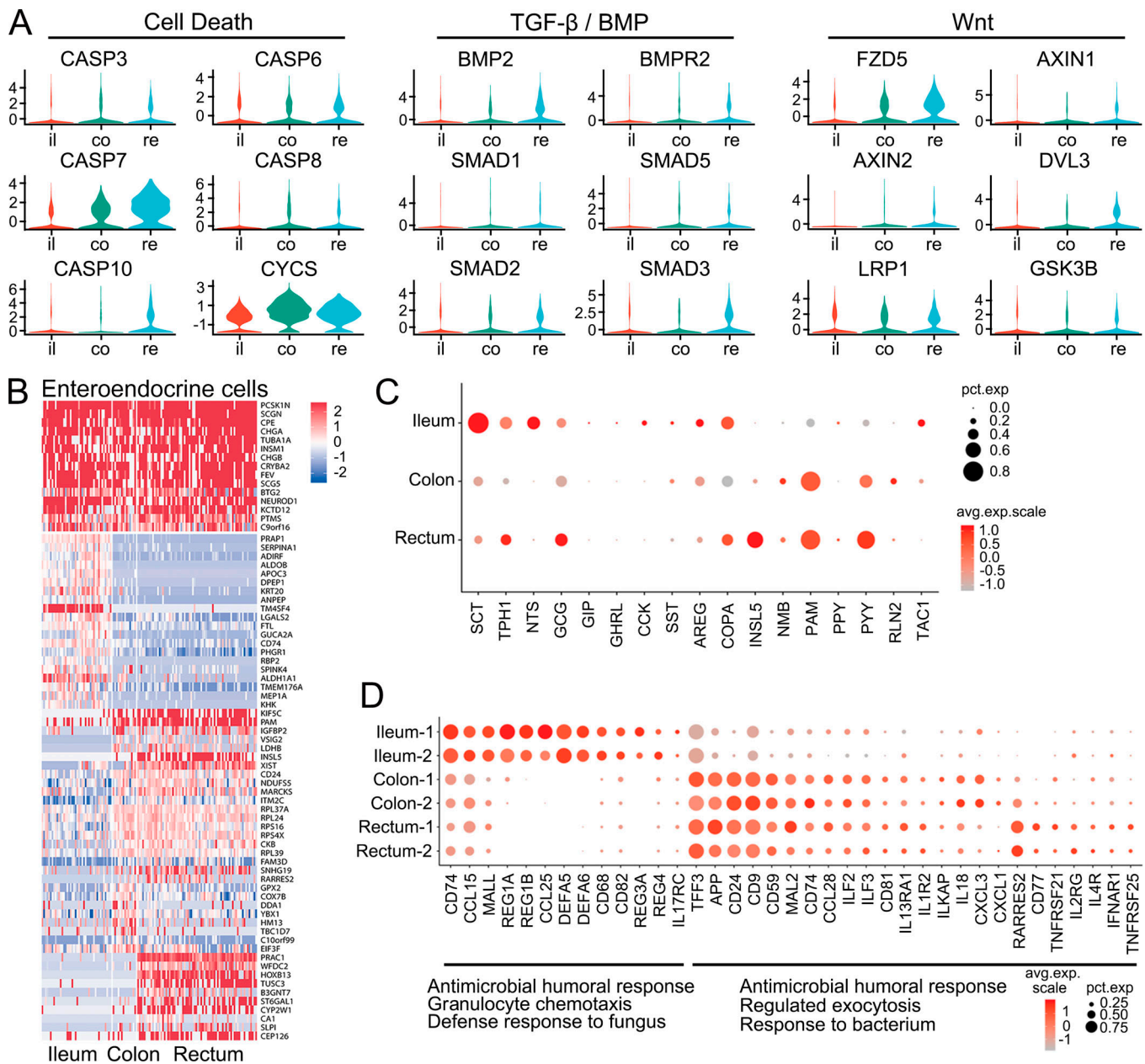


Figure 3. Differential expression of signaling molecules, hormones, and immunity-related genes. (A) Violin plots showing expression distributions of signaling genes in enterocytes from the ileum (il), colon (co), and rectum (re). **(B)** Expression heatmap of signature genes in enteroendocrine cells. **(C)** Expression patterns of specific hormone genes in enteroendocrine cells from different intestine segments. **(D)** Expression patterns of immune-related genes in cells from different intestine segments and gene functional enrichments. Each dot represents a gene, of which the color saturation indicates the average expression level (scaled by Z-score) in an intestine segment and the size indicates the percentage of cells expressing the gene.

transcriptome analysis, we found that NUSAP1 (nucleolar and spindle associated protein 1), which is up-regulated in colorectal cancer (Han et al., 2018a), was specifically expressed in the TA cluster in the ileum, colon, and rectum, just like KI67 (Fig. 5 A and Fig. S5, A and B). However, immunofluorescence analysis showed that almost all NUSAP1⁺ cells were costained with KI67⁺ cells, while 45% KI67⁺ cells were NUSAP1⁺ (Fig. 5, B and C). Gene Ontology (GO) analysis of the genes enriched in NUSAP1⁺ cells unveiled that these cells were highly proliferative (Fig. S5 C). Unlike PCNA, NUSAP1 expression did not overlap with LGR5 (Figs. 5 A and S5 D). In addition, Nusap1 was

not colocalized with Lgr5 in mouse intestine (Fig. S5 E). Taken together, these observations suggest that NUSAP1 may serve as a potential specific marker of a subset of TA cells.

The main function of goblet cells is to secrete mucus that protects the epithelial membrane. Interestingly, we found that the genes involved in calcium transport were highly expressed in goblet cells of all three segments, while the genes related to the vitamin metabolic process were found in the goblet cells of ileum and the genes related to salmonella infection in the colon (Fig. 5 D and Table S8). Specifically, ITLN1 (interectin-1/omectin-1), which binds to microbial glycans and is involved in innate

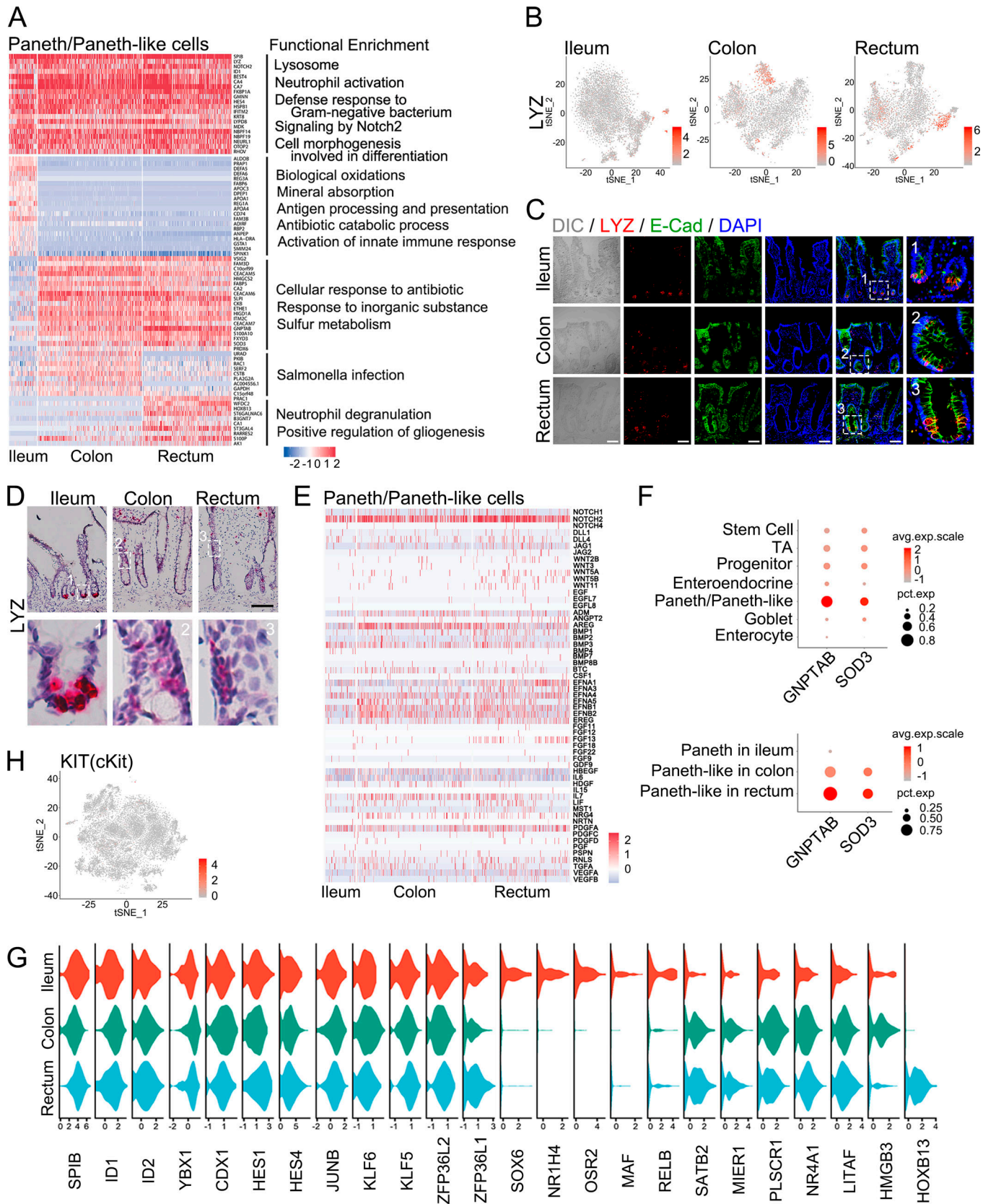


Figure 4. **PLCs identified in the human colon and rectum.** (A) Expression heatmap of signature genes in PLCs and gene functional enrichments. (B) Lysozyme (LYZ) expression (indicated by color saturation) in single cells from the ileum, colon, and rectum. (C) Immunofluorescence was performed to confirm LYZ expression in the ileum, colon and rectum. Scale bars, 100 μ m. (D) smFISH results of LYZ in human ileum, colon, and rectum. Scale bar, 100 μ m. (E) mRNA expression levels of growth factors in PCs and PLCs in the ileum and PLCs in the colon and rectum. (F) Specific expression of GNPTAB and SOD3 in PCs and PLCs of the human large intestine. (G) Violin plots showing expression distributions of transcription factors in PCs or PLCs. (H) Kit expression in all 14,537 cells.

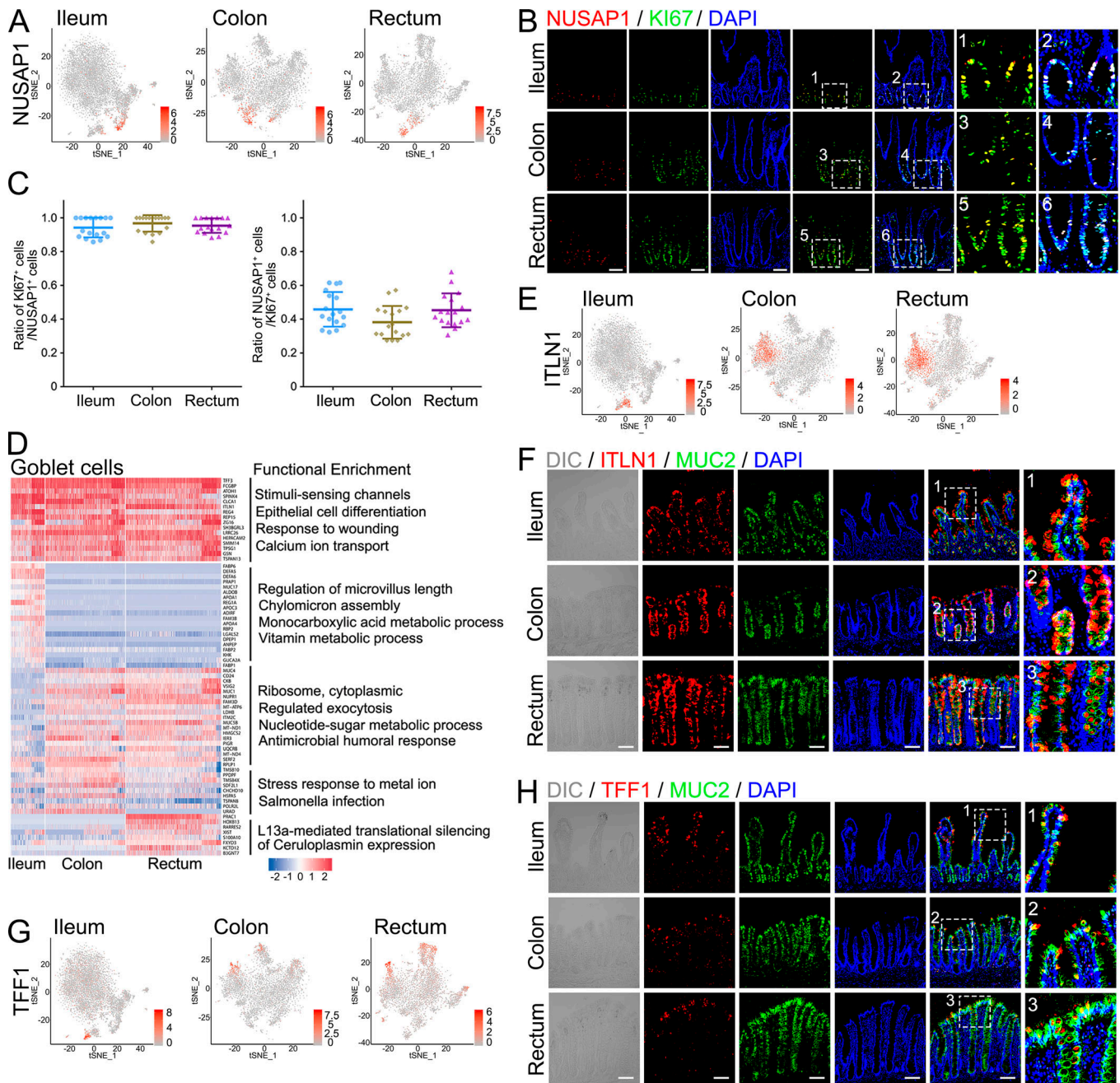


Figure 5. **Potential new markers of human TA cells and goblet cells.** (A) NUSAP1 expression (indicated by color saturation) in single cells from the ileum, colon, and rectum. (B) Immunofluorescence was performed to confirm the NUSAP1 expression in the ileum, colon, and rectum. Scale bars, 100 μ m. (C) Statistics of the NUSAP1 immunofluorescence data shown in B. The proportion of Ki67⁺/NUSAP1⁺ cells and NUSAP1⁺/Ki67⁺ cells were quantified. Data are represented as mean \pm SD of $n = 17$. (D) Expression heatmap of signature genes in goblet cells and gene functional enrichments. (E) ITLN1 expression (indicated by color saturation) in single cells from the ileum, colon, and rectum. (F) Immunofluorescence was performed to confirm ITLN1 expression. Scale bars, 100 μ m. (G) TFF1 expression (indicated by color saturation) in single cells from the ileum, colon, and rectum. (H) Immunofluorescence was performed to confirm the TFF1 expression. Scale bars, 100 μ m.

immunity (Wesener et al., 2017), was specifically expressed in all goblet cells of the ileum, colon, and rectum (Fig. 5, E and F; and Fig. S5 F). Immunofluorescence analysis revealed that ITLN1⁺ cells were costained with MUC2 and distributed along the villus in the ileum and crypts in the colon and rectum (Fig. 5 F), suggesting that ITLN1 is a potential new marker of human goblet cells. However, unexpectedly, *Itln1* was only found in mouse PCs

(Fig. S5 G), suggesting a major difference between human and mouse goblet cells.

TFF1 encodes a Trefoil factor peptide that plays an important role in response to gastrointestinal mucosa injury and inflammation. As an isoform of TFF1, TFF3 was found in mouse and human goblet cells (Aihara et al., 2017; Haber et al., 2017), while TFF1 is expressed in a subset of human goblet cells, but not in

mouse intestinal cells (Fig. 5 G and Fig. S5, H and I). Moreover, TFF1 protein was found only in the villus of the ileum and in the top zone of crypts of the colon and rectum (Fig. 5 H), suggesting that these cells may represent mature goblet cells. Indeed, the signature genes of TFF1⁺ goblet cells were highly enriched by the function of antigen processing via MHC class (Fig. S5 J), and the MHC-related genes (HLA-A, HLA-B, HLA-C, and HLA-E) were indeed highly expressed in TFF1⁺ goblet cells (Fig. S5 K). Interestingly, DEFA5 and DEFA6, both of which are expressed in PCs, were enriched in ileal goblet cells (Fig. 5 D), suggesting the antibiotic function of these two cell types. Reg4, which is a marker gene of enteroendocrine cells in the mouse intestine (Haber et al., 2017), was found in human goblet cells (Fig. 5 D).

Cell-type variations of gene expression in the human and mouse ilea

To gain a better understanding of the differences between the human and mouse intestine, we compared our data with the published transcriptome data of the mouse ileum (Haber et al., 2017). A total of 6,187 single cells from human ileum and 3,927 single cells from mouse ileum were combined and subjected to unsupervised graph-based clustering based on their gene expression profiles. As shown in Fig. 6 (A and B), the overall gene expression pattern in major cell types was conserved between human and mouse. Cell cycle analysis showed that while mouse stem cells were mainly in S and G2/M phase, human stem cells were mainly in G1 phase (Fig. 6, C and D). This is consistent with the slow cycling of human colon stem cells in the mouse xenograft system for normal human colon organoids (Sugimoto et al., 2018). In contrast, TA cells were mainly in S and G2/M phase in both human and mouse (Fig. 6, C and D).

Next, we also examined the conservativeness of the marker genes between human and mouse ilea. In addition to the markers used for cell clustering, such as TMTM37 in enterocyte cells, TFF3 in goblet cells, and CHGB in enteroendocrine cells, many other signature genes were also conserved in both human and mouse, such as FEV and VWA5B2 in enteroendocrine cells and REPI5 and BCAS1 in goblet cells (Fig. 6 E). LYZ, whose homologous genes are *Lyz1* and *Lyz2* in mouse, is a well-known marker of PCs (Sato et al., 2011), and they were expressed in PLCs in human and mouse large intestines. Interestingly, we also noticed that some genes showed heterogeneities between species in the same cell type. For instance, stem cells from human and mouse ilea shared known markers, such as *LGR5*, *SMOC2*, *ASCL2*, and *RGMB* (Fig. 6, E and F), but *ZFP36L1* and *PDZK1IP1* were only found in human stem cells, while *SP5* and *RGCC* were found only in mouse stem cells (Fig. 6, F and G). Furthermore, some genes were enriched in one cell type of human ileum but might exist in another cell type in mouse. For example, TFF3 was expressed in both mouse and human goblet cells, but TFF1 was enriched only in human goblet cells and could not be detected in mouse intestinal cells (Figs. 5 D and S5 I). *ITLN1*, which is expressed in mouse PCs, was enriched in human goblet cells (Figs. 5 F and S5 G). In summary, these observations further confirm the conservative marker genes in human and mouse intestinal epithelial cells and also reveal special cell type signatures with distinct expression pattern across human and mouse ilea.

Discussion

As the organ of nutrient digestion and absorption, microbe defense, and endocrine function, the pathophysiological processes and related regulations of the intestine have been extensively studied. However, many important questions still remain unclear. For example, the functional differences among cells of the same type in different intestine segments are poorly understood. In this study, using single-cell RNA-seq, we surveyed the gene expression profiles of the epithelium in the human ileum, colon, and rectum at single-cell resolution for the first time. Our data revealed the differential functions of nutrient absorption in these segments. We confirmed the presence of PLCs in the large intestine and found different gene expression patterns between human and mouse. In addition, potential new markers were identified for human TA cells and goblet cells. These results provide the basis for a better appreciation of human intestine cell constitution and functions as well as further investigation of enterocolitis and intestinal tumorigenesis.

Our data unveiled the differential expression of nutrient absorption-related genes in human ileum, colon, and rectum. High expression of the genes related to transport of lipid, bile salt, vitamin, and water in the ileum indicates that the absorption of these nutrients is mainly accomplished in the small intestine, which is consistent with an earlier report (Verkman et al., 2014). Although mean expression of the genes related to the transport of amino acids, sugars, and inorganic and organic solutes was similar among the three segments, the expression of individual transporters varied in different segments, suggesting there may be preferential absorption of different nutrients or metabolites in different parts of the intestine. Further investigation is needed to obtain a clearer landscape of nutrient absorption in human intestine.

PLCs have been recently reported in rat colon and human fetal large intestine (Gao et al., 2018; Mantani et al., 2014). In addition, a subset of c-Kit⁺ goblet cells that might have the equivalent function of PCs have been described in the mouse colon (Rothenberg et al., 2012). Our data provided compelling evidence of the existence of PLCs in the human large intestine and showed that these cells express genes related to microbe defense and niche factors to sustain Lgr5⁺ stem cells.

Mouse models have been widely used to investigate the mechanisms of human diseases and test drug toxicity and efficacy. Comprehensive assessment of the differences and similarities between mouse and human is a key for the proper application of mouse models. By comparing the transcriptomes of mouse and human ileum epithelial cells, we found different signature gene expression patterns in mouse and human ileum. For instance, we found that *ITLN1* and *Reg4* were enriched in human goblet cells, but not in Paneth and enteroendocrine cells as reported in the mouse ileum (Haber et al., 2017). Understanding the precise gene expression difference in mouse and human cells would surely help to establish better mouse models for human diseases.

Materials and methods

Human intestine tissue collection and ethics statement

Intestine mucosa were freshly sampled at least 10 cm away from the tumor border in six surgically resected specimens from six

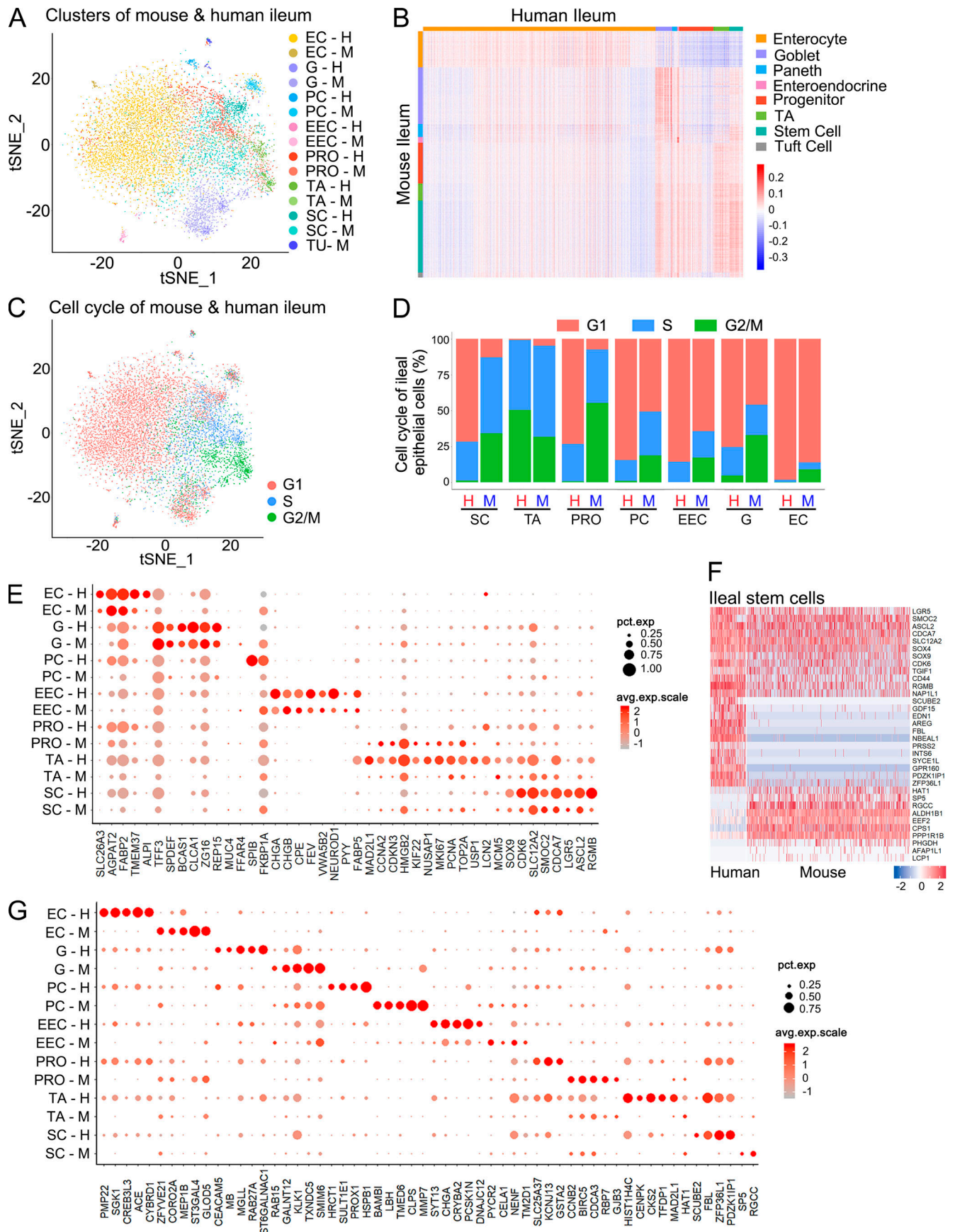


Figure 6. **Comparison of gene expression profiles between the epithelial cells of the human and mouse ilea.** (A) Unsupervised graph-based clustering of single cells from human ($n = 6,187$) and mouse ileum ($n = 3,927$) based on their gene expression profiles. (B) Similarity matrix showing the Pearson correlations between each pair of single cells from human ($n = 6,187$) and mouse ileum ($n = 3,927$) based on their gene expression profiles. (C) Cell cycle analysis. The calculated cell cycle phase of each cell is shown for all 10,114 ileal cells. (D) Cell cycle quantity of human and mouse ileal cells. (E) Conserved signature genes of each cluster in human and mouse ilea. (F) Signature genes of human and mouse ileal stem cells. (G) Human- or mouse-specific signature genes of each cell type in the ilea. Each dot represents a gene, of which the color saturation indicates the average expression level (scaled by Z-score) in a specific cell type, and the size indicates the percentage of cells expressing the gene. EC, enterocytes; EEC, enteroendocrine cells; G, goblet cells; H, humans; M, mice; PRO, progenitor cells; SC, stem cells; TU, tuft cells.

patients who had been diagnosed with intestine tumors at Peking University Third Hospital, Beijing, China. All samples were obtained with informed consent, and the study was approved by the Peking University Third Hospital Medical Science Research Ethics Committee (M2018083). All relevant ethical regulations of Peking University Third Hospital Medical Science Research Ethics Committee were followed.

cDNA library construction and single-cell RNA-seq

Intestinal tissues were washed in cold HBSS several times to remove mucus, blood cells and muscle tissue. Connective tissue was scraped away carefully. Then, epithelial tissue was cut into small pieces (5 mm) and incubated in 5 mM EDTA in HBSS for 30 min at 4°C. The pieces were transferred into cold HBSS and vigorously suspended to obtain fractions. Mesenchymal and immune cells were further removed by discarding supernatant after centrifugation (10 s at 200 rpm). Then, epithelial tissue was enriched through centrifugation (3 min at 1,000 rpm). The sediment was incubated in 2 mg/ml collagenase I (Sigma-Aldrich) in Advanced DMEM/F12 for 15 min at 37°C. After centrifugation (3 min at 1,000 rpm), the sediment was incubated in Tryple (Invitrogen) for 20 min at 37°C to obtain single-cell suspension. The cell suspension was stained with propidium iodide (PI; 5 µg/ml), and PIPI-negative single cells were sorted by FACS (BeckMan). Single cells were captured in the 10X Genomics Chromium Single Cell 3' Solution, and RNA-seq libraries were prepared following the manufacturer's protocol (10X Genomics). The libraries were subjected to high-throughput sequencing on an Illumina HiSeq X Ten PE150 platform, and 150-bp paired-end reads were generated.

Process and quality control of the single-cell RNA-seq data

The raw sequencing reads were first demultiplexed using Illumina bcl2fastq software to generate 150-bp paired-end read files in FASTQ format. The reads were then aligned to the GRCh38 human reference genome using the Cellranger toolkit (version 2.1.0) provided by 10X Genomics. The exonic reads uniquely mapped to the transcriptome were then used for unique molecular identifier (UMI) counting. Selection and filtering of the droplet barcodes for single cells were done using the Cellranger toolkit as described before (Haber et al., 2017; Kinchen et al., 2018). In brief, the 99th percentile of the total UMI counts divided by 10 was used as cutoff for calling of single cells. Subsequently, the filtered single cells and their UMI count matrices were imported into R package "Seurat" (version 2.3.2) for further analysis (Satija et al., 2015). After discarding the genes expressed in fewer than three cells, low-quality cells were

further filtered if they expressed ≤ 200 genes. Furthermore, the cells with $>50\%$ of the genes from the mitochondrial were also discarded. Finally, mesenchyme, immune, and hematopoietic cells were removed based on these marker genes LSP1, MZB1, VIM, CD52, CD78B, and COL3A1. CD45 was not detected in our results.

Data normalization and batch correction

Library size normalization was performed using Seurat NormalizeData. Specifically, the global-scaling normalization method "LogNormalize" normalized the gene expression measurements for each cell by the total expression, multiplied by a scaling factor (10,000 by default), and the results were log-transformed. Next, the six batches of single-cell RNA-seq data were subjected to batch correction, as described previously (Mayer et al., 2018). In brief, the canonical correlation analysis (CCA) strategy was used to find linear combinations of features across datasets that are maximally correlated. The shared correlation structure conserved among the six datasets from the ileum, colon, and rectum were identified. Based on the shared structure, all six batches of data were finally pooled into a single object for downstream analyses (Butler et al., 2018; Hardoon et al., 2004). Batch distributions for each dataset were visualized using t-distributed stochastic neighbor embedding (t-SNE) plots.

Unsupervised clustering analysis

The R package Seurat was used to combine linear and nonlinear dimensionality reduction algorithms for unsupervised clustering of single cells. Specifically, first, highly variable genes were identified by the FindVariableGenes function, and average expression and dispersion for each gene were calculated. Subsequently, CCA was performed based on the variable genes in the six intestine samples. The canonical correlation vectors then projected each dataset into the maximally correlated subspaces for downstream analysis. Graph-based clustering was performed, which allocated cells in a K-nearest neighbor graph structure based on high correlation strength CCA. The cells were then iteratively clustered, and the modularity was optimized with the Louvain algorithm. Finally, we used t-SNE to place cells with similar local neighborhoods in high-dimensional space or low-dimensional space based on scaled expression of variable genes to visualize the clustering results of all the cells.

Differential gene expression analysis

To identify signature genes of each cell type, the functions FindAllMarkers and FindMarkers in Seurat were used with the

following configurations: min.pct = 0.10, thresh.use = 0.25, test.use = “roc”. For a given cluster, FindAllMarkers identified positive markers compared with all other cells. The receiver operating characteristic test was used to return the “classification power” for any individual marker (ranging from 0 [random] to 1 [perfect]), and for each cluster, five genes with high area under the curve score were identified as candidate cell-type signature genes. All differentially expressed genes as positive markers of specific cell clusters are listed in Tables S1–S8. Expression heatmaps of the signature genes for each cluster are shown in Fig. 1. Similarly, the function FindMarkers was used for identification of signature genes by comparing the cell type of interest to another specific group of cells (e.g., intestine-segment-specific expression of immunity-related genes). Differential expression analysis of transcription factors was performed for the full list of human transcription factors obtained from the Animal Transcription Factor Database (<http://bioinfo.life.hust.edu.cn/AnimalTFDB/>).

Cell cycle analysis

Cell cycle stage annotation of each cell was performed using the Cell Cycle Scoring function in Seurat, which assigns each cell a score based on the expression of 43 marker genes for G2/M phase and 54 marker genes for S phase (Table S9; Buettner et al., 2015; Macosko et al., 2015; Tirosch et al., 2016).

Gene expression correlation between human and mouse ileum cells

Single-cell transcriptome data of mouse ileum was obtained from Gene Expression Omnibus (GEO) accession no. GSE92332 (Haber et al., 2017). Altogether, we compared gene expression matrices of 6,187 human ileum cells in this study to the 3,927 mouse ileum cells, which were subjected to the same process of quality control and filtering. We only considered the homologous genes between human and mouse, which eventually generated a scaled expression matrix for 11,157 genes of 10,114 cells. For each pair of cells from human and mouse, the Pearson correlation was calculated with the scaled expression data of the genes in the two cells.

Comparison between human and mouse ileum cells

We obtained single-cell transcriptome data of mouse ileum from GEO accession no. GSE92332. Altogether, we analyzed gene expression matrices of 6,187 human ileal cells in this study and 3,927 mouse ileal cells after removing the low-quality cells using the same filtering strategy with the human data. We considered the expression data of all 11,157 homologous genes with identical gene names between the human and mouse datasets and then performed CCA as implemented in the Seurat (Butler et al., 2018) to combine 10,114 human and mouse ileal cells together. t-SNE, cell cycle, and differential expression analyses were performed using the same methods described above.

Immunofluorescence and immunohistochemistry

Immunofluorescence and immunohistochemistry were performed as previously described (Qi et al., 2017). Briefly, human intestinal tissues were washed in cold HBSS to remove muscle tissue, fixed with 4% formaldehyde solution for 2 h at 4°C, and

dehydrated in 30% sucrose solution at 4°C overnight. Next, the tissue was embedded in optimal cutting temperature compound and stored at –80°C. The sections were prepared with vibrating blade microtome (HM650; Microm) and permeabilized with PBST solution (3% BSA and 0.1% Triton X-100 in PBS) for 2 h at room temperature. Then, the sections were incubated overnight with the primary antibody at 4°C. The fluorescein-labeled secondary antibodies (1:300; Life Technologies) for immunofluorescence or secondary horseradish peroxidase-conjugated anti-rabbit antibody (1:200; Invitrogen) for immunohistochemistry were added for 2 h at room temperature. Confocal laser scanning (FV3000; Olympus) or 3,3'-diaminobenzidine development (CytoMatom; Dako) was used to detect the staining signals.

Animals and mouse intestine sections

C57BL/6J mice were obtained from the Laboratory Animal Research Center of Tsinghua University. Lgr5-EGFP mice were obtained from The Jackson Laboratory. All mice were housed at a specific pathogen-free experimental animal facility at the Laboratory Animal Research Center of Tsinghua University with water and food ad libitum and a 12-h/12-h night/daylight cycle. All mice were backcrossed into the C57BL/6 genetic background for at least 10 generations. C57BL/6 and Lgr5-EGFP mice aged 8–10 wk were used to obtain intestine. Animals were then euthanized, and tissue was processed immediately. All animal experiments were conducted in accordance with the relevant animal regulations with approval of the Institutional Animal Care and Use Committee of Tsinghua University.

Antibodies

Rabbit anti-LYZ (1:200, ab108508; Abcam), mouse anti-MUC2 (1:200, ab11197; Abcam), rabbit anti-NUSAP1 (1:100, 12024-1-AP; Proteintech), mouse anti-Ki67 (1:300, 9449s; CST), mouse anti-E-Cad (1:1,000, 610182; BD Biosciences), rabbit anti-ITLN1 (1:50, 11770-1-AP; Proteintech), rabbit anti-TFF1 (1:50, 13734-1-AP; Proteintech), rabbit anti-APOB (1:50, 20578-1-AP; Proteintech), rabbit anti-APOA4 (1:100, 17996-1-AP; Proteintech), rabbit anti-SLC26A2 (1:200, 27759-1-AP; Proteintech), rabbit anti-SCNN1B (1:200, 14134-1-AP; Proteintech), rabbit anti-SLC35A1 (1:400, 16342-1-AP; Proteintech), rabbit anti-FABP6 (1:500, 13781-1-AP; Proteintech), rabbit anti-SLC38A1 (1:100, 12039-1-AP; Proteintech), and rabbit anti-SLC44A1 (1:100, 14687-1-AP; Proteintech).

Single-molecule in situ hybridization (smFISH)

Human intestinal tissues were fixed with 4% formaldehyde solution for 2 h at 4°C and dehydrated in 30% sucrose solution at 4°C overnight. Next, the tissue was embedded in optimal cutting temperature compound and stored at –80°C. The sections were prepared with vibrating blade microtome (HM650; Microm) and endogenous peroxidase blocking was performed by RNAscope Hydrogen Peroxide (322335; ACD) for 10 min at room temperature. Then, RNAscope Protease Plus (322331; ACD) was used for 10 min at 40°C before probe hybridization. ITLN1 probe (549701; ACD) and LYZ probe (421441; ACD) were hybridized for 2 h at 40°C, AMP 1–6, and signal detection was performed as described in the user manual (322350; ACD). Finally, the slides were

counterstained by 50% hematoxylin, and images were obtained by a Nikon 90i microscope (Nikon).

Quantitative RT-PCR

RNeasy Mini Kit (Qiagen) was used to extract total RNA, and cDNA was obtained by Revertra Ace (Toyobo). Then, real-time PCR reactions were performed in triplicate on a LightCycler 480 (Roche). Primers of selected genes are listed in Table S10.

Human intestinal organoid culture

Human intestinal tissue was washed in cold HBSS and removed muscle tissue. Then, epithelial tissue was cut into small pieces (5 mm) and incubated in 5 mM EDTA in HBSS for 30 min at 4°C. The pieces were transferred into cold HBSS and vigorously suspend to get fraction, and epithelial tissue was enriched through centrifugation (3 min at 1,000 rpm). Crypts were then embedded in Matrigel (BD Biosciences) and seeded on a 24-well plate. After polymerization, crypt culture medium (Advanced DMEM/F12 (12634028; Thermo Scientific) supplemented with penicillin/streptomycin (15140122; Thermo Scientific), GlutaMAX-I (35050061; Thermo Scientific), N2 (17502048; Thermo Scientific), B27 (17504044; Thermo Scientific), and N-acetylcysteine (Sigma-Aldrich) containing EGF (50 ng/ml; Invitrogen), Noggin (100 ng/ml; R&D), R-spondin1 (500 ng/ml; R&D), CHIR-99021 (5 μM; Selleck), A-83-01 (0.5 μM; Cayman), SB202190 (10 μM; Selleck), Gastrin (1 nM; Tocris), Y27632 (10 μM; Enzo), PGE2 (2.5 μM; Selleck), and Nicotinamide (10 mM; Sigma-Aldrich) was added.

Nutrient uptake assay

For amino acid uptake, 50 μl medium was selected 24 h later after the first passage, which then mixed with 200 μl -80°C methyl alcohol. The mixture was stored at -80°C for at least 2 h and then performed centrifugation (15 min at 12,000 rpm, 4°C). 100 μl supernatant was extracted to detect the amino acid changes compared with blank group by liquid chromatography mass spectrometry (LC-MS; Q Exactive; Thermo Scientific). To find out amino acid uptake per cell, organoids were incubated in Tryple (Invitrogen) for 20 min at 37°C to obtain a single-cell suspension. The cell suspension was stained by PI (5 μg/ml), and the live cell number was analyzed by FACS (BeckMan) for PI-negative cells. The amino acid uptake per cell was calculated by combining amino acid changes and cell number. For choline, succinic and citric acid uptake, 20 mM choline (C805027; MACLIN), succinic acid (S817854; MACLIN), and citric acid (C805019; MACLIN) were added to the crypt culture medium 8 h later after the first passage. 50 μl medium was selected 24 h later to detect uptake changes, and the cell number was counted by FACS for PI-negative cells as described above. Organic solute uptake per cell was calculated by combining amino acid changes and cell number. For sugar uptake, GlutaMAX-I was removed from the medium 8 h later after the first passage. Then, 20 mM fructose (S5176; Selleck), galactose (S3849; Selleck), and mannose (S5763; Selleck) were added to the medium, and 50 μl medium was selected 24 h later to detect uptake changes, and the cell number was counted by FACS for PI-negative cells as described above. Sugar uptake per cell was calculated by

combining amino acid changes and cell number. The fold change from ileum, colon, and rectum organoids was calculated by comparing with ileum organoids.

Statistical analysis

All experiments with quantitation were performed independently at least three times with three replicates within each experiment, and data are represented as mean ± SD. Then statistical differences were calculated using ordinary two-way ANOVA followed by Tukey's multiple comparisons test; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. All statistical analysis was performed with GraphPad Prism 7 (win).

Data availability

All data have been deposited in the GEO under accession no. GSE125970 and in the file "Single-cell transcriptome analysis of adult human ileum, colon and rectum" (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE125970>).

Code availability

R markdown scripts enabling the main steps of the analysis to be performed are available from the corresponding authors on reasonable request.

Online supplemental material

Fig. S1 shows general information of clinical samples and annotations of cell types of single-cell RNA-seq data. Fig. S2 shows expression patterns of cell markers. Fig. S3 shows characterization of stem cells, TA cells, progenitor cells, and transcription factor analysis. Fig. S4 shows expression patterns of transporter genes and validation by immunofluorescence or immunohistochemistry. Fig. S5 shows specific expression of NUSAP1⁺ cells in TA cells and *Itln1* and *Tff1* in mouse intestine. Table S1 shows all cell type-specific genes. Table S2 shows ileum, colon, and rectum cell type-specific genes. Table S3 shows stem cell and TA subset genes. Table S4 shows progenitor subset-specific genes. Table S5 shows enterocyte cell subset-specific genes. Table S6 shows enteroendocrine cell subset-specific genes. Table S7 shows PLC subset signature genes. Table S8 shows goblet cell subset signature genes. Table S9 shows signature genes involved in the cell cycle. Table S10 shows quantitative PCR primers.

Acknowledgments

We thank Drs. Ligong Chen and Xin Zhou for critical reading of the manuscript, Yuxin Sun for information consolidation, and the Metabolomics Facility at Tsinghua University for LC-MS analyses.

This work was supported by the National Key Research and Development Program of China (grant 2017YFA0103601 to Y.-G. Chen and grant 2016YFC0906001 to X. Yang) and the National Natural Science Foundation of China (grant 31330049 to Y.-G. Chen and grants 81472855 and 91540109 to X. Yang).

The authors declare no competing financial interests.

Author contributions: Y. Wang and Y.-G. Chen designed the study and analyzed the data; Y. Wang performed the experiments; W. Song and X. Yang performed the bioinformatics

analysis and analyzed the data; J. Wang and W. Fu provided samples and selected clinical information and analyzed the data; T. Wang and X. Xiong helped with functional experiments; Z. Qi helped with single-cell isolation; and Y. Wang, W. Song, X. Yang, and Y.-G. Chen wrote the manuscript.

Submitted: 20 June 2019

Revised: 28 August 2019

Accepted: 18 October 2019

References

- Aihara, E., K.A. Engevik, and M.H. Montrose. 2017. Trefoil Factor Peptides and Gastrointestinal Function. *Annu. Rev. Physiol.* 79:357–380. <https://doi.org/10.1146/annurev-physiol-021115-105447>
- Arora, K., J.M. Sequeira, and E.V. Quadros. 2017. Maternofetal transport of vitamin B₁₂: role of TCbIR/CD320 and megalin. *FASEB J.* 31:3098–3106. <https://doi.org/10.1096/fj.201700025R>
- Bogdan, A.R., M. Miyazawa, K. Hashimoto, and Y. Tsuji. 2016. Regulators of Iron Homeostasis: New Players in Metabolism, Cell Death, and Disease. *Trends Biochem. Sci.* 41:274–286. <https://doi.org/10.1016/j.tibs.2015.11.012>
- Buettner, F., K.N. Natarajan, F.P. Casale, V. Proserpio, A. Scialdone, F.J. Theis, S.A. Teichmann, J.C. Marioni, and O. Stegle. 2015. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* 33:155–160. <https://doi.org/10.1038/nbt.3102>
- Butler, A., P. Hoffman, P. Smibert, E. Papalexi, and R. Satija. 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36:411–420. <https://doi.org/10.1038/nbt.4096>
- Choi, E.K., T.T. Nguyen, N. Gupta, S. Iwase, and Y.A. Seo. 2018. Functional analysis of SLC39A8 mutations and their implications for manganese deficiency and mitochondrial disorders. *Sci. Rep.* 8:3163. <https://doi.org/10.1038/s41598-018-21464-0>
- Clevers, H., and E. Batlle. 2013. SnapShot: the intestinal crypt. *Cell.* 152: 1198–1198.e2. <https://doi.org/10.1016/j.cell.2013.02.030>
- Clevers, H.C., and C.L. Bevins. 2013. Paneth cells: maestros of the small intestinal crypts. *Annu. Rev. Physiol.* 75:289–311. <https://doi.org/10.1146/annurev-physiol-030212-183744>
- Coady, M.J., B. Wallendorff, D.G. Gagnon, and J.Y. Lapointe. 2002. Identification of a novel Na⁺/myo-inositol cotransporter. *J. Biol. Chem.* 277: 35219–35224. <https://doi.org/10.1074/jbc.M204321200>
- Dalerba, P., T. Kalisky, D. Sahoo, P.S. Rajendran, M.E. Rothenberg, A.A. Leyrat, S. Sim, J. Okamoto, D.M. Johnston, D. Qian, et al. 2011. Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat. Biotechnol.* 29:1120–1127. <https://doi.org/10.1038/nbt.2038>
- de Oliveira, M.R. 2015. Vitamin A and Retinoids as Mitochondrial Toxicants. *Oxid. Med. Cell. Longev.* 2015:140267. <https://doi.org/10.1155/2015/140267>
- Di Ciaula, A., G. Garruti, R. Lunardi Baccetto, E. Molina-Molina, L. Bonfrate, D.Q. Wang, and P. Portincasa. 2017. Bile Acid Physiology. *Ann. Hepatol.* 16:s4–s14. <https://doi.org/10.5604/01.3001.0010.5493>
- Eide, D.J. 2004. The SLC39 family of metal ion transporters. *Pflugers Arch.* 447: 796–800. <https://doi.org/10.1007/s00424-003-1074-3>
- Gao, S., L. Yan, R. Wang, J. Li, J. Yong, X. Zhou, Y. Wei, X. Wu, X. Wang, X. Fan, et al. 2018. Tracing the temporal-spatial transcriptome landscapes of the human fetal digestive tract using single-cell RNA-sequencing. *Nat. Cell Biol.* 20:721–734. <https://doi.org/10.1038/s41556-018-0105-4>
- Gehart, H., and H. Clevers. 2018. Tales from the crypt: new insights into intestinal stem cells. *Nat. Rev. Gastroenterol. Hepatol.*
- Georgiev, D., D. Arion, J.F. Enwright, M. Kikuchi, Y. Minabe, J.P. Corradi, D.A. Lewis, and T. Hashimoto. 2014. Lower gene expression for KCNS3 potassium channel subunit in parvalbumin-containing neurons in the prefrontal cortex in schizophrenia. *Am. J. Psychiatry.* 171:62–71. <https://doi.org/10.1176/appi.ajp.2013.13040468>
- Gierl, M.S., N. Karoulias, H. Wende, M. Strehle, and C. Birchmeier. 2006. The zinc-finger factor *Insm1* (IA-1) is essential for the development of pancreatic beta cells and intestinal endocrine cells. *Genes Dev.* 20: 2465–2478. <https://doi.org/10.1101/gad.381806>

- Goncalves, A., S. Roi, M. Nowicki, A. Dhaussy, A. Huertas, M.J. Amiot, and E. Reboul. 2015. Fat-soluble vitamin intestinal absorption: absorption sites in the intestine and interactions for absorption. *Food Chem.* 172:155–160. <https://doi.org/10.1016/j.foodchem.2014.09.021>
- Grün, D., A. Lyubimova, L. Kester, K. Wiebrands, O. Basak, N. Sasaki, H. Clevers, and A. van Oudenaarden. 2015. Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature.* 525:251–255. <https://doi.org/10.1038/nature14966>
- Haber, A.L., M. Biton, N. Rogel, R.H. Herbst, K. Shekhar, C. Smillie, G. Burgin, T.M. Delorey, M.R. Howitt, Y. Katz, et al. 2017. A single-cell survey of the small intestinal epithelium. *Nature.* 551:333–339. <https://doi.org/10.1038/nature24489>
- Han, G., Z. Wei, H. Cui, W. Zhang, X. Wei, Z. Lu, and X. Bai. 2018a. NUSAP1 gene silencing inhibits cell proliferation, migration and invasion through inhibiting DNMT1 gene expression in human colorectal cancer. *Exp. Cell Res.* 367:216–221. <https://doi.org/10.1016/j.yexcr.2018.03.039>
- Han, X., R. Wang, Y. Zhou, L. Fei, H. Sun, S. Lai, A. Saadatpour, Z. Zhou, H. Chen, F. Ye, et al. 2018b. Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell.* 173:1307. <https://doi.org/10.1016/j.cell.2018.05.012>
- Hardoon, D.R., S. Szedmak, and J. Shawe-Taylor. 2004. Canonical correlation analysis: an overview with application to learning methods. *Neural Comput.* 16:2639–2664. <https://doi.org/10.1162/0899766042321814>
- Heneghan, J.F., A. Akhavein, M.J. Salas, B.E. Shmukler, L.P. Karniski, D.H. Vidorpe, and S.L. Alper. 2010. Regulated transport of sulfate and oxalate by SLC26A2/DTDST. *Am. J. Physiol. Cell Physiol.* 298: C1363–C1375. <https://doi.org/10.1152/ajpcell.00004.2010>
- Hummeler, E., and F. Beermann. 2000. Scnn1 sodium channel gene family in genetically engineered mice. *J. Am. Soc. Nephrol.* 11(Suppl 16):S129–S134.
- Jones, R.J., S. Dickerson, P.M. Bhende, H.J. Delecluse, and S.C. Kenney. 2007. Epstein-Barr virus lytic infection induces retinoic acid-responsive genes through induction of a retinol-metabolizing enzyme, DHRS9. *J. Biol. Chem.* 282:8317–8324. <https://doi.org/10.1074/jbc.M608667200>
- Kinchen, J., H.H. Chen, K. Parikh, A. Antanaviciute, M. Jagielowicz, D. Fawcner-Corbett, N. Ashley, L. Cubitt, E. Mellado-Gomez, M. Attar, et al. 2018. Structural Remodeling of the Human Colonic Mesenchyme in Inflammatory Bowel Disease. *Cell.* 175:372–386.e17. <https://doi.org/10.1016/j.cell.2018.08.067>
- Kobayashi, K., D.S. Sinasac, M. Iijima, A.P. Boright, L. Begum, J.R. Lee, T. Yasuda, S. Ikeda, R. Hirano, H. Terazono, et al. 1999. The gene mutated in adult-onset type II citrullinaemia encodes a putative mitochondrial carrier protein. *Nat. Genet.* 22:159–163. <https://doi.org/10.1038/9667>
- Kunzelmann, K., and M. Mall. 2002. Electrolyte transport in the mammalian colon: mechanisms and implications for disease. *Physiol. Rev.* 82: 245–289. <https://doi.org/10.1152/physrev.00026.2001>
- Lin, L., S.W. Yee, R.B. Kim, and K.M. Giacomini. 2015. SLC transporters as therapeutic targets: emerging opportunities. *Nat. Rev. Drug Discov.* 14: 543–560. <https://doi.org/10.1038/nrd4626>
- Macosko, E.Z., A. Basu, R. Satija, J. Nemes, K. Shekhar, M. Goldman, I. Tirosh, A.R. Bialas, N. Kamitaki, E.M. Martnersteck, et al. 2015. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell.* 161:1202–1214. <https://doi.org/10.1016/j.cell.2015.05.002>
- Mantani, Y., M. Nishida, H. Yuasa, K. Yamamoto, E. Takahara, T. Omotehara, K.G. Udayanga, J. Kawano, T. Yokoyama, N. Hoshi, and H. Kitagawa. 2014. Ultrastructural and histochemical study on the Paneth cells in the rat ascending colon. *Anat. Rec. (Hoboken).* 297:1462–1471. <https://doi.org/10.1002/ar.22937>
- Mayer, C., C. Hafemeister, R.C. Bandler, R. Machold, R. Batista Brito, X. Jaglin, K. Allaway, A. Butler, G. Fishell, and R. Satija. 2018. Developmental diversification of cortical inhibitory interneurons. *Nature.* 555:457–462. <https://doi.org/10.1038/nature25999>
- Moor, A.E., Y. Harnik, S. Ben-Moshe, E.E. Massasa, M. Rozenberg, R. Eilam, K. Bahar Halpern, and S. Itzkovitz. 2018. Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation along the Intestinal Villus Axis. *Cell.* 175:1156–1167.e15. <https://doi.org/10.1016/j.cell.2018.08.063>
- Murphy, K.G., and S.R. Bloom. 2006. Gut hormones and the regulation of energy homeostasis. *Nature.* 444:854–859. <https://doi.org/10.1038/nature05484>
- Naik, R., and S. Galande. 2019. SATB family chromatin organizers as master regulators of tumor progression. *Oncogene.* 38:1989–2004. <https://doi.org/10.1038/s41388-018-0541-4>
- Nicklin, P., P. Bergman, B. Zhang, E. Triantafellow, H. Wang, B. Nyfeler, H. Yang, M. Hild, C. Kung, C. Wilson, et al. 2009. Bidirectional transport of amino acids regulates mTOR and autophagy. *Cell.* 136:521–534. <https://doi.org/10.1016/j.cell.2008.11.044>

- Nilsson, R., I.J. Schultz, E.L. Pierce, K.A. Soltis, A. Naranuntarat, D.M. Ward, J.M. Baughman, P.N. Paradkar, P.D. Kingsley, V.C. Culotta, et al. 2009. Discovery of genes essential for heme biosynthesis through large-scale gene expression analysis. *Cell Metab.* 10:119–130. <https://doi.org/10.1016/j.cmet.2009.06.012>
- Peterson, L.W., and D. Artis. 2014. Intestinal epithelial cells: regulators of barrier function and immune homeostasis. *Nat. Rev. Immunol.* 14: 141–153. <https://doi.org/10.1038/nri3608>
- Qi, Z., Y. Li, B. Zhao, C. Xu, Y. Liu, H. Li, B. Zhang, X. Wang, X. Yang, W. Xie, et al. 2017. BMP restricts stemness of intestinal Lgr5⁺ stem cells by directly suppressing their signature genes. *Nat. Commun.* 8:13824. <https://doi.org/10.1038/ncomms13824>
- Raja, M., T. Puntheeranurak, P. Hinterdorfer, and R. Kinne. 2012. SLC5 and SLC2 transporters in epithelia-cellular role and molecular mechanisms. *Curr. Top. Membr.* 70:29–76. <https://doi.org/10.1016/B978-0-12-394316-3.00002-8>
- Reboul, E. 2015. Intestinal absorption of vitamin D: from the meal to the enterocyte. *Food Funct.* 6:356–362. <https://doi.org/10.1039/C4FO00579A>
- Rothenberg, M.E., Y. Nusse, T. Kalisky, J.J. Lee, P. Dalerba, F. Scheeren, N. Lobo, S. Kulkarni, S. Sim, D. Qian, et al. 2012. Identification of a cKit(+) colonic crypt base secretory cell that supports Lgr5(+) stem cells in mice. *Gastroenterology.* 142:1195–1205.e6. <https://doi.org/10.1053/j.gastro.2012.02.006>
- Sanger, G.J., and K. Lee. 2008. Hormones of the gut-brain axis as targets for the treatment of upper gastrointestinal disorders. *Nat. Rev. Drug Discov.* 7:241–254. <https://doi.org/10.1038/nrd2444>
- Satija, R., J.A. Farrell, D. Gennert, A.F. Schier, and A. Regev. 2015. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33: 495–502. <https://doi.org/10.1038/nbt.3192>
- Sato, T., J.H. van Es, H.J. Snippert, D.E. Stange, R.G. Vries, M. van den Born, N. Barker, N.F. Shroyer, M. van de Wetering, and H. Clevers. 2011. Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature.* 469:415–418. <https://doi.org/10.1038/nature09637>
- Schuijers, J., J.P. Junker, M. Mokry, P. Hatzis, B.K. Koo, V. Sasselli, L.G. van der Flier, E. Cuppen, A. van Oudenaarden, and H. Clevers. 2015. Ascl2 acts as an R-spondin/Wnt-responsive switch to control stemness in intestinal crypts. *Cell Stem Cell.* 16:158–170. <https://doi.org/10.1016/j.stem.2014.12.006>
- Song, Z. 2013. Roles of the nucleotide sugar transporters (SLC35 family) in health and disease. *Mol. Aspects Med.* 34:590–600. <https://doi.org/10.1016/j.mam.2012.12.004>
- Sugimoto, S., Y. Ohta, M. Fujii, M. Matano, M. Shimokawa, K. Nanki, S. Date, S. Nishikori, Y. Nakazato, T. Nakamura, et al. 2018. Reconstruction of the Human Colon Epithelium In Vivo. *Cell Stem Cell.* 22:171–176.e5. <https://doi.org/10.1016/j.stem.2017.11.012>
- Suhre, K., H. Wallaschofski, J. Raffler, N. Friedrich, R. Haring, K. Michael, C. Wasner, A. Krebs, F. Kronenberg, D. Chang, et al. 2011. A genome-wide association study of metabolic traits in human urine. *Nat. Genet.* 43: 565–569. <https://doi.org/10.1038/ng.837>
- Tabula Muris Consortium. 2018. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature.* 562:367–372. <https://doi.org/10.1038/s41586-018-0590-4>
- Tappenden, K.A. 2014. Pathophysiology of short bowel syndrome: considerations of resected and residual anatomy. *JPEN J. Parenter. Enteral Nutr.* 38(1, Suppl):14S–22S. <https://doi.org/10.1177/0148607113520005>
- Tazawa, S., T. Yamato, H. Fujikura, M. Hiratochi, F. Itoh, M. Tomae, Y. Takemura, H. Maruyama, T. Sugiyama, A. Wakamatsu, et al. 2005. SLC5A9/SGLT4, a new Na⁺-dependent glucose transporter, is an essential transporter for mannose, 1,5-anhydro-D-glucitol, and fructose. *Life Sci.* 76:1039–1050. <https://doi.org/10.1016/j.lfs.2004.10.016>
- Tirosh, I., B. Izar, S.M. Prakadan, M.H. Wadsworth II, D. Treacy, J.J. Trombetta, A. Rotem, C. Rodman, C. Lian, G. Murphy, et al. 2016. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science.* 352:189–196. <https://doi.org/10.1126/science.aad0501>
- Torrents, D., J. Mykkänen, M. Pineda, L. Feliubadaló, R. Estévez, R. de Cid, P. Sanjurjo, A. Zorzano, V. Nunes, K. Huoponen, et al. 1999. Identification of SLC7A7, encoding y⁺LAT-1, as the lysinuric protein intolerance gene. *Nat. Genet.* 21:293–296. <https://doi.org/10.1038/6809>
- Traiffort, E., S. O'Regan, and M. Ruat. 2013. The choline transporter-like family SLC44: properties and roles in human diseases. *Mol. Aspects Med.* 34:646–654. <https://doi.org/10.1016/j.mam.2012.10.011>
- Tremaroli, V., and F. Bäckhed. 2012. Functional interactions between the gut microbiota and host metabolism. *Nature.* 489:242–249. <https://doi.org/10.1038/nature11552>
- Verkman, A.S., M.O. Anderson, and M.C. Papadopoulos. 2014. Aquaporins: important but elusive drug targets. *Nat. Rev. Drug Discov.* 13:259–277. <https://doi.org/10.1038/nrd4226>
- Wesener, D.A., A. Dugan, and L.L. Kiessling. 2017. Recognition of microbial glycans by soluble human lectins. *Curr. Opin. Struct. Biol.* 44:168–178. <https://doi.org/10.1016/j.sbi.2017.04.002>
- Wood, I.S., and P. Trayhurn. 2003. Glucose transporters (GLUT and SGLT): expanded families of sugar transport proteins. *Br. J. Nutr.* 89:3–9. <https://doi.org/10.1079/BJN2002763>
- Wright, E.M. 2013. Glucose transport families SLC5 and SLC50. *Mol. Aspects Med.* 34:183–196. <https://doi.org/10.1016/j.mam.2012.11.002>
- Zhang, Z., and Z. Liu. 2016. Paneth cells: the hub for sensing and regulating intestinal flora. *Sci. China Life Sci.* 59:463–467. <https://doi.org/10.1007/s11427-016-5018-5>
- Zhang, P., J. Gao, C. Pu, G. Feng, L. Wang, L. Huang, and Y. Zhang. 2017. ApoM/HDL-C and apoM/apoA-I ratios are indicators of diabetic nephropathy in healthy controls and type 2 diabetes mellitus. *Clin. Chim. Acta.* 466:31–37. <https://doi.org/10.1016/j.cca.2017.01.006>
- Zhang, Y., Y. Zhang, K. Sun, Z. Meng, and L. Chen. 2019. The SLC transporter in nutrient and metabolic sensing, regulation, and drug development. *J. Mol. Cell Biol.* 11:1–13. <https://doi.org/10.1093/jmcb/mjy052>
- Zhou, W., S.D. Ye, C. Chen, and W. Wang. 2018. Involvement of RBP4 in Diabetic Atherosclerosis and the Role of Vitamin D Intervention. *J. Diabetes Res.* 2018:7329861. <https://doi.org/10.1155/2018/7329861>
- Zorn, A.M., and J.M. Wells. 2009. Vertebrate endoderm development and organ formation. *Annu. Rev. Cell Dev. Biol.* 25:221–251. <https://doi.org/10.1146/annurev.cellbio.042308.113344>