nature cell biology

RESOURCE

https://doi.org/10.1038/s41556-018-0105-4

Corrected: Publisher Correction

Tracing the temporal-spatial transcriptome landscapes of the human fetal digestive tract using single-cell RNA-sequencing

Shuai Gao[1,9], Liying Yan[1,2,9], Rui Wang[1,3,9], Jingyun Li[4,5,9], Jun Yong[1,2], Xin Zhou[6], Yuan Wei[1,2], Xinglong Wu[4,5], Xiaoye Wang[1,2], Xiaoying Fan[1], Jie Yan[1,2], Xu Zhi[1,2], Yun Gao[1], Hongshan Guo[1], Xiao Jin[7], Wendong Wang[6], Yunuo Mao[1,2], Fengchao Wang[8], Lu Wen[1], Wei Fu[6], Hao Ge[7]*, Jie Qiao[1,2,4]* and Fuchou Tang[1,3,4]*

The development of the digestive tract is critical for proper food digestion and nutrient absorption. Here, we analyse the main organs of the digestive tract, including the oesophagus, stomach, small intestine and large intestine, from human embryos between 6 and 25 weeks of gestation as well as the large intestine from adults using single-cell RNA-seq analyses. In total, 5,227 individual cells are analysed and 40 cell types clearly identified. Their crucial biological features, including developmental processes, signalling pathways, cell cycle, nutrient digestion and absorption metabolism, and transcription factor networks, are systematically revealed. Moreover, the differentiation and maturation processes of the large intestine are thoroughly investigated by comparing the corresponding transcriptome profiles between embryonic and adult stages. Our work offers a rich resource for investigating the gene regulation networks of the human fetal digestive tract and adult large intestine at single-cell resolution.

The gastrointestinal (GI) tract, also called the digestive tract, is a highly organized complex organ system in humans[1–3]. The GI tract performs many important functions, including food digestion[4], nutrient absorption[5], endocrine secretion[6] and resistance to microorganism invasion[7]. At the fourth week of human embryo gestation, the primitive gut can be divided into three distinct regions—the foregut, midgut and hindgut—which will contribute to different components of the GI tract. The human GI tract further differentiates into the oesophagus, stomach, small intestine (S-Intes) and large intestine (L-Intes)[2,3]. Disorders of the digestive tract during early embryo development may lead to various neonatal GI diseases, such as congenital intestinal atresia and necrotizing enterocolitis[8,9]. GI cancer accounts for approximately 35% of global cancer-related mortalities[10,11]. Therefore, systematic investigations of the gene expression network during the development of the human GI tract are urgently needed.

Recently, we and others have developed a wide variety of single-cell RNA-seq methods that can accurately investigate the transcriptomes of mammalian cells at single-cell resolution[12–19]. The characteristics of each cell can be identified according to its transcriptional profile[20,21]. In addition, a growing toolkit of bioinformatics analysis algorithms, such as Seurat[22], StemID[23,24] and Monocle[25,26], has greatly improved the analyses of single-cell RNA-seq data for the de novo identification of the stem cell signature and elucidation of the transcriptome dynamics during the process of cell differentiation[24,27]. Using single-cell RNA-seq, breakthroughs concerning the stomach and intestine have emerged[24,28,29]. However, the transcriptomic regulations in human embryonic digestive systems have not been systematically analysed, reflecting the limited availability of clinical samples.

Results

Single-cell transcriptome analysis of the human fetal digestive tract. To systematically investigate the gene expression pattern during human fetal GI tract development, the digestive tracts of 15 human embryos ranging from 6 weeks to 25 weeks were analysed using single-cell RNA-seq technique (Fig. 1a). The oesophagus, stomach, S-Intes and L-Intes—the four main organs of the digestive tract—were isolated from 14 of these 15 embryos. The four dissected organs were dissociated, and individual cells were isolated. In total, 5,290 single cells covering different cell types were selected (Supplementary Table 1). On average, 2.9 million reads were sequenced for each individual cell, generating 5.0 Tb of data for these single-cell samples. After rigorous filtration[30,31], 4,089 individual cells were obtained for subsequent analysis (Supplementary Table 1). On average, each single cell expressed 5,098 genes (median number) and 276,247 mRNA molecules (median number).

After combining t-distributed stochastic neighbour embedding (t-SNE) and affinity propagation clustering analysis (Supplementary Table 2), we excluded 331 immune cells (Groups 1, 25, 45 and 46),

[1]Department of Obstetrics and Gynecology, Beijing Advanced Innovation Center for Genomics, College of Life Sciences, Third Hospital, Peking University, Beijing, China. [2]Biomedical Institute for Pioneering Investigation via Convergence & Key Laboratory of Assisted Reproduction, Ministry of Education, Beijing, China. [3]Ministry of Education Key Laboratory of Cell Proliferation and Differentiation, Peking University, Beijing, China. [4]Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China. [5]Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. [6]Department of General Surgery, Peking University Third Hospital, Beijing, China. [7]Beijing International Center for Mathematical Research, Peking University, Beijing, China. [8]National Institute of Biological Sciences (NIBS), Beijing, China. [9]These authors contributed equally: Shuai Gao, Liying Yan, Rui Wang and Jingyun Li
*e-mail: haoge@pku.edu.cn; jie.qiao@263.net; tangfuchou@pku.edu.cn

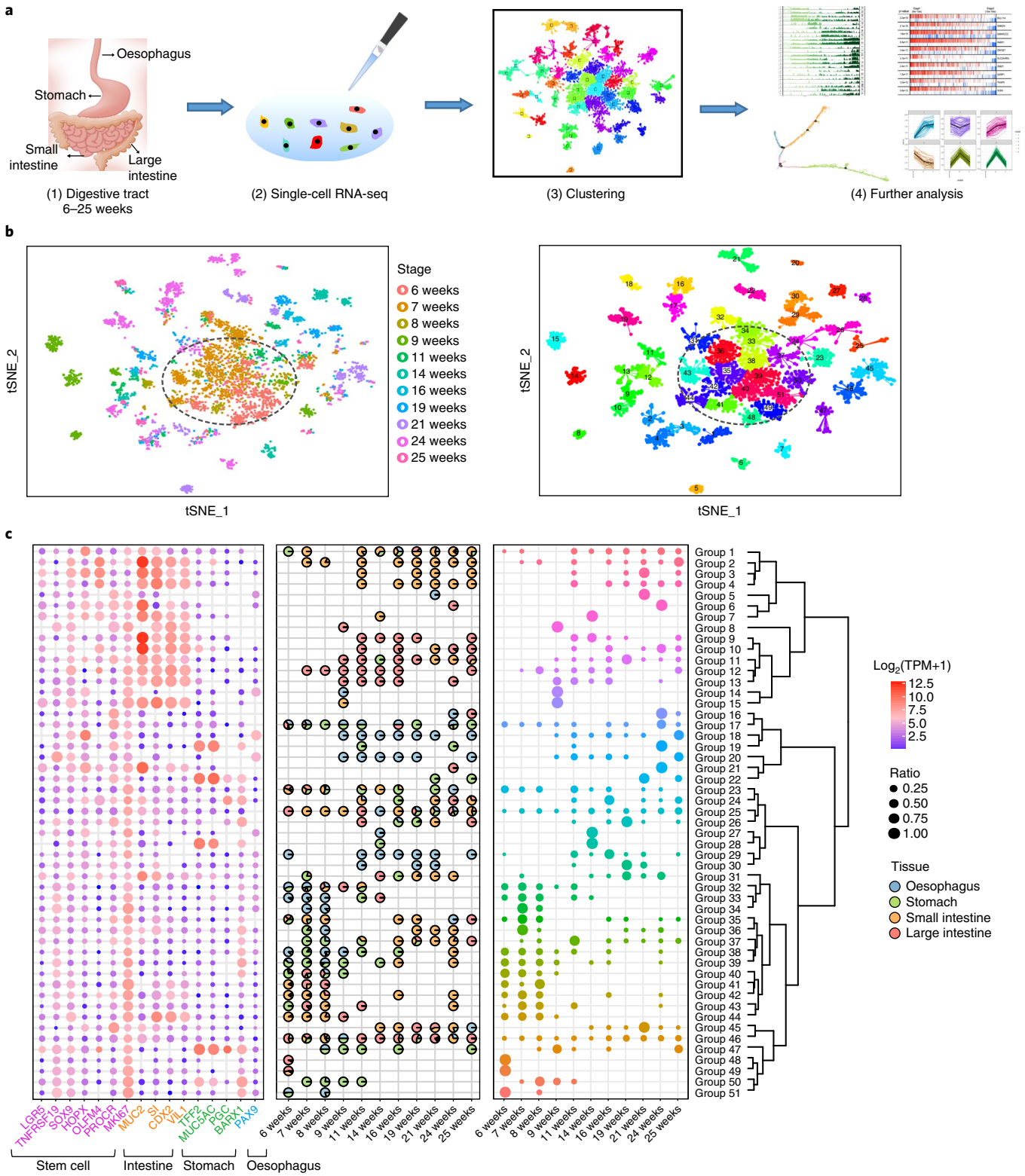


Fig. 1 | Global analysis of single-cell expression profiles of human fetal digestive system. a, Overview of the experimental and bioinformatics analysis workflow. **b**, Left, t-SNE analysis of the transcriptomes of all stage digestive tract cells. Right, unsupervised clustering of all digestive tract cells using affinity propagation. Each cell is coloured according to the gestation stage from 6 weeks to 25 weeks. Most cells within the black dashed oval belong to embryos at 6–10 weeks. In total, 4,089 cells were analysed by single-cell RNA sequencing. **c**, Hierarchical clustering of identified groups in the t-SNE analysis. Left, Average expression level of canonical markers in each group. The colour of each dot represents the average expression level from low (blue) to high (red), and the size of each dot represents the ratio of the positive cells for each gene. Middle, Pie plots showing the organ source of cells for each group at different developmental stages. Right, Dot plot showing the stage distribution of each group.

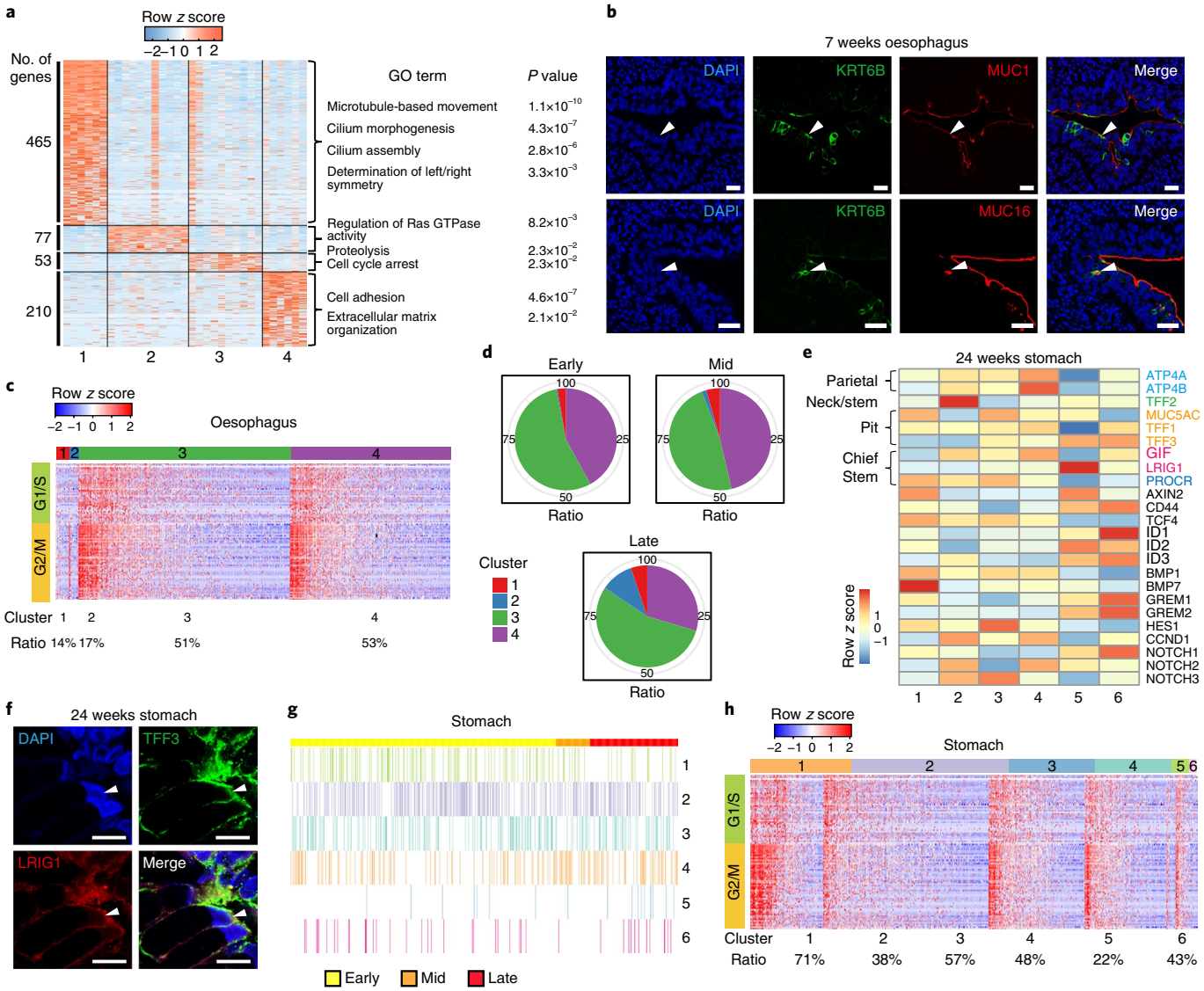


Fig. 2 | Cell type identification and development tracking of human fetal oesophagus and stomach. a, Heatmap and GO-enriched terms of cell-type-specific genes of a 25 week oesophagus. $n = 44$ cells. The colour key from blue to orange indicates the relative expression levels from low to high, respectively. **b**, Co-immunostaining of KRT6B with MUC1 and co-immunostaining of KRT6B with MUC16 in the 7 weeks oesophagus were performed to verify the existence of the oesophagus Cluster 2 cell type. White arrowheads indicate KRT6B$^+$MUC1$^+$ and KRT6B$^+$MUC16$^+$ cells. Scale bars, 25 μm. All experiments were independently repeated twice, with similar results. **c**, Heatmap showing expression levels of cell-cycle-related genes in each oesophagus cluster. **d**, Pie plots showing the ratio of each cell type at different developmental stages in the oesophagus. In **c** and **d**, 773 cells were analysed. **e**, Heatmap showing average expression levels of well-defined cell type markers and signalling-pathway-related genes in each cluster (containing at least five cells) of a 24 week stomach. $n = 98$ cells. Genes with expression exceeding $2(\log_2(\mathrm{TPM}+1))$ in at least one cluster are shown. The colour key from blue to red indicates relative expression levels from low to high. **f**, Immunostaining of TFF3 (green) and LRIG1 (red) in 24 weeks stomach to verify the cell type of Cluster 5 shown in **e**. White arrowheads indicate TFF3$^+$LRIG1$^+$ cells. Scale bars, 25 μm. The experiment was independently repeated twice, with similar results. **g**, Distribution of cell types along the stomach differentiation axis. Cells are first ordered according to developmental stage and then based on pseudotime. **h**, Heatmap showing expression levels of cell cycle-related genes in each stomach cluster. In **c** and **h**, cells are ordered according to the average expression levels of the cell-cycle-related genes within each cluster. The colour key from blue to red indicates expression levels from low to high. The cell cycle index of each cell type is shown at the bottom of the heatmap. In **g** and **h**, 854 cells were analysed.

43 muscle cells (Group 5), 103 haemocytes (Groups 23) and 268 endothelial cells (Groups 16, 17 and 26) from subsequent analyses according to the expression of cell-type-specific markers (Supplementary Fig. 1a-d). Our data revealed that most of the individual cells from the four organs between 6 and 10 weeks had the common features of being highly proliferating and progenitor-like (Fig. 1b and Supplementary Fig. 1e,f). However, after 11 weeks, cells from different digestive tract organs were clearly separated, indicating that these cells had further differentiated into distinct cell types (Fig. 1c). Most of the cells in Groups 1 to 5 were from the S-Intes, whereas L-Intes cells were primarily distributed in Groups 9 to 13, as confirmed by the high expression of intestine-specific marker genes[32] such as *MUC2* and *CDX2*. Oesophageal cells between 11 and 25 weeks were primarily distributed in Groups 17 to 20, with high expression of oesophagus marker genes[33,34] such as *PAX9*. Cells in Groups 19, 22, 28 and 47 were primarily from the stomach, as confirmed by the specific expression of

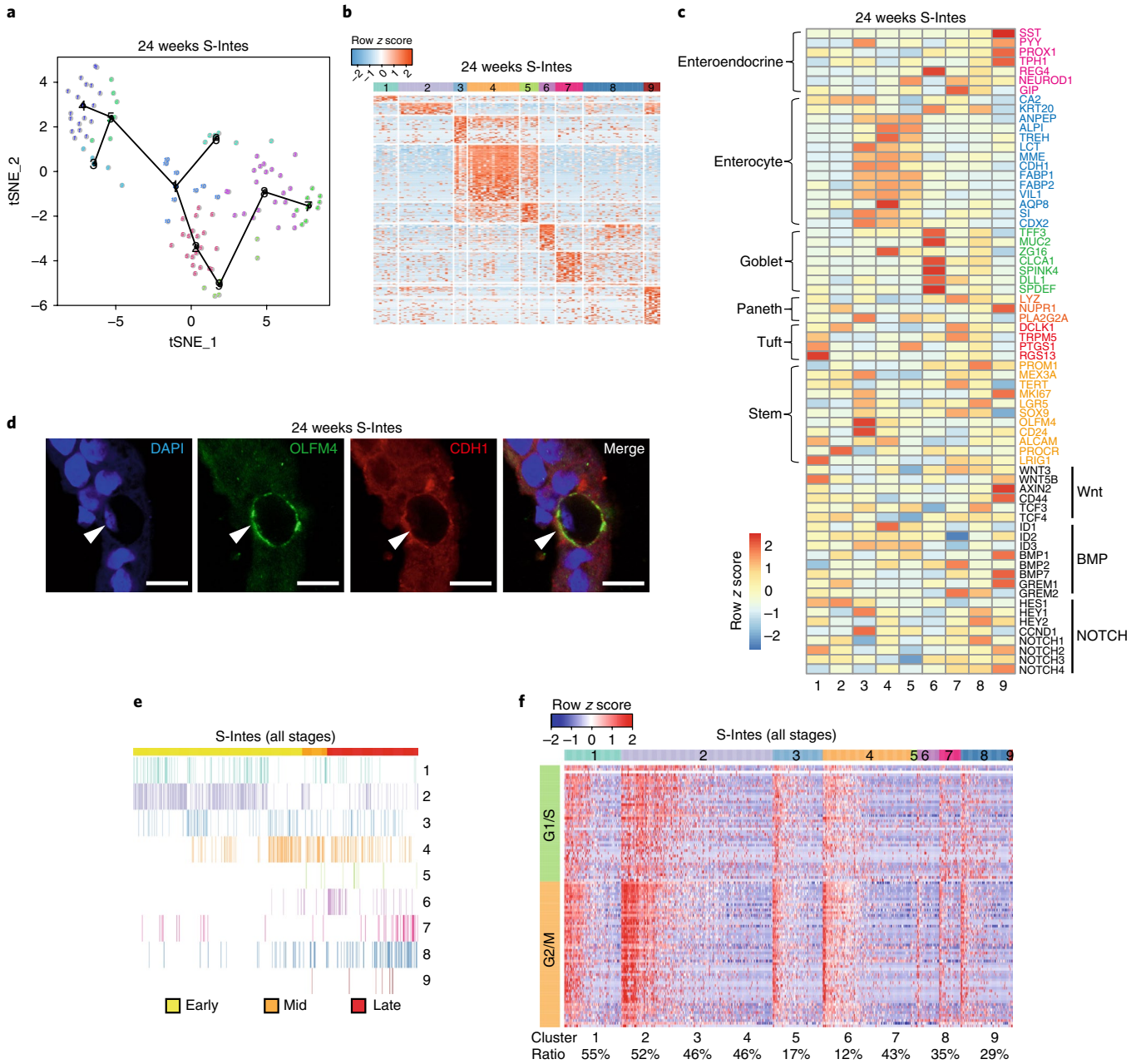


Fig. 3 | Cell type identification and development tracking of human fetal small intestine. a, Inter-cluster links of S-Intes clusters visualized in two-dimensional coordinates. The black solid line indicates a minimum spanning tree connecting the cluster centres. Clusters containing fewer than five cells are not shown. **b**, Heatmap showing the cluster-specific genes of 24 week S-Intes. The colour key from blue to orange indicates relative expression levels from low to high. In **a** and **b**, 88 cells were analysed. **c**, Heatmap showing the averaged expression of well-known cell type specific markers and selected-signalling-pathway-related genes in the nine clusters of 24 weeks S-Intes. $n = 88$ cells. Genes with expression exceed $2(\log_2(\text{TPM}+1))$ in at least one cluster are shown. Colour from blue to red represents expression levels from low to high. **d**, Immunostaining of OLFM4 (green) and CDH1 (red) in 24 weeks S-Intes to verify the existence of S-Intes cluster 3 and 4 cells. White arrowheads indicates OLFM4$^+$CDH1$^+$ cells. Scale bars, 25 μm. The experiment was independently repeated twice, with similar results. **e**, Distribution of all S-Intes cell types along the S-Intes differentiation axis. Cells are first ordered based on developmental stage and then according to pseudotime. **f**, Heatmap showing expression levels of cell-cycle-related genes in each S-Intes cluster. Cells are ordered according to average expression level of cell-cycle-related genes within each cell type. The colour key from blue to red indicates expression levels from low to high. The cell cycle index of each cell type is shown at the bottom of the heatmap. Progenitor cell types (Clusters 1 to 4) belonging to the early stage show higher ratios of actively cycling cells than differentiated cell types. In **e** and **f**, 868 cells were analysed.

stomach marker genes[35,36] such as *TFF2* and *MUC5AC*. Interestingly, adult stem cell markers *LGR5*, *SOX9*, *OLFM4* and *HOPX*[37,38] were highly expressed in most cells after 11 weeks. Overall, t-SNE analysis revealed the general dynamic changes in the transcriptomic patterns of these four digestive tract organs during human fetal development.

Pseudotime analyses of temporal gene expression patterns of the fetal digestive tract. By setting the embryonic developmental stage for each individual cell, we can accurately explore the maturation speed, cell type differentiation and transcriptome patterns of different organs. Using Monocle2, we constructed a pseudotime for

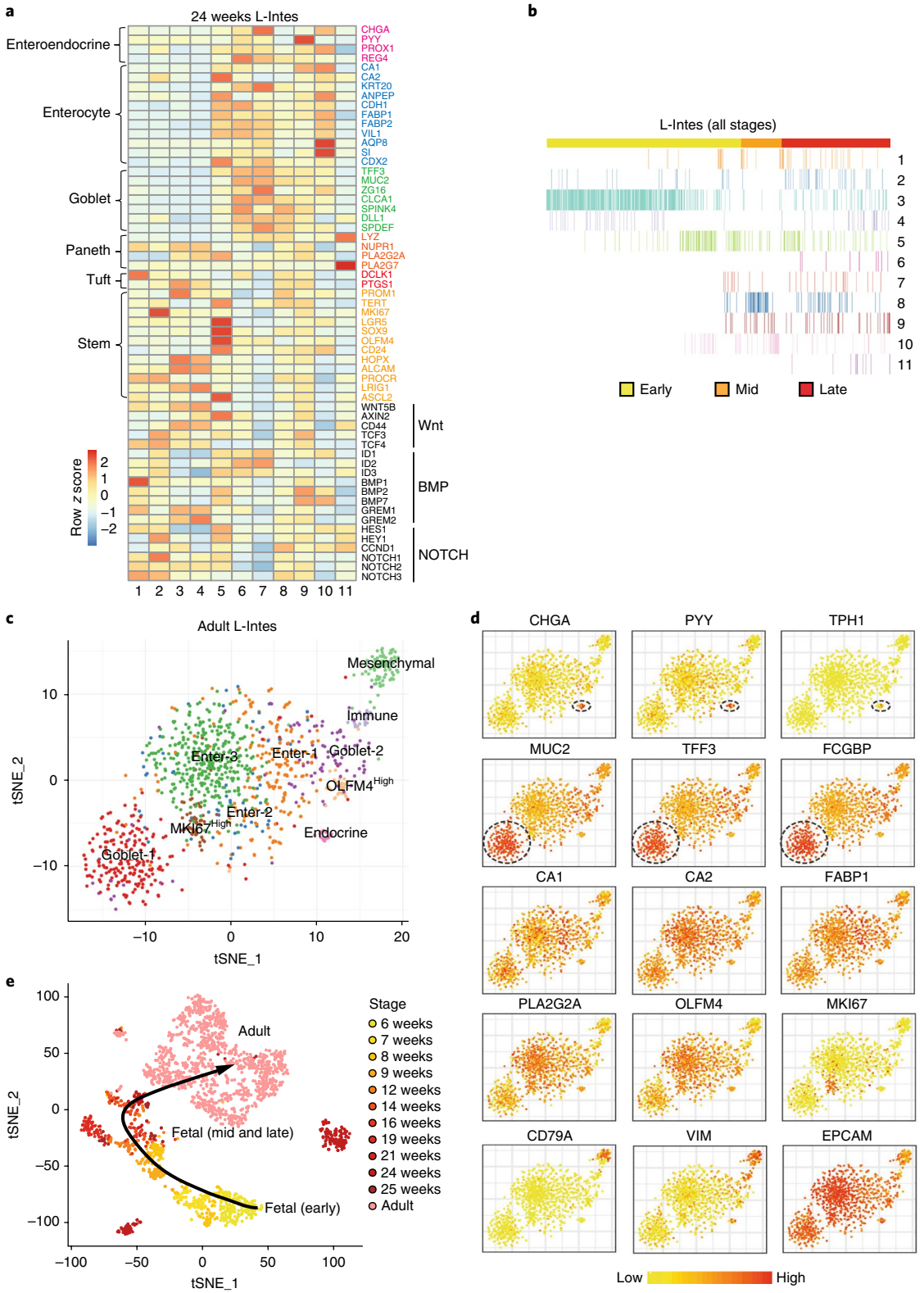


Fig. 4 | Cell type identification of fetal and adult large intestine and comparison of the cell type features between fetal and adult stages. a, Heatmap showing averaged expression of well-known cell type markers and selected signalling-pathway-related genes in the 11 clusters of the 24 weeks L-Intes. $n = 175$ cells. Genes with expression exceeding $2(\log_2(\text{TPM}+1))$ in at least one cluster are shown. Colours from blue to red represent expression levels from low to high. **b**, Distribution of all L-Intes cell types along the L-Intes differentiation axis. Cells are first ordered based on developmental stage and then according to pseudotime. $n = 849$ cells. **c**, t-SNE plot of all adult L-Intes cells. $n = 1{,}138$ cells. Each colour represents one cell cluster, and each point represents one cell. **d**, Expression levels of selected well-known marker genes are plotted onto the t-SNE map. $n = 1{,}138$ cells. Colour key from yellow to red indicates relative expression levels from low to high. Dashed ovals highlight the cells with the high expression levels of selected well-known marker genes. **e**, t-SNE plot of all stages of fetal L-Intes cells and all adult L-Intes cells. $n = 2{,}218$ cells. Each point represents one individual cell, and the points are coloured according to development stage.

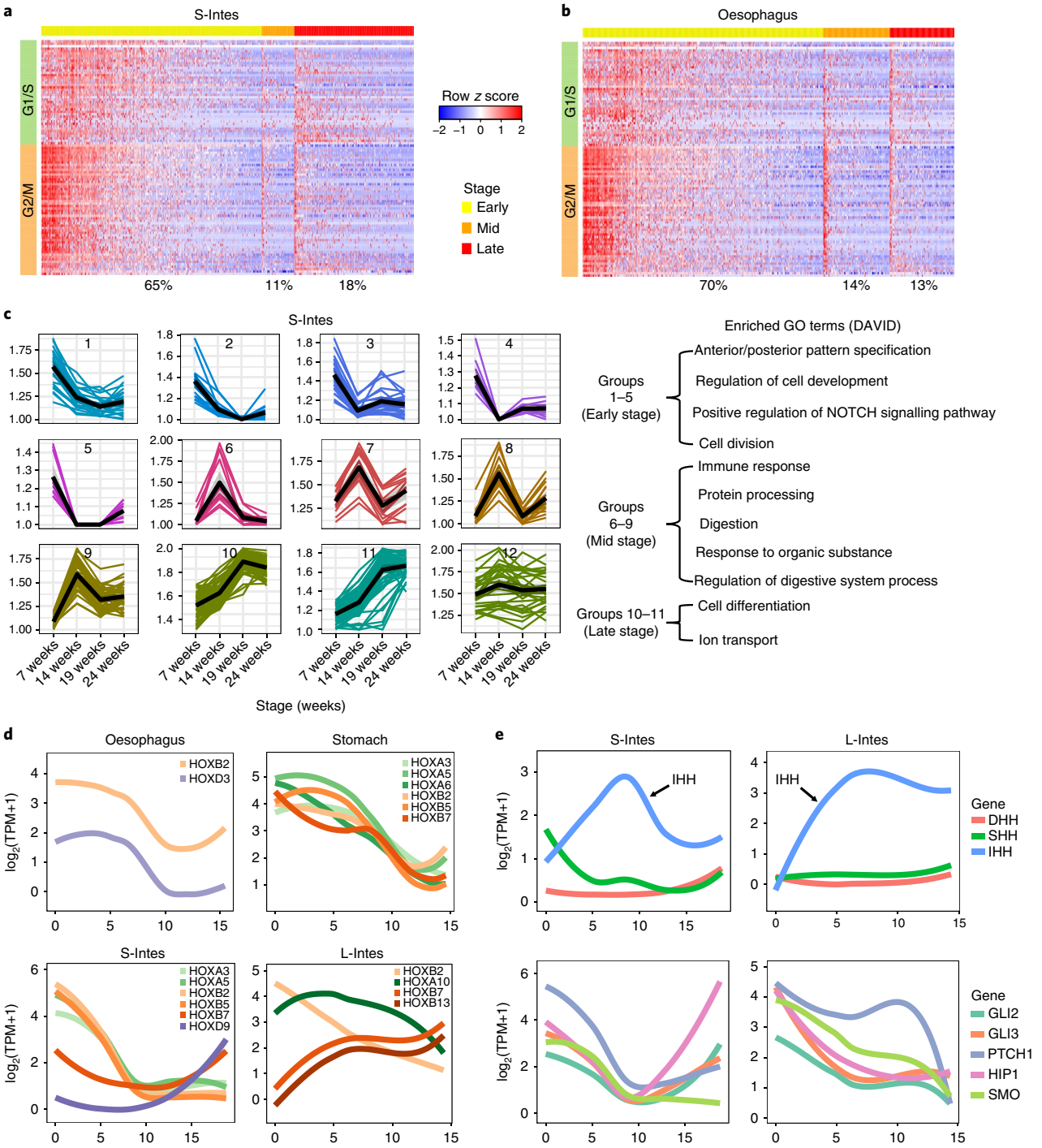


Fig. 5 | Transcriptome dynamics and selected signalling pathway in the development of human fetal intestinal tract. a, Heatmap showing expression patterns of cell-cycle-related genes in all S-Intes cells from the early, mid and late developmental stages. Cells were ordered by average expression levels of cell-cycle-related genes within each developmental stage. The colour key from blue to red indicates expression levels from low to high. The cell cycle index of each cell type is shown at the bottom of the heatmap. Early, $n=514$ cells; Mid, $n=76$ cells; Late, $n=278$ cells. **b**, Heatmap showing the expression patterns of cell-cycle-related genes in all oesophagus cells from the early, mid and late developmental stages. The colour key from blue to red indicates expression levels from low to high. The cell cycle index of each cell type is shown at the bottom of the heatmap. Early, $n=500$ cells; Mid, $n=138$ cells; Late, $n=135$ cells. **c**, Clustering of stage-specific gene expression tendencies during S-Intes development and corresponding enriched GO terms. Each sub-plot represents the expression tendencies of the group of stage-specific genes. Solid black and coloured lines represent expression tendencies of all genes and each gene, respectively. For example, Groups 1 to 5 present gene expression tendencies to be highly expressed in the early stage (7 weeks). Further GO analysis indicates that the early-stage-specific genes in Groups 1 to 5 are associated with cell proliferation and the NOTCH signalling pathway. By the mid stage, the enriched genes in Groups 6 to 9 are involved in digestion and the immune response. Late-stage specific genes (Groups 10 and 11) are mainly enriched for GO terms of cell differentiation. 7 weeks, $n=205$ cells; 14 weeks, $n=54$ cells; 19 weeks, $n=41$ cells; 24 weeks, $n=113$ cells. **d**, Smoothed-spline showing expression patterns of HOX family genes along the developmental pseudotime of each organ. Oesophagus, $n=773$ cells; stomach, $n=854$ cells; S-Intes, $n=868$ cells; L-Intes, $n=849$ cells. **e**, Smoothed-spline showing expression patterns of the Hedgehog signalling-pathway-related genes along the intestinal developmental pseudotime. S-Intes, $n=868$ cells; L-Intes, $n=849$ cells.

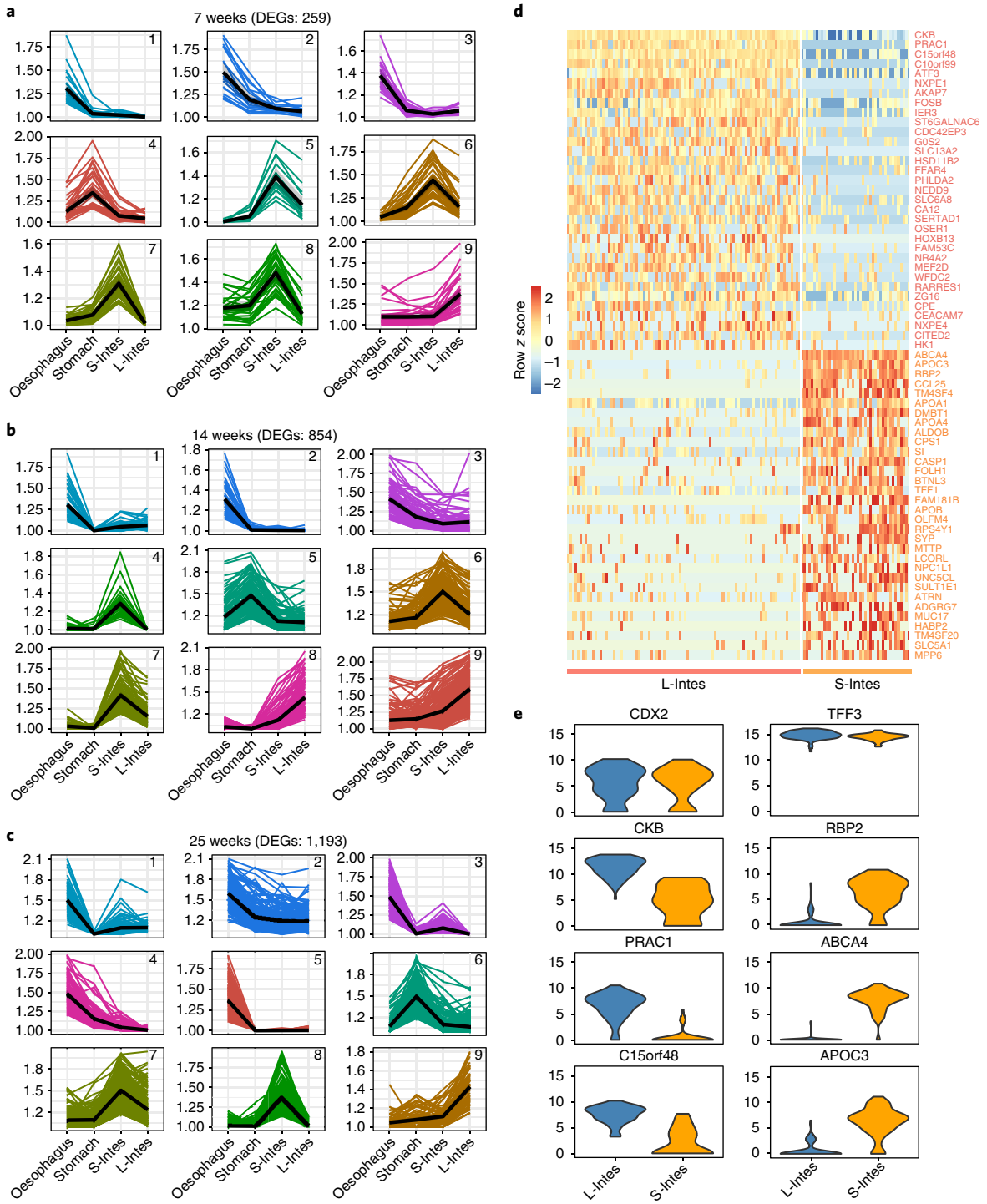


Fig. 6 | Identification of organ-specific genes. a–c, Clustering of genes with similar expression patterns by comparing their expression signatures among the four digestive organs of the 7, 14 and 25 weeks embryos, respectively. Solid black and coloured lines represent expression tendencies of all genes and each gene, respectively. Moreover, the patterns also show those groups of genes highly expressed in only one organ, such as Groups 1 to 3 (oesophagus), 4 (stomach), 5 to 8 (S-Intes) and 9 (L-Intes) in 7 weeks. Total numbers of DEGs are also shown. 7 weeks: oesophagus ($n=168$ cells), stomach ($n=233$ cells), S-Intes ($n=205$ cells), L-Intes ($n=118$ cells); 14 weeks: oesophagus ($n=65$ cells), stomach ($n=52$ cells), S-Intes ($n=54$ cells), L-Intes ($n=57$ cells); 25 weeks: oesophagus ($n=42$ cells), stomach ($n=25$ cells), S-Intes ($n=44$ cells), L-Intes ($n=32$ cells). **d**, Heatmap shows the DEGs between S-Intes goblet cells and L-Intes goblet cells. **e**, Violin plot showing expression patterns of S-Intes goblet-cell-specific genes and L-Intes goblet-cell-specific genes. The density of violin plots are scales to a maximum of 1 by setting 'scale = area' and all violins have the same maximum width. For example, a highly expressed gene of L-Intes, *PRAC1*, was reported to be specifically expressed in the prostate, rectum and distal colon, where it may play a regulatory role in cell nuclei. *RBP2* was specifically expressed in the S-Intes goblet cells, which was shown to be necessary for the growth, reproduction and differentiation of epithelium tissues. In **d** and **e**, S-Intes goblet cells, $n=77$; L-Intes goblet cells, $n=92$.

each organ. The pseudotimes were well matched with the actual developmental stages of the embryos analysed (Supplementary Fig. 2a and Supplementary Table 3). After rebuilding the temporal series, we further investigated the dynamic patterns of well-known marker genes in each organ along the temporal axis (Supplementary Fig. 2b-e). There results were verified at the protein level using western blotting (Supplementary Fig. 2f). Importantly, the four organs showed distinct maturation patterns (Supplementary Fig. 2g,h).

Cell type identification and developmental tracking of the fetal oesophagus. To better track the developmental route and accurately identify cell types in each organ, we defined the early (6–11 weeks), mid (12–18 weeks) and late (19–25 weeks) stages of gestation. The analysis pipeline is shown in Supplementary Fig. 2i. By applying the StemID and K-nearest neighbour (KNN) algorithm, we identified four cell types in the 25 weeks oesophagus (Fig. 2a and Supplementary Table 4). Based on the well-defined cell-type marker genes[39,40], cells in Cluster 1 were classified as ciliated epithelial cells, as these cells highly expressed *NME5* and *DNAI1* (Supplementary Fig. 3a), as further confirmed by gene ontology (GO) analysis of Cluster-1-specific genes (Fig. 2a). Cluster 2 cells were defined as KRT6B^{+} secretory progenitor cells because they highly expressed both basal cell markers (*KRT6A* and *KRT6B*) and secretory cell markers (*MUC1* and *MUC16*) (Supplementary Fig. 3a). Immunofluorescence staining clearly identified the MUC1^{+}KRT6B^{+} cells and MUC16^{+}KRT6B^{+} cells at the columnar epithelium layer (Fig. 2b). Clusters 3 and 4 were identified as FGFR1^{Low}NME5^{-} epithelial cells and FGFR1^{High}NME5^{-} epithelial cells, respectively. Moreover, Clusters 3 and 4 showed a higher percentage of actively cycling cells than Clusters 1 and 2 (Fig. 2c). Next, using KNN, we found that these cell types appeared and generally remained relatively constant throughout the entire 6–25 weeks developmental process (Fig. 2d and Supplementary Table 3).

Notably, we observed that *ANXA1* was specifically expressed in almost all oesophageal cells, except a few cells from the early stage (Supplementary Fig. 3b-d)[41,42]. Interestingly, the results of further analysis implied that *ANXA1* was expressed in the epithelial cells of the oesophagus, and the endothelial cells of the S-Intes, indicating that it may play distinct roles in the fetal oesophagus and S-Intes (Supplementary Fig. 3e-h).

Cell type identification and developmental tracking of the fetal stomach. By applying the same analysis strategy, we identified six cell types for the 24 weeks stomach (Fig. 2e and Supplementary Tables 3 and 4). According to well-defined marker genes[43–46], Clusters 1 and 2 were multipotent progenitor cells because they both expressed stem cell markers (*PROCR* in Cluster 1 and *TFF2* in Cluster 2) and terminally differentiated cell markers (*ATP4A* and *ATP4B* for parietal cells and *MUC5AC* and *TFF1* for pit cells). We defined Cluster 3 cells as HES1^{High} progenitor cells. The cells in Cluster 4 were identified as parietal progenitor cells. Cluster 5 highly expressed *LRIG1* (stem cells) and *TFF3* (pit cells), and were probably the progenitors of the pit cells, which were further verified through immunofluorescence staining (Fig. 2f). Cells of Cluster 6 were pit cells (*TFF1* and *TFF3*).

To trace the emergence and differentiation of these six cell types, we allocated all stomach cells to the pseudotime path constructed by Monocle2. We observed that all of these cell types were distributed from 6 to 25 weeks (Fig. 2g). The proliferation index of each cluster is shown in Fig. 2h.

Cell type identification and developmental tracking of the fetal small intestine. To investigate the expression patterns of the canonical cell type marker genes during human fetal S-Intes development, we performed immunofluorescence staining of the marker genes for the main cell types at 7 and 24 weeks (Supplementary Fig. 4). We observed that most of these genes were expressed at the putative mesenchymal layer but not the columnar epithelium at 7 weeks, although these cells were sparse and rare. At 24 weeks, combined with the formation of crypt-villus axis, the cell types expressing the canonical marker genes were located at the epithelial layer, as expected.

The transcriptomes of the 24 weeks S-Intes were analysed using the StemID algorithm (Fig. 3a,b and Supplementary Table 4). Based on the expression pattern of marker genes, nine types of S-Intes cell were clearly identified (Fig. 3c). We observed that cells in Cluster 1 were tuft progenitor cells because they highly expressed both stem cell markers *ALCAM* and *LRIG1* and tuft cell markers *TRPM5*, *PTGS1* and *RGS13*. Cluster 2 comprised PROCR^{+} progenitor cells. Additionally, cells in Clusters 3 and 4 were defined as two enterocyte progenitor subtypes with high expression of enterocyte markers and stem cell markers, as further confirmed by immunofluorescence staining (Fig. 3d). We classified Cluster 6 as goblet cells. The cells in Cluster 7 were secretory progenitor cells because they highly expressed both stem cell markers and markers of several secretory cell types (*GIP*, *LYZ* and *DCLK1*). We further determined that cells in Cluster 8 were LGR5^{+} stem cells and Cluster 9 cells tended to be enteroendocrine cells (*PYY*, *PROX1* and *TPH1*).

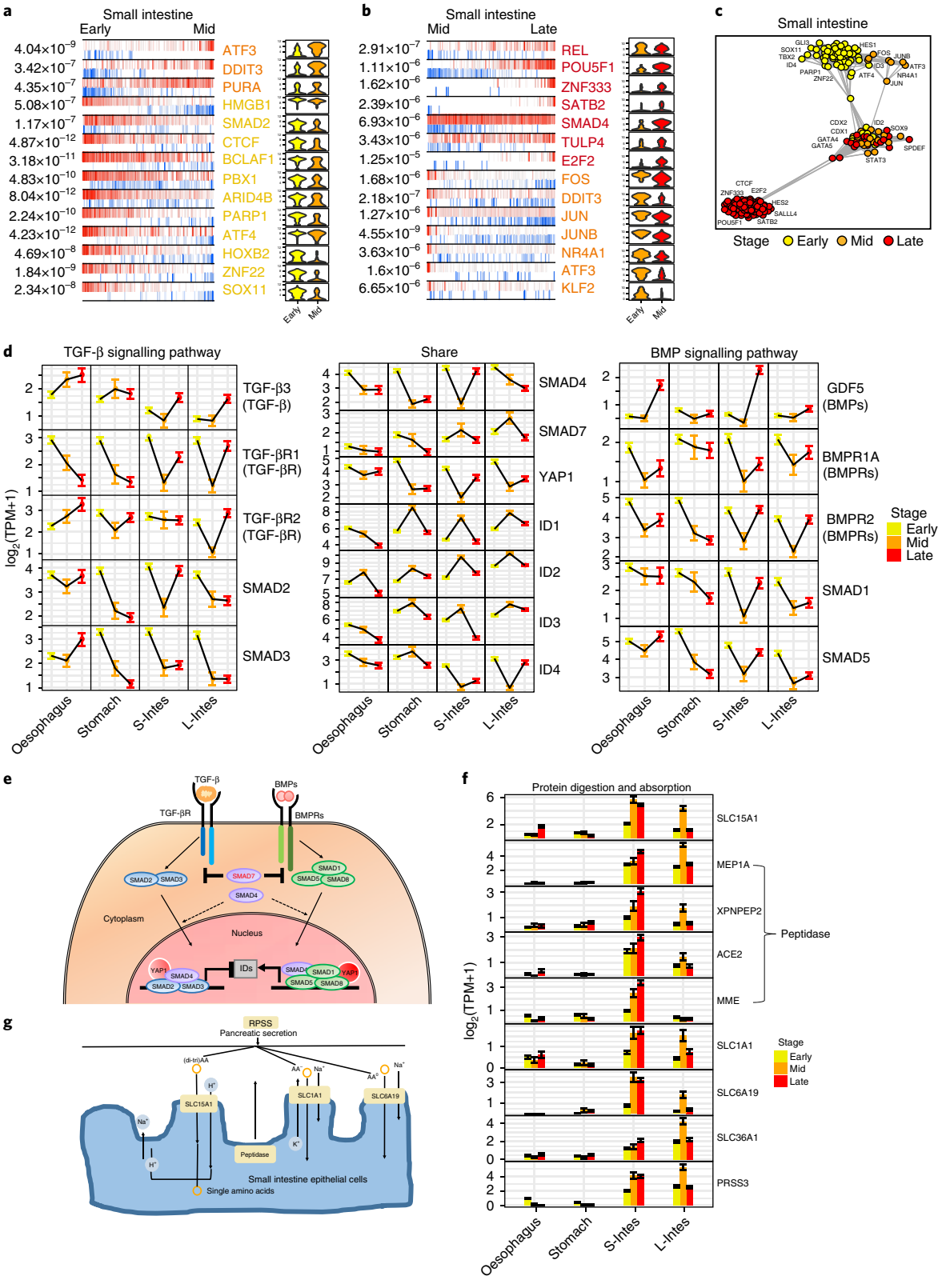
These nine cell types display clear developmental stage-specific patterns (Fig. 3e). Notably, most of the progenitor cells in Clusters 1 to 3 were from the early developmental stage. Most Goblet cells (Cluster 6) and enteroendocrine cells (Cluster 9) appeared in the late developmental stages. Interestingly, the reported LGR5^{+} stem cells (Cluster 8) in the adult S-Intes were established in the mid and late stage of gestation. Moreover, the results of the expression patterns of cell-cycle-related genes suggest that cell proliferation in the early developmental stage is, in general, more active than in the late stage (Fig. 3f).

Cell type identification and developmental tracking of the fetal large intestine. We performed immunofluorescence staining of

Fig. 7 | Inferred master regulators and signal pathway regulation of the small intestine. a, Top candidate master transcription regulators during S-Intes development from early stage ($n=514$ cells) to mid stage ($n=76$ cells). **b**, Top candidate master transcription regulators during the S-Intes development from mid stage ($n=76$ cells) to late stage ($n=278$ cells). In **a** and **b**, violin plots show expression levels of each master transcription regulator at early, mid and late stages, respectively. The density of violin plots are scaled to a maximum of 1 by setting 'scale = area' and all violins have the same maximum width. The precise *P* values are listed in the left-hand side. **c**, TF correlation network during S-Intes lineage progression ($n=868$ cells). Nodes (TFs) with more than three edges are shown, with each edge representing a high correlation (>0.3) between connected TFs. Correlation values are derived from Pearson pairwise correlation. Yellow, orange and red circles represent highest average expression levels of genes in early, mid and late stages, respectively. **d**, Graph showing expression patterns of genes involved in the transforming growth factor beta (TGF-β) and bone morphogenetic protein (BMP) signalling pathways in each organ along the developmental process. Oesophagus, $n=773$ cells; stomach, $n=854$ cells; S-Intes, $n=868$ cells; L-Intes, $n=849$ cells. Colours represent developmental stages (yellow, early stage; orange, mid stage; red, late stage). Data are presented as mean ± s.e.m. **e**, Diagram showing regulation of the TGF-β and BMP signalling pathways on inhibitor of DNA binding (ID) genes. **f**, Bar plots showing average expression levels of genes that play roles in protein digestion and absorption along the development process of the four organs. Oesophagus, $n=773$ cells; stomach, $n=854$ cells; S-Intes, $n=868$ cells; L-Intes, $n=849$ cells. Colours represent developmental stages (yellow, early stage; orange, mid stage; red, late stage). Data are presented as mean ± s.e.m. **g**, Diagram showing genes involved in S-Intes protein digestion and absorption.

the marker genes for the main intestine cell types at 7 weeks and 24 weeks (Supplementary Fig. 5a). The expression patterns of these genes in L-Intes were generally similar to those in S-Intes.

Similarly, as shown in Fig. 4a, 11 cell types were identified in the L-Intes from a 24 weeks embryo (Supplementary Tables 3 and 4). Cluster 1 was identified as DCLK1^{+} progenitor cells and Cluster 2 was defined as MKI67^{+} progenitor cells. We classified Clusters 3 and 4 as PROM1^{High} and PROM1^{Low} progenitor cells, respectively. The cells in Cluster 5 were LGR5^{+} stem cells. According to the expression of the secretory progenitor marker *DLL1* and the goblet progenitor marker *SPDEF*, we defined the cells in Cluster 6 as secretory progenitor cells, Cluster 7 as goblet progenitor cells and



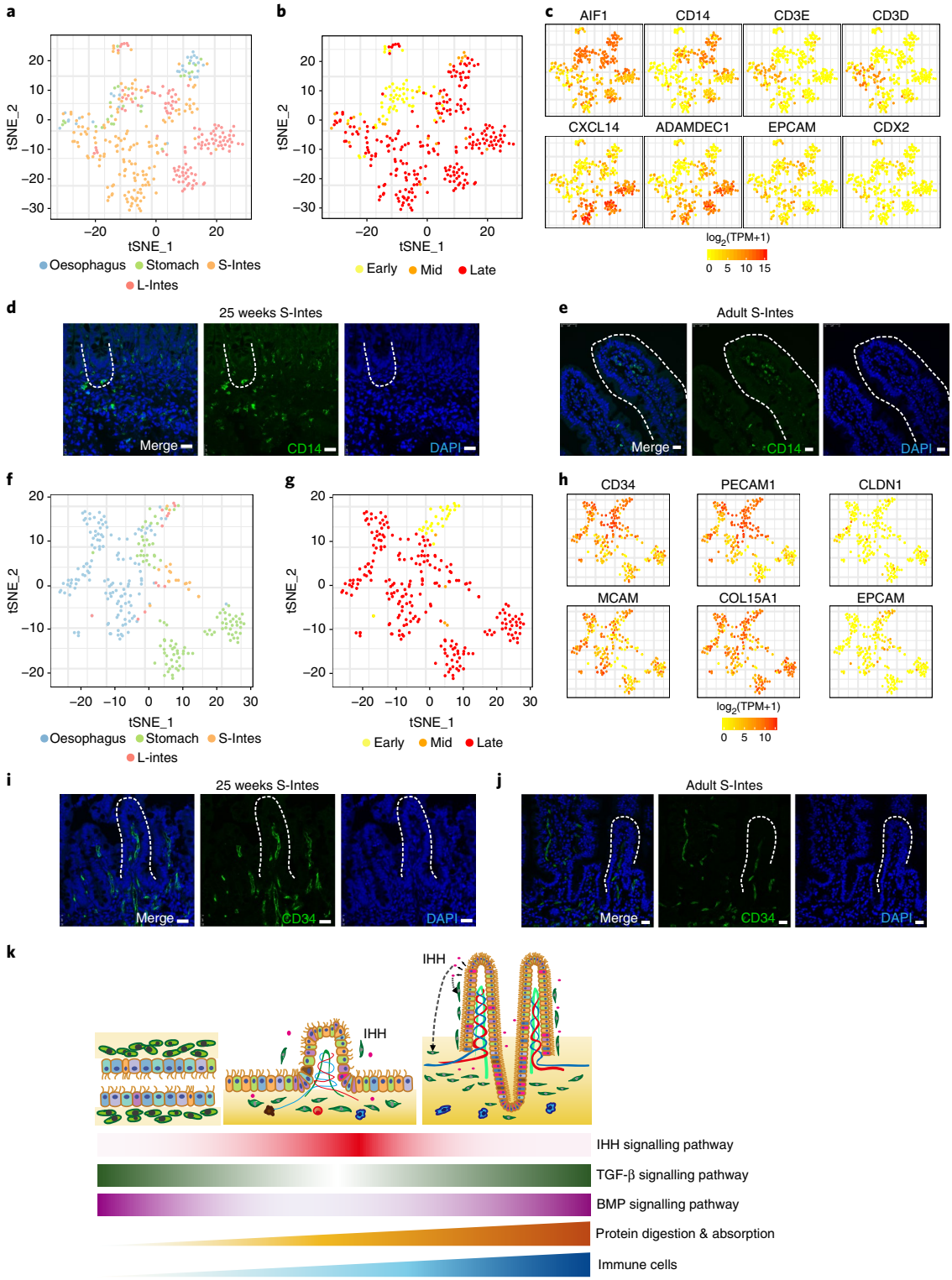


Fig. 8 | Gene expression analysis of immune and endothelial cell and summary of small intestine development. a, t-SNE map of immune cells. $n = 331$ cells. Cells are coloured according to organ source. **b**, Unsupervised clustering of immune cells. $n = 331$ cells. Cells are coloured according to stage source. **c**, Expression patterns of immune and epithelial cell markers are shown in a t-SNE map. $n = 331$ cells. Colour represents expression levels from low (yellow) to high (red). **d**,**e**, Immunostaining of CD14 (macrophage cell marker) in 25 weeks S-Intes (**d**) and adult S-Intes (**e**). Dashed lines indicate crypt (**d**) and villus (**e**). Scale bars, 25 μm. All experiments were independently repeated twice, with similar results. **f**, t-SNE map of endothelial cells. $n = 268$ cells. Cells are coloured according to organ source. **g**, t-SNE map of endothelial cells. $n = 268$ cells. Cells are coloured according to stage source. **h**, Expression patterns of endothelial cell marker genes *CD34*, *PECAM1* and *MCAM* and epithelial cell marker genes *EPCAM* and *CLDN1* are shown in a t-SNE map. $n = 268$ cells. Colour represents expression levels from low (yellow) to high (red). **i**,**j**, Immunostaining of CD34 (endothelial cell marker) in the 25 weeks S-Intes (**i**) and adult S-Intes (**j**). Dashed lines indicate the villus. Scale bars, 25 μm. All experiments were independently repeated twice, with similar results. **k**, Schematic diagram showing morphology changes, signalling pathway regulation and cell–cell interactions during the S-Intes developmental process. DAPI, 4′,6-diamidino-2-phenylindole.

Table 1 | Summary of the critical features of each cell type feature identified in the four fetal digestive organs

	Clusters	Cell Type	Stage	No. Cells	Cycling	Markers
Oesophagus	1	Ciliated Epithelial		26	23%	MME5, DNAI1, CES1, ENPP5
	2	KRT6B^{+} Secretory Progenitor		18	17%	KRT6B, HOPX, MUC21, SPNS2
	3	FGFR1^{Low}NME5^{-} Epithelial		415	51%	MBNL3, TCF4
	4	FGFR1^{High}NME5^{-} Epithelial		314	53%	LUM, FBN1, FGFR1
Stomach	1	PROCR^{High} Progenitor		139	71%	PPIH, SGK1, GPT2
	2	Progenitor at Neck		316	38%	TFF2, ECE1, PROM1
	3	HES1^{High}Progenitor		183	57%	LINC01578, PLXDC1
	4	Parietal Progenitor		156	48%	ATP4A, ATP4B, CFH, FGD5
	5	Pit Progenitor		18	22%	LRIG1, AXIN2, CD44, ACTC1
	6	Pit Cell		42	43%	ID1, MBD1, GREM, RSPO2
S-Intes	1	Tuft Progenitor		110	55%	DCLK1, TRPM5, RGS13, H1F0
	2	PROCR^{+} Progenitor		293	52%	PROCR, MFAP4
	3	Enterocyte Progenitor -1		97	46%	CA2, OLFM4, TSPAN8, FUK
	4	Enterocyte Progenitor -2		171	46%	APOB, APOB3, AFP
	5	Enterocyte		12	17%	ALPI, THER, MME, CDHR5
	6	Goblet Cell		43	12%	TFF3, MUC2, SPINK4, FCGBP
	7	Secretory Progenitor		42	43%	GIP, LYZ, GAD2
	8	LGR5^{+} Stem Cells		93	35%	LGR5, EPYC, CLDN18
	9	Enteroendocrine		7	29%	TPH1, PYY, AXIN2, CPZ
L-Intes	1	DCLK1^{+} Progenitor		46	20%	DCLK1, SPARC, MGP
	2	MKI67^{+} Progenitor		53	53%	HEG1, MKI67, CDK1
	3	PROM1^{High} Progenitor		58	52%	PTGS1, PROM1, CCBE1, ARMC9
	4	PROM1^{Low} Progenitor		348	54%	LRIG1, ALCAM, SGCA
	5	LGR5^{+} Stem Cells		126	47%	LGR5, SOX9, OLFM4, CA2, LEFTY1
	6	Secretory Progenitor		9	22%	DLL1, SPINK4, STAB1
	7	Goblet Progenitor		25	20%	TFF3, ZG16, SPDEF, CHGA, FOXA3
	8	Goblet Cell		67	1.5%	CDH1, REG4, ZG16
	9	Enteroendocrine		42	50%	PYY, ANKED20A9P, FRAS1
	10	Enterocyte Cell		65	18%	AQP8, SI, LAD1
	11	Paneth-Like Cell		10	10%	LYZ, PLA2G7, SRGN

Stage: Early Mid- Late

Cluster 8 as goblet cells. Cluster 7 cells were further confirmed by immunofluorescence staining of *CHGA* and *TFF3* in 24 weeks fetal L-Intes, adult L-Intes and 24 weeks S-Intes (Supplementary Fig. 5b). We determined that the cells in Clusters 9 and 10 were enteroendocrine and enterocyte cells, respectively[47]. Cluster 11 cells were identified as Paneth-Like cells[48].

Subsequently, the developmental progress of these eleven cell types was investigated (Fig. 4b). Of note, progenitor cells in Clusters 1 to 4 initially emerged in the early stage, and most of them distributed in the late stage except Cluster 3. The differentiated cell types (Clusters 8 to 11) were primarily distributed in the mid to late stages. Moreover, most of these cell types belonging to the early

stage or stem/progenitor cells, such as Clusters 2 to 5, tend to exhibit a higher proliferation index than differentiated cell types (Clusters 8, 10 and 11) (Supplementary Fig. 5c).

Cell type identification of adult large intestine and comparison between fetal and adult large intestine. Similarly, the results of immunofluorescence staining confirmed that all of the marker genes for major types of intestine epithelial cell expressed in the adult L-Intes (Supplementary Fig. 5a).

Subsequently, to systematically explore the transcriptome dynamics from fetal to adult stage, 1,138 individual cells from adult L-Intes were analysed using single-cell RNA-seq after filtration. According to the t-SNE analysis and expression patterns of known marker genes, we identified 10 cell types, including enteroendocrine cells (Cluster A), enterocytes (Clusters B to D), OLFM4^{High} stem cells (Cluster E), goblet cells (Cluster F and G), MKI67^{High} cells (Cluster H), mesenchymal cells (Cluster I) and immune cells (Cluster J) (Fig. 4c,d and Supplementary Fig. 6a,b).

Next, using principal component analysis (PCA), we observed that the analysis not only distinguished the fetal L-Intes from the adult L-Intes, but also indicated that the adult L-Intes cells were more similar to the late fetal stage L-Intes cells, as expected (Fig. 4e and Supplementary Fig. 6c). We also investigated the differentially expressed genes (DEGs) between the fetal and adult L-Intes and performed GO analysis (Supplementary Table 5).

Dynamic gene expression patterns of the digestive tract along the differentiation and spatial axis during human fetal development. First, the expression profiles of cell-cycle-related genes revealed that the cell cycle activity of the S-Intes initially decreased and then slightly increased during human fetal development (Fig. 5a). The oesophagus showed gradual decreases in cell proliferation activity from early to late stages (Fig. 5b).

Next, the four organs from 7, 14, 19 and 24 weeks were selected to investigate the dynamic gene expression patterns of the human fetal digestive tract during development. Additionally, DEGs for each developmental stage were clustered in each organ based on the gene expression patterns to identify more representative stage-specific marker genes (Fig. 5c, Supplementary Fig. 7a-c and Supplementary Table 6). In the S-Intes, GO analysis indicated that the early stage cells (Groups 1 to 5) are probably actively proliferating progenitor cells, the digestive functions and immune components were established at the mid stage (Groups 6 to 9), whereas more differentiated and more mature cell types were detected in the late stage (Groups 10 and 11) (Fig. 5c).

Subsequently, we observed that HOX genes showed clear spatial-temporal specific expression patterns during digestion tract development[49] (Fig. 5d). To investigate the regulation patterns of the Hedgehog signalling pathway[50,51], we found that only ligand IHH was highly expressed by the S-Intes and L-Intes epithelial cells (EPCAM^{High} cells), which was different from the expression pattern in the mouse[50,52,53] (Fig. 5e). According to the expression pattern of Hedgehog signalling genes (*PTCH1*, *SMO* and *GLI2/3*), we speculated that the reported paracrine pattern of the Hedgehog signalling pathway may also exist in the human fetal S-Intes[54,55] (Fig. 5e and Supplementary Fig. 7d-f). Further analysis showed that Hedgehog signalling tends to indirectly activate the transition between epithelial and mesenchymal states in human fetal S-intes by crosstalk with other signal pathways[56] (Supplementary Fig. 7d,e).

To further explore the dynamic processes among these organs as development proceeded, we investigated the DEGs among the four digestive organs from three embryos (7, 14 and 25 weeks) (Fig. 6a–c and Supplementary Table 7). We observed that more DEGs among different organs were detected at later stages (1,193) than at earlier stages (259), indicating that these organs became more different from each other at the later stage. Next, we selected those groups of genes highly expressed in only one organ, such as Groups 1 to 3 (oesophagus), 4 (stomach), 5 to 8 (S-Intes) and 9 (L-Intes) at 7 weeks (Fig. 6a). These DEGs specifically revealed the organ-specific signatures of these four organs at each stage analysed (Supplementary Table 7). To investigate the differences of the same cell types in different organs, we compared the transcriptome profiles between S-Intes goblet cells and L-Intes goblet cells (Fig. 6d,e). Goblet cells in both the S-Intes and L-Intes showed similar expression patterns regarding *CDX2* and *TFF3*. However, 65 DEGs existed between S-Intes and L-Intes goblet cells. Goblet cells in the S-Intes specifically express genes, such as *ABCA4* and *APOC3*, that play important roles in digestion and absorption.

Transcription factor regulatory network during human fetal digestive tract development. All 1,568 known human transcription factors (TFs) were analysed at the three developmental stages[57]. We identified a clear set of developmental-stage-specific TFs that may play critical roles for the in vivo development of these organs. To further explore the regulation pattern at all three developmental stages, we constructed a TF correlation network to visualize the gene regulation change occurring during the development process of the S-Intes through the early–mid–late stage. The details of potentially important TFs in these organs are shown in Fig. 7a–c and Supplementary Fig. 8.

Accordingly, we systematically investigated the expression patterns of genes related to the TGF-β and BMP signalling pathways during human fetal digestive tract development (Fig. 7d,e). As expected, the regulation of each pathway showed consistent dynamic changes in the S-Intes and L-Intes during development. In detail, we observed that *TGF-β3*, the receptor *TGF-βR1*, and the complex of *SMAD2/3* with the common mediators *SMAD4* and *YAP1* showed decreased expression from the early to mid stages and then increased expression in the late stage during S-Intes development. In addition, inhibitory *SMAD7* showed the opposite expression pattern[58]. Meanwhile, the BMP signalling pathway genes were all downregulated in the mid stage and subsequently upregulated in the late stage. Furthermore, we observed that the expression levels of *ID1/2/3* were probably inhibited through the TGF-β signalling pathway whereas *ID4* was probably activated via the BMP signalling pathway during human fetal development[59,60].

Notably, most of the genes involved in protein digestion and absorption, particularly the peptidase genes, were specifically upregulated in the S-Intes from the early to late stages (Fig. 7f,g). Interestingly, in the L-Intes, expression of these peptidase genes increased from the early to mid stages but decreased in the late stage, implying that functional fate diversification between the S-Intes and L-Intes may occur in the late stage during human embryo development.

Analysis of immune and endothelial cells in human fetal digestive tract development. As shown in Fig. 1, we isolated immune cells from the t-SNE analysis. According to the expression patterns of known marker genes, macrophages (*AIF1* and *CD14*), T cells (*CD3E* and *CD3D*), and monocytes/dendritic cells (*CXCL14* and *ADAMDEC1*) were identified (Fig. 8a–c). Importantly, these immune cells tended to emerge in the late stage, and most of these cells were located in the S-Intes and L-Intes (Fig. 8a,b). Moreover, immunofluorescence staining showed that CD14^{+} cells (macrophages) located in the basal layer of the 25 weeks S-Intes and distributed throughout the villus lacuna of the adult S-Intes, indicating the sharp distinction of the immune cells in the fetal and adult S-Intes (Fig. 8d,e).

The endothelial cells were also analysed by t-SNE. Generally, we observed that most of these endothelial cells emerged in the late stage in the oesophagus and stomach (Fig. 8f,g). Moreover, these cells showed high expression of endothelial markers but low expression of the epithelial markers (Fig. 8h). Immunofluorescence

staining clearly showed that endothelium cells of 25 weeks S-Intes located in and distributed throughout the villus lacuna, similar to that in the adult S-Intes (Fig. 8i,j).

Discussion

In summary, this study systematically analyses the gene expression profiles of human fetal digestive tract and adult L-Intes in vivo at single-cell resolution. We first identified 30 different cell types in the fetal oesophagus (4), stomach (6), S-Intes (9) and L-Intes (11). Detailed information, including cell identity, developmental stage, cell number, cell cycle index and cell-type-specific markers are summarized in Table 1. Moreover, we identified 10 types of cell in adult L-Intes and uncovered the transcriptomic features and developmental dynamics from fetal to adult stages. Second, we revealed the regulation pattern of signalling pathways such as IHH, TGF-β and BMP during human fetal development (Fig. 8k). In addition, the physiological functions involved in the digestion and absorption of proteins were initially established and then specifically enhanced in the S-Intes (Fig. 8k). Third, we identified immune cells, including T cells, B cells and macrophages, in the embryonic digestion tract, and most of them arise in the late stages of S-Intes and L-Intes development.

Our findings offer a comprehensive transcriptome resource for human fetal digestive tract development as well as the large intestine in adults at single-cell resolution. This analysis of the gene expression patterns of each organ in both temporal and spatial dimensions provides guidelines to estimate the status of cells differentiated from pluripotent stem cells in vitro. It may facilitate the diagnosis and treatment of digestive-tract-related diseases, such as GI cancer.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41556-018-0105-4.

Received: 12 October 2017; Accepted: 17 April 2018;
Published online: 25 May 2018

References

1. Zorn, A. M. & Wells, J. M. Vertebrate endoderm development and organ formation. *Annu. Rev. Cell Dev. Biol.* **25**, 221–251 (2009).
2. Kraus, M. R. & Grapin-Botton, A. Patterning and shaping the endoderm in vivo and in culture. *Curr. Opin. Genet. Dev.* **22**, 347–353 (2012).
3. Sherwood, R. I., Chen, T. Y. & Melton, D. A. Transcriptional dynamics of endodermal organ formation. *Dev. Dyn.* **238**, 29–42 (2009).
4. Schedl, H. P. & Clifton, J. A. Solute and water absorption by the human small intestine. *Nature* **199**, 1264–1267 (1963).
5. Schulz, M. D. et al. High-fat-diet-mediated dysbiosis promotes intestinal carcinogenesis independently of obesity. *Nature* **514**, 508–512 (2014).
6. Ugolev, A. Influence of the surface of the small intestine on enzymatic hydrolysis of starch by enzymes. *Nature* **188**, 588–589 (1960).
7. Maynard, C. L., Elson, C. O., Hatton, R. D. & Weaver, C. T. Reciprocal interactions of the intestinal microbiota and immune system. *Nature* **489**, 231–241 (2012).
8. Palm, N. W. et al. Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell* **158**, 1000–1010 (2014).
9. Jervis, E. L. & Levin, R. J. Anatomic adaptation of the alimentary tract of the rat to the hyperphagia of chronic alloxan-diabetes. *Nature* **210**, 391–393 (1966).
10. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
11. Park, Y. H. & Kim, N. Review of atrophic gastritis and intestinal metaplasia as a premalignant lesion of gastric cancer. *J. Cancer Prev.* **20**, 25–40 (2015).
12. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-seq by multiplexed linear amplification. *Cell Rep.* **2**, 666–673 (2012).
13. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
14. Klein, A. M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
15. Tang, F. et al. RNA-seq analysis to capture the transcriptome landscape of a single cell. *Nat. Protoc.* **5**, 516–535 (2010).
16. Fan, X. et al. Single-cell RNA-seq transcriptome analysis of linear and circular RNAs in mouse preimplantation embryos. *Genome Biol.* **16**, 148 (2015).
17. Jaitin, D. A. et al. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779 (2014).
18. Islam, S. et al. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res* **21**, 1160–1167 (2011).
19. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
20. Svensson, V. et al. Power analysis of single-cell RNA-sequencing experiments. *Nat. Methods* **14**, 381–387 (2017).
21. Ziegenhain, C. et al. Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* **65**, 631–643 (2017).
22. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
23. Grun, D. et al. De novo prediction of stem cell identity using single-cell transcriptome data. *Cell Stem Cell* **19**, 266–277 (2016).
24. Grun, D. et al. Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525**, 251–255 (2015).
25. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
26. Qiu, X. et al. Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **14**, 309–315 (2017).
27. Joost, S. et al. Single-cell transcriptomics reveals that differentiation and spatial signatures shape epidermal and hair follicle heterogeneity. *Cell Syst.* **3**, 221–237 (2016).
28. Schumacher, M. A. et al. The use of murine-derived fundic organoids in studies of gastric physiology. *J. Physiol.* **593**, 1809–1827 (2015).
29. Haber, A. L. et al. A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333–339 (2017).
30. Petropoulos, S. et al. Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell* **165**, 1012–1026 (2016).
31. Mohammed, H. et al. Single-cell landscape of transcriptional heterogeneity and cell fate decisions during mouse early gastrulation. *Cell Rep.* **20**, 1215–1228 (2017).
32. Noah, T. K., Donahue, B. & Shroyer, N. F. Intestinal development and differentiation. *Exp. Cell Res.* **317**, 2702–2710 (2011).
33. Kearns, N. A. et al. Generation of organized anterior foregut epithelia from pluripotent stem cells using small molecules. *Stem Cell Res.* **11**, 1003–1012 (2013).
34. Peters, H. et al. Isolation of the Pax9 cDNA from adult human esophagus. *Mamm. Genome* **8**, 62–64 (1997).
35. Quante, M., Marrache, F., Goldenring, J. R. & Wang, T. C. TFF2 mRNA transcript expression marks a gland progenitor cell of the gastric oxyntic mucosa. *Gastroenterology* **139**, 2018–2027 (2010).
36. Noguchi, T. K. et al. Generation of stomach tissue from mouse embryonic stem cells. *Nat. Cell Biol.* **17**, 984–993 (2015).
37. Goodell, M. A., Nguyen, H. & Shroyer, N. Somatic stem cell heterogeneity: diversity in the blood, skin and intestinal stem cell compartments. *Nat. Rev. Mol. Cell Biol.* **16**, 299–309 (2015).
38. Barker, N. Adult intestinal stem cells: critical drivers of epithelial homeostasis and regeneration. *Nat. Rev. Mol. Cell Biol.* **15**, 19–33 (2014).
39. Rishniw, M. et al. Molecular aspects of esophageal development. *Ann. NY Acad. Sci.* **1232**, 309–315 (2011).
40. Yu, W. Y., Slack, J. M. & Tosh, D. Conversion of columnar to stratified squamous epithelium in the developing mouse oesophagus. *Dev. Biol.* **284**, 157–170 (2005).
41. Gao, Y., Chen, Y., Xu, D., Wang, J. & Yu, G. Differential expression of ANXA1 in benign human gastrointestinal tissues and cancers. *BMC Cancer* **14**, 520 (2014).
42. Onozawa, H. et al. Annexin A1 is involved in resistance to 5-FU in colon cancer cells. *Oncol. Rep.* **37**, 235–240 (2017).
43. Kim, T. H. & Shivdasani, R. A. Stomach development, stem cells and disease. *Development* **143**, 554–565 (2016).
44. McCracken, K. W. et al. Modelling human development and disease in pluripotent stem-cell-derived gastric organoids. *Nature* **516**, 400–404 (2014).
45. Bartfeld, S. et al. In vitro expansion of human gastric epithelial stem cells and their responses to bacterial infection. *Gastroenterology* **148**, 126–136 (2015).
46. Wang, D. et al. Identification of multipotent mammary stem cells by protein C receptor expression. *Nature* **517**, 81–84 (2015).
47. de Lau, W. et al. Peyer's patch M cells derived from $Lgr5^+$ stem cells require SpiB and are induced by RankL in cultured 'miniguts'. *Mol. Cell Biol.* **32**, 3639–3647 (2012).
48. Almohazey, D. The ErbB3 receptor tyrosine kinase negatively regulates Paneth cells by PI3K-dependent suppression of Atoh1. *Cell Death Differ.* **24**, 855–865 (2017).
49. Beck, F. Homeobox genes in gut development. *Gut* **51**, 450–454 (2002).
50. Merchant, J. L. Hedgehog signalling in gut development, physiology and cancer. *J. Physiol.* **590**, 421–432 (2012).

51. Kim, B. M. & Choi, M. Y. New insights into the role of Hedgehog signaling in gastrointestinal development and cancer. *Gastroenterology* **137**, 422–424 (2009).
52. Litingtung, Y., Lei, L., Westphal, H. & Chiang, C. Sonic hedgehog is essential to foregut development. *Nat. Genet.* **20**, 58–61 (1998).
53. Motoyama, J. et al. Essential function of Gli2 and Gli3 in the formation of lung, trachea and oesophagus. *Nat. Genet.* **20**, 54–57 (1998).
54. van den Brink, G. R. et al. Indian Hedgehog is an antagonist of Wnt signaling in colonic epithelial cell differentiation. *Nat. Genet.* **36**, 277–282 (2004).
55. van den Brink, G. R. Hedgehog signaling in development and homeostasis of the gastrointestinal tract. *Physiol. Rev.* **87**, 1343–1375 (2007).
56. Katoh, Y. & Katoh, M. Hedgehog signaling, epithelial-to-mesenchymal transition and miRNA (review). *Int J. Mol. Med* **22**, 271–275 (2008).
57. Li, L. et al. Single-cell RNA-seq analysis maps development of human germline cells and gonadal niche interactions. *Cell Stem Cell* **20**, 891–892 (2017).
58. Siegel, P. M., Shu, W. & Massague, J. Mad upregulation and Id2 repression accompany transforming growth factor (TGF)-beta-mediated epithelial cell growth suppression. *J. Biol. Chem.* **278**, 35444–35450 (2003).
59. Wang, R. N. et al. Bone morphogenetic protein (BMP) signaling in development and human diseases. *Genes Dis.* **1**, 87–105 (2014).
60. Qi, Z. BMP restricts stemness of intestinal Lgr5(+) stem cells by directly suppressing their signature genes. *Nat. Commun.* **8**, 13824 (2017).

Acknowledgements

This work was supported by grants from the National Natural Science Foundation of China (31625018, 31230047, 31601177, 81521002), the Ministry of Science and Technology of China (2017YFA0102702), a General Financial Grant from the China Postdoctoral Science Foundation (2017M610703) and Shanghai Science and Technology Development Funds (16YF140940).

Author contributions

F.T. and J.Q. conceived and supervised the project. S.G., L.Y., and J.L. performed the experiments, including single cell collection, library construction, immunofluorescence staining and so on, with the help of Ju.Y., Xi.Z., X.W., Y.W., X.Y.W, X.F., Ji.Y., Xu.Z., Y.G., S.H.G., W.W., C.W., Y.M., W.F. and L.W. Under the supervision of H.G. and with the help of X.J., R.W. performed bioinformatics analysis. S.G., F.T., R.W., J.L. and L.Y. wrote the manuscript with help from all of the authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41556-018-0105-4.

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

Correspondence and requests for materials should be addressed to H.G. or J.Q. or F.T.

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

Methods

Ethics statement. The embryonic part of the project was designed to study the gene expression network for the development of human fetal gut through RNA-seq transcriptome analyses. The study first analysed the four main organs of the digestive tract, including the oesophagus, stomach, small intestine and large intestine, from multiple human embryos between 6 and 25 weeks of gestation (weeks after estimated fertilization time) using single-cell RNA-seq analyses. The adult part of the project was designed to study the gene expression network of the adult human large intestine through RNA-seq transcriptome analyses. The adult cells were collected from two adults and the transcriptome dynamics of the large intestine cells from human fetal to adult stages were systematically investigated at single-cell resolution. All of the above mentioned protocols have been reviewed and approved by the Ethics Committee of Peking University Third Hospital (2012SZ013, IRB00006761-M2016170). Informed consent was obtained from all couples (who could not continue pregnancies because of severe pregnancy disease or severe fetal anomaly). Before giving consent, persons donating embryos were provided with all the necessary information, including an introduction to this research, a means to receive counselling, as well as the risk, gain and right of participation. No financial inducements were offered for donation. Human fetal stem/progenitor cell work (identification and analysis of fetal stem/progenitor cells in Figs. 2 to 4, Supplementary Figs. 3a, 5b,c and 6 and Supplementary Tables 3 to 5) complied with the 2016 Guidelines for Stem Cell Research and Clinical Translation issued by the International Society for Stem Cell Research (ISSCR). All relevant ethical regulations of the Ethics Committee of Peking University Third Hospital (IRB00006761-M2016170) were followed for the adult human research participants.

Human embryo collection, tissue dissection and single-cell capturing. The current study collected human embryos aged from 6 weeks to 25 weeks and L-Intes was obtained from two adults. For single-cell RNA-sequencing of fetal tissues, two embryos per stage were collected for 6, 7 and 8 weeks samples, and one embryo was collected for each of the other stages (9–25 weeks) due to the scarcity of materials available (except that one pair of twins were collected at 21 weeks). For analysis of adult large intestine tissues, samples were collected from two donors. The stages of human embryos examined in the present study were calculated from the estimated fertilization time. In total, 15 embryos were collected and, for each human embryo, the main organs of the digestive system, including the oesophagus, stomach, small intestine (S-Intes) and large intestine (L-Intes), were dissected. However, for one of the embryos ('embryo-1') at 7 weeks only the stomach and L-Intes (Supplementary Table 1) were isolated because of fragmentation of the sample. These four organs at 6–9 weeks were still tiny, and the substructures within each tissue were indistinguishable, so all parts of each organ were dissociated for single-cell collection. For S-Intes and L-Intes after 10 weeks, we isolated the cells from ileum of S-Intes and from rectum of L-Intes. Moreover, the villus and crypt were separated by shaving off the villus carefully with a surgical blade. The mesenteric matrix of the S-Intes and L-Intes was torn off carefully under a microscope. The corpus of the stomach was used to collect single cells. The selected tissue pieces were washed with PBS several times. The selected tissue pieces were then digested into single cells using 2 mg ml^{-1} collagenase/dispase (Roche) for 30–40 min at 37 °C. The cell suspension was then centrifuged at 800*g* for 5 min at 4 °C, and the cell pellets were resuspended in 1% BSA. Finally, we used a mouth pipette to pick all types and sizes of living single cells, which were immediately transferred to cold scRNA-seq lysis buffer.

Single-cell lysis and reverse transcription. We used a modified STRT-seq[18] protocol to construct the libraries. Initially, single cells were placed in a 200 µl Eppendorf tube containing 2 µl of lysis buffer with 0.05 µl RNase Inhibitor (40 U µl^{-1}, Invitrogen), 0.095 µl Triton X-100 (10%, Sigma-Aldrich), 0.5 µl dNTP (10 mM, Fermentas), 1.105 µl nuclease-free water, 0.1 µl ERCC (1:2,000,000 dilution of ERCC RNA Spike-In Mix, Life Technologies) and 0.15 µl barcode primer (10 µM, 5′-TCAGACGTGTGCTCTTCCGATCTXXXXXXXXNNNNNNNNT25-3′; X8 is an 8 bp predesigned barcode sequence, N8 is an 8 nt random unique molecular identifier (UMI) sequence) with mouth pipetting, and lysed at 72 °C for 3 min. Subsequently, 2.85 µl of reverse transcription mix containing 0.25 µl SuperScript II reverse transcriptase (200 U µl^{-1}, Invitrogen), 1 µl Superscript II first-strand buffer (5×, Invitrogen), 0.125 µl RNase Inhibitor (40 U µl^{-1}, Invitrogen), 0.25 µl dithiothreitol (DTT) (0.1 M, Invitrogen), 1 µl betaine (5 M, Sigma-Aldrich), 0.03 µl $MgCl_2$ (1 M, Sigma-Aldrich), 0.145 µl nuclease-free water, 0.05 µl TSO Primer (100 µM, 5′-AAGCAGTGGTATCAACGCAGAGTACATrGrG+G-3′; 'rG', riboguanosines; '+G', locked nucleic acid (LNA)-modified guanosine) were added to each tube with the lysed single cell. The mixture was incubated at 25 °C for 5 min, followed by 42 °C for 60 min, 50 °C for 30 min and 70 °C for 10 min.

cDNA preamplification. Each tube included 7.5 µl PCR mix containing 6.25 µl KAPA HiFi HotStart ReadyMix (2×, KAPA), 0.25 µl primer (10 µM, 5′-AAGCAGTGGTATCAACGCAGAGT-3′), 0.75 µl 3′P2 primer (10 µM, 5′-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC-3′) and 0.25 µl nuclease-free water, followed by incubation at 95 °C for 3 min, 4 cycles at 98 °C for 20 s, 65 °C for 30 s, and 72 °C for 5 min), with 10–16 cycles at 98 °C for 20 s, 67 °C for 15 s and 72 °C for 5 min), with a final step at 72 °C for 5 min. The preamplified cDNAs tagged with different barcodes were pooled, and the pooled DNA was purified twice using 0.8× AMPure XP beads (Beckman).

Tagging the 3′ ends with biotin and fragmentation. Biotin PCR was performed using primer (5′-AAGCAGTGGTATCAACGCAGAGT-3′) and biotin index primers. Biotin PCR mixture containing 25 µl KAPA HiFi HotStart ReadyMix (2×, KAPA), 0.4 µM IS primer (5′-AAGCAGTGGTATCAACGCAGAGT-3′), 2 µl of biotin primer (10 µM, 5′-/biotin/CAAGCAGAAGACGGCATACGAGAT/index/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC-3′), 30–40 ng of pooled DNA and nuclease-free water to a total volume of 50 µl. After PCR, the 3′ ends of the DNA were tagged with biotin. Subsequently, the PCR products were purified twice. The biotin-tagged DNA was fragmented to ~300 bp and purified once.

Biotin enrichment. Dynabeads MyOne streptavidin C1 (Invitrogen) was used to enrich the biotin-tagged 3′ ends.

Library construction. The end-repair and dA-tailing of the DNA fragments and ligation of the adaptors to the DNA fragments were performed according to the KAPA Hyper Prep Kits with PCR Library Amplification/Illumina series (KAPA, KK8054). After the adaptor ligation step, the final PCR was performed with 25 µl of KAPA HiFi HotStart ReadyMix (2×, KAPA), 0.3 µM Illumina QP2 primer (5′-CAAGCAGAAGACGGCATACGA-3′) and 0.3 µM short universal primer (5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC-3′), adaptor-ligated DNA and nuclease-free water to a final volume of 50 µl. The libraries were sequenced with Illumina HiSeq 4000 for 150 bp paired-end reads.

Immunofluorescence staining. Small pieces of oesophagus, stomach, S-Intes and L-Intes dissected from embryos at different weeks were fixed in 4% paraformaldehyde overnight at 4 °C. After washing three times with PBS, tissues were dehydrated in 20% sucrose and embedded in Tissue-Tek OCT. The slices were then permeabilized with 0.5% PBS-Triton for 30 min and blocked with 10% donkey serum or fetal bovine serum (Visteck) in 0.5% PBS-Triton for 1.5 h at room temperature. The slices were incubated with primary antibodies diluted in blocking solution overnight at 4 °C and incubated in secondary antibodies diluted in blocking solution for 2 h at room temperature. DAPI (4′,6-diamidino-2-phenylindole, blue) was used to stain the cell nuclei. Finally, tissue slices were mounted for confocal laser scanning microscope analysis.

Western blotting. S-Intes and L-Intes from fetuses at 7, 14 and 24 weeks were first ground to fine powder with a mortar and pestle in liquid nitrogen. They were then collected and lysed in lysis buffer (50 mM Tris–HCl, pH 7.5; 150 mM KCl; 1% Triton X–100; 1 mM EDTA) containing a protease inhibitor cocktail (Roche Diagnostics). After ultrasonic extraction, equal amounts of protein were loaded onto 8% polyacrylamide gels and electrophoresed. The blots were incubated overnight with anti-OLFM4 antibody (1:1,000, Abcam), anti-MUC2 antibody (1:1,000, Abcam) and anti-GAPDH (1:10,000, Sigma). Enhanced chemiluminescence peroxidase-labelled anti-mouse or anti-rabbit antibodies (1:2,000, Amersham) were used for detection.

Single-cell RNA-seq data preprocessing and reads mapping. Paired-end sequencing reads were initially split according to the cell-specific barcodes in reads 2. Simultaneously, the UMI sequences in reads 2 were attached to reads 1 and subsequently used for further analysis. After the TSO or poly(A) sequence was removed from reads 1, quality control was performed to eliminate adapter contamination and low-quality bases. Next, the cleaned sequencing reads were mapped to the human genome (hg19) using TopHat (version 2.0.14), and only uniquely mapped reads were kept[61]. The abundance of the transcripts was estimated by counting the uniquely mapped reads to each gene with HTSeq, and reads with duplicated UMIs of each gene were excluded[62].

Low-quality cell filtering. To filter out low-quality cells, stringent quality filter criteria were applied to each cell, and 77% of fetal cells were maintained for further analysis (4,089 cells post-filtering of 5,290 sequenced cells) and 60% of adult cells passed the filter criteria (1,138 out of 1,891 cells). First, at least 3,000 genes for fetal samples and 2,000 genes for adult samples were detected from each cell. Second, the second maximum pairwise and Pearson correlations were required to be greater than 0.6 (ref. [30]). Third, the mapping ratio of the reads should be higher than 20%. Fourth, the ratio of reads mapping to the mitochondria should be lower than 40%. The single-cell sequencing data were analysed twice (in one set analysed as above; in the other set analysed as above but the percentage of mapping reads and ratio of reads mapping to the mitochondria were not filtered) using two different quality filtrations, and similar results were obtained.

Nonlinear dimensional reduction (t-SNE) analysis and unsupervised clustering using affinity propagation of all quality-filtered cells. The Seurat package was applied to analyse our $\log_2(TPM+1)$ expression data. Only genes detected in at least three cells with expression levels higher than 1 were maintained for subsequent analysis[22]. Highly variable genes were observed using the

'mean.var.plot' function with setting parameters y.cutoff = 2, x.low.cutoff = 2, fxn.x = expMean, fxn.y = logVarDivMean, and these highly variable genes were used to perform PCA. Principal components that can significantly distinguish cells were first selected by jackstraw analysis with 200 replications, after which these principal components were used to perform t-SNE analysis through the run_tsne function of the Seurat package with 2,000 iterations[22,63]. Affinity propagation is an algorithm that could identify clusters with much lower error rate and did so in shorter time to cluster the cells. This algorithm is suitable for our data (consisting of RNA-sequencing data of a large number of cells) because of its high accuracy and because it is less time-consuming to run. Thus, cell populations were defined using the APcluster function with negative squared Euclidean similarity matrix in APcluster, an R package for affinity propagation clustering, which identifies clusters with a low error rate and in a short time[64,65]. To ensure the robustness towards differences in the overall transcriptomes between cells, a two-dimensional t-SNE matrix was used for affinity propagation clustering. After clustering the cells by affinity propagation, we want to explore the relationships among different cell clusters. Unsupervised hierarchical clustering has the advantage that it shows the relationship of clusters much more intuitively than a t-SNE plot. Thus, we subsequently applied hierarchical clustering of the clusters defined using APcluster to explore the relationship of each cluster. On comparing the t-SNE plot to hierarchical clustering, the results of these two methods are, in general, found to be consistent with each other. The cells within the same clusters in t-SNE plots also tend to gather together in the tree of hierarchical clustering. Additionally, as shown in Fig. 1c, there were seven clusters (Groups 6, 7, 8, 14, 15, 27 and 28) from just one fetal sample. We assumed that they were potentially caused by the drastic differentiation and changes in gene expression during human fetal development.

Identification of epithelial non-associated cells. We focused on epithelial cell developmental progression. Thus, we calculated the average expression levels of the marker genes in each cluster and calculated the ratio of cells expressing specific genes in each cluster using the 'dot.plot' function in the Seurat package. Subsequently, we used colour to represent the expression levels and circle sizes to represent the percentage of expressed cells. Based on the detection of cell-type-specific markers, we removed muscle, erythrocytes, immune and endothelial cells from the data set.

DEGs and cluster biomarker identification. For cluster numbers higher than two, cluster-specific marker genes were identified by running the 'find_all_markers' function with parameters thresh.test = 2 and test.use = 'roc.' To identify differentially expressed genes between two clusters, we used the 'find.markers' function with thresh.test = 2 and test.use = 'roc'. All differentially expressed genes of specific clusters are listed in Supplementary Tables 4 to 7. GO enrichment analysis was performed using the DAVID website (https://david.ncifcrf.gov/).

Single-cell trajectories in the pseudotime analysis of each organ. Monocle2 is software used to order cells according to progression along a learned trajectory, and this program tracks changes as a function of progress along the trajectory, termed 'pseudotime'[25,26]. To discover the expression patterns along the development axis of each organ in the GI tract, we reconstructed single-cell trajectories using Monocle2. We first obtained the differentially expressed genes of 7, 14, 16 and 24 weeks of each organ using the Seurat 'find_all_markers' function and subsequently used these genes as input for Monocle2. To largely improve the quality of the trajectory, we called the reduceDimension function with max_components = 2 in Monocle2 to apply a dimensionality reduction to the input data set. After the expression data were projected into a lower-dimensional space, we used the orderCells function to order the cells. To confirm the correct ordering, we coloured the cell according to embryo weeks and observed that the pseudotime matched well with the physiological development time.

Pseudotime normalization of four organs. The four organs were collected from the same embryo, indicating no bias in the physiological developmental time, so the starting time and end time were the same in these four organs. To make the pseudotimes of different organs comparable, we normalized this parameter to the same timescale (1–100) by dividing the largest pseudotime of each organ and subsequently multiplying by 100. After normalizing the pseudotime, we explored and compared the developmental patterns of the four organs.

Fetal cell-type identification with the StemID algorithm. Because t-SNE plots of all cells from the human fetal digestive tract are complex, our major analyses of the cell types were based on StemID analysis and KNN analysis[23,24]. The late stage developmental digestive organs are supposed to be more mature and contain more differentiated cell types than organs from the early stage, so the percentage of stem cells at late stages may be much lower compared with early stages. Thus, we applied StemID to the 25 weeks oesophagus and 24 weeks stomach, S-Intes and L-Intes to identify not only abundant cell types but also rare cell types at 24 or 25 weeks. We initially set up the S4 class object of StemID with the 'SCseq' function using our filtered data as input. We subsequently clustered the data using k-medoids clustering with the 'clusterexp' function by setting parameters FUNcluster = 'kmedoids', SE. method = 'Tibs2001SEmax', bootnr = 50, do.gap = FALSE and cln = 0. After the 'clusterexp' method was executed, by calling the method 'findoutliers' with a default setting, outlier cells were identified and the final clusters were redefined. To visualize the partitioning of the cell population into clusters, the t-SNE map was computed by running the 'comptsne' function. To construct the lineage tree of cell populations, we initialized the lineage tree object by executing the 'Ltree' function and subsequently using the 'compentropy' method to compute the StemID score. Running the function 'projcells' projected the cells onto all inter-cluster links with setting cthr = 4 and nmode = FALSE. The randomization could subsequently be computed by applying the function 'projback' with a default setting. Next, to assemble the lineage tree, the function 'lineagetree' was performed with setting pthr = 0.01. Subsequently, we visualized the lineage tree by running the 'plotmapprjections' function to show the projection of cells onto the inter-cluster links in the t-SNE space. After defining the cell types in the 25 weeks oesophagus, 24 weeks stomach, S-Intes and L-Intes, we used the Seurat package to observe each cell-type-specific markers using the 'find_all_markers' function.

KNN algorithm to allocate single cells to previously identified clusters. After defining cell types in 24/25 weeks embryos, we used the cell-type-specific genes observed in the 24/25 weeks data by applying the knn function of the KNN method (R package) to allocate other stages of cells to previously identified clusters. To improve the accuracy of knn, we used 24/25 weeks samples as a verification data set by using several k and a different number of markers to allocate cells into specific groups. The default value of the parameter 'l' was defined as 0. Then, by comparing the performance at different parameters, we chose k and the number of markers with the best performance. Finally, we set ($k = 10$, $n = 2$), ($k = 7$, $n = 6$), ($k = 8$, $n = 7$) and ($k = 7$, $n = 4$) for oesophagus, stomach, S-Intes and L-Intes, respectively. We used only clusters with more than five cells observed using StemID as training samples and other cells of the same organ as test samples.

Cell cycle analysis. Sets of 43 G1/S and 55 G2/M genes were used in the cell cycle analysis[19,66]. To calculate the ratio of actively proliferating cells of each feature, such as different organs and different developmental stages, we first calculated the total expression levels of all 98 cell cycle genes in each single cell, and then only cells with mean expression level higher than $2(\log_2(\mathrm{TPM}+1))$ were regarded as actively proliferating. We also applied the Cyclone function in scran R package to allocate single cells to G1, G2/M and S phases based on gene expression data[67]. Similar results showed a high coincidence of the two methods.

Adult large intestine cell type identification and comparison with fetal large intestine. The t-SNE analysis was initially conducted using the SCENIC package based on the gene regulation network of adult samples. The cells were clustered based on the regulon binary activity matrix by using hierarchical clustering. DEGs of identified clusters were observed using the Seurat package with the 'find_all_markers' function. By comparing fetal and adult large intestine cells, we initially excluded the immune cells of adult samples and subsequently conducted PCA and t-SNE analysis using the Seurat package. High-loading gene expression levels along the PCA score were drawn using the 'pcheatmap' function.

Transcription master regulators analysis. The algorithm ARACNe is one of most effective tools for accurately reconstructing gene regulatory networks from large-scale molecular profile data sets. The ARACNe-AP implements an adaptive partitioning (AP) strategy for estimating the mutual information (MI), which dramatically improves the computational performance over previous methods. The regulation networks were first built by applying ARACNe-AP software[68]. Running ARACNe-AP involved three key steps: MI threshold estimation, MI network reconstruction and building a consensus network. The full list of human TFs downloaded from Animal TFDB (http://bioinfo.life.hust.edu.cn/AnimalTFDB/) was used during the analysis[69,70]. After obtaining the regulation network, we performed master regulator analysis using the ssmarina package implemented using the MARINa algorithm[68].

Gene expression tendency clustering. To identify clusters of genes with similar expression profiles, we used K-means clustering by running the 'csCluster' function in the cummeRbund R package. 'csCluster' uses the 'pam' method from the clustering package to perform partitioning around the medoids, and the Jensen–Shannon distance matrix was used for clustering.

TF correlation network construction. Covariance network analysis and visualization were achieved using igraph implemented in R (http://igraph.sf.net). We first computed a pairwise correlation matrix for TFs annotated in Animal TFDB and TFs with a correlation of greater than 0.3, while at least three other TFs were retained for further analysis. We used the pairwise correlation of selected TFs as input into the function 'graph.adjacency' in the igraph package to generate a weighted network graph. The networks were visualized using the Fruchterman–Teingold layout. The vertices represent selected TFs and the edges link the vertices with pairwise correlations >0.25. Orange and red represent the highest average expression in early, mid and late stages, respectively.

Statistics and reproducibility. Statistical significance was calculated by two-sided Mann–Whitney U-test. *P* values lower than 0.01 were considered statistically significant. All error bars in this manuscript represent standard error of the mean (s.e.m.). DEGs were identified by the Seurat package. Pseudotime was calculated by the Monocle2 package as described. Boxes indicate upper and lower quartiles (25th and 75th percentiles), and whiskers indicate the 1.5 interquartile range. The middle hinge represent the median number (50th percentile). The density of the violin plots was scaled to a maximum of 1 by setting 'scale = area', and all violins have the same maximum width. For single-cell sequencing experiments, the two embryos at the same developmental stage (for 6, 7 and 8 W) were considered two biological replicates as they were dissected separately and the single-cell sequencing was performed separately. Single-cell sequencing data were analysed twice using two different quality filtrations and similar results were obtained. All immunostaining experiments were performed twice with similar results, and the data shown in Figs. 2b,f, 3d and 8d,e,i,j and Supplementary Figs. 3c,e,h, 4 and 5a,b are representative images. Western blotting data in Supplementary Fig. 2f are representative images of four independent experiments.

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

Data availability. Single-cell RNA-seq data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under accession codes GSE103239. All other data supporting the findings of this study are available from the corresponding authors on reasonable request.

References

61. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-seq. *Bioinformatics* **25**, 1105–1111 (2009).
62. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
63. Chung, N. C. & Storey, J. D. Statistical significance of variables driving systematic variation in high-dimensional data. *Bioinformatics* **31**, 545–554 (2015).
64. Bodenhofer, U., Kothmeier, A. & Hochreiter, S. APCluster: an R package for affinity propagation clustering. *Bioinformatics* **27**, 2463–2464 (2011).
65. Frey, B. J. & Dueck, D. Clustering by passing messages between data points. *Science* **315**, 972–976 (2007).
66. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
67. Scialdone, A. et al. Computational assignment of cell-cycle stage from single-cell transcriptome data. *Methods* **85**, 54–61 (2015).
68. Lachmann, A., Giorgi, F. M., Lopez, G. & Califano, A. ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinformatics* **32**, 2233–2235 (2016).
69. Zhang, H. M. et al. AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors. *Nucleic Acids Res.* **43**, 76–81 (2015).
70. Zhang, H. M. et al. AnimalTFDB: a comprehensive animal transcription factor database. *Nucleic Acids Res.* **40**, 144–149 (2012).

nature research

Corresponding author(s): Fuchou Tang

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

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

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	tophat version 2.0.14, samtools version 0.1.19 , python 2.7 , HTSeq 0.6.1 ,R version 3.4.0, ARACNe, R package Seurat 1.4.0.16, monocle 2.4.0, ggplot2(2.2.1), cluster(2.0.6), RColorBrewer(1.1-2), RaceID_StemID, apcluster(1.4.3), igraph(1.0.1), scran (1.4.5), SCORPIUS, diffusionMap(1.1-0), ssmaria, LSM 510 META microscope (Zeiss)
Data analysis	Generally, The Seurat package was applied to analyze our log2(TPM+1) expression data. Only genes detected in at least three cells with expression levels higher than 1 were maintained for subsequent analysis. Highly variable genes were observed using the 'mean.var.plot' function with setting parameters y.cutoff=2, x.low.cutoff=2, fxn,x=expMean, fxn.y=logVarDivMean, and these highly variable genes were used to perform PCA (principal component analysis). Principal components that can significantly distinguish cells were first selected by jackstraw analysis with 200 replications, after which these principal components were used to perform t-SNE analysis through the run_tsne function of the Seurat package with 2,000 iterations. Affinity propagation is an algorithm that could identify clusters with much lower error rate and did so in shorter time to cluster the cells. The algorithm is suitable for our data consisting of RNA-sequencing data of large number of cells due to high accuracy and less time-consuming. Thus, cell populations were defined using the APcluster function with negative squared Euclidean similarity matrix in APcluster, an R package for affinity propagation clustering, which identifies clusters with a low error rate and in a short time . To ensure the robustness toward differences in the overall transcriptomes between cells, the two-dimension t-SNE matrix was used for affinity propagation clustering. After clustering the cells by affinity propagation, we want to explore the relationship among different cell clusters. The unsupervised hierarchical clustering has the advantage that it shows the relationship of clusters much more intuitive than t-SNE plot. Thus, we subsequently applied hierarchical clustering of the clusters defined using APcluster to explore the relationship of each cluster. For cluster numbers higher than two, the cluster-specific marker genes were identified by running the 'find_all_markers' function with

parameters thresh.test=2 and test.use='roc.' To identify differentially expressed genes between two clusters, we used the 'find.markers' function with thresh.test = 2 and test.use ='roc'.
Monocle2 is software used to order cells according to progression along a learned trajectory, and this program tracks changes as a function of progress along the trajectory, termed 'pseudotime'.
We applied StemID to the 25W esophagus and 24W stomach, S-Intes and L-Intes to identify not only abundant cell types but also rare cell types at 24W or 25W. We initially set up the S4 class object of StemID with 'SCseq' function using our filtered data as input. We subsequently clustered the data using k-medoids clustering with the 'clusterexp' function by setting parameters FUNcluster='kmedoids', SE. method='Tibs2001SEmax', bootnr=50, do.gap=FALSE, and cln=0. After the 'clusterexp' method was executed, by calling the method 'findoutliers' with a default setting, the outlier cells were identified and the final clusters were redefined. To visualize the partitioning of the cell population into clusters, the t-SNE map was computed by running the 'comptsne' function. To construct the lineage tree of cell populations, we initialized the lineage tree object by executing the 'Ltree' function and subsequently using the 'compentropy' method to compute the StemID score. Running the function 'projcells' projected the cells onto all inter-cluster links with setting cthr=4 and nmode=FALSE. The randomization could subsequently be computed by applying the function 'projback' with a default setting. Next, to assemble the lineage tree, the function 'lineagetree' was performed with setting pthr=0.01. Subsequently, we visualized the lineage tree by running the 'plotmapprjections' function to show the projection of cells onto the inter-cluster links in the t-SNE space. After defining the cell types in the 25W esophagus, 24W stomach, S-Intes and L-Intes, we used the Seurat package to observe each cell-type-specific markers using the 'find_all_markers' function.
After defining cell types in 24W/25W embryos, we used the cell-type-specific genes observed in the 24W/25W data by applying the knn function of the KNN method (R Package) to allocate other stages of cells to previously identified clusters.
Sets of 43 G1/S and 55 G2/M genes were used in the cell cycle analysis. Cyclone function in scran R package was used to allocate single cell to G1, G2/M and S phases based on the gene expression data.
The regulation networks were first built by applying ARACNe-AP software. Then we performed master regulator analysis using the ssmarina package implemented using the MARINa algorithm.
To identify clusters of genes with similar expression profiles, we used K-means clustering by running the 'csCluster' function in the cummeRbund R package.
Immunostaining cells mounted on slides were observed on a LSM 510 META microscope (Zeiss) using Plan Neofluar 63×/1.4 Oil DIC objective.

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Single-cell RNA–seq data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under accession codes GSE103239. All other data supporting the findings of this study are available from the corresponding author on reasonable request.

Field-specific reporting

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

☒ Life sciences ☐ Behavioural & social sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences

Study design

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

Sample size	To systematically investigate the gene expression pattern during human fetal GI tract development, the digestive tracts of 15 human embryos ranging from 6W to 25W were analyzed using single-cell RNA-seq analysis. The esophagus, stomach, small and large intestine—the four main organs of the digestive tract—were isolated from 14 of these 15 embryos. In total, 5,290 fetal GI single cells covering different cell types were selected and 4,089 cells passed quality control. Subsequently, to systematically explore the transcriptome dynamics from fetal to adult stage, 1,138 individual cells from adult L-Intes were analyzed using single-cell RNA-seq after filtration.
Data exclusions	We retain cells that the number of genes > 3,000 and top correlation coefficients with any other two cells > 0.6, the mapping ratio of the reads should be higher than 20%,the ratio of reads mapping to the mitochondria should be lower than 40% and the rest of cells were excluded.
Replication	All attempts at replication were successful.
Randomization	All single cells were allocated into different groups according to their development stages.

Blinding

The investigators were not blinded to group allocation during data collection and/or analysis. The main purpose of the manuscript is to investigate the gene expression profiles of human fetal digestive tract according to the actual human fetal developmental stage. Single cells of collected samples were randomly picked with mouth pipette, and library preparation and sequencing was randomized to avoid batch effects. The RNA-seq data can nearly cover whole transcriptiome and do not cause any bias.

Materials & experimental systems

Policy information about availability of materials

n/a	Involved in the study
☒	☐ Unique materials
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Research animals
☐	☒ Human research participants

Antibodies

Antibodies used

anti-CDH1;Abcam;ab24590;1:100;Immunofluorescent analysis of intestine in fetal and early stage. http://www.abcam.cn/cd31-antibody-p2b1-ab24590.html

anti-VILLIN;Abcam;ab201989;GR286745-7;1:100;Immunofluorescent analysis of intestine in fetal and early stage.http://www.abcam.cn/villin-antibody-3e5g11-n-terminal-ab201989.html

anti-MUC2;Abcam;ab11197;1:100 for Immunofluorescent analysis, 1:1000 for Western blotting;Immunofluorescent and Western boltting analysis of intestine in fetal and early stage. http://www.abcam.cn/muc2-antibody-9961-ab11197.html

anti-TFF3;Abcam;ab101099;GR131940-2;1:100;Immunofluorescent analysis of intestine in fetal and early stage.http://www.abcam.cn/trefoil-factor-3-antibody-ab101099.html

anti-CHGA;Abcam;ab715;GR301307-5,1:100;Immunofluorescent analysis of intestine in fetal and early stage.http://www.abcam.cn/chromogranin-a-antibody-lk2h10-phe5-ab715.html

anti-PYY;Abcam;ab22663;GR306687-5;1:200;Immunofluorescent analysis of intestine in fetal and early stage.http://www.abcam.cn/peptide-yy-antibody-ab22663.html

anti-LYZ;Abcam;ab108508;GR291230-6;1:200;Immunofluorescent analysis of intestine in fetal and early stage.http://www.abcam.cn/lysozyme-antibody-epr29942-ab108508.html

anti-DCLK1;Abcam;ab31704;GR310497-1;1:100;Immunofluorescent analysis of intestine in fetal and early stage.http://www.abcam.cn/dcamkl1-antibody-ab31704.html

anti-SOX9;Abcam;GR298822-4;1:100;Immunofluorescent analysis of intestine in fetal and early stage.http://www.abcam.cn/sox9-antibody-3c10-ab76997.html

anti-OLFM4;Abcam;GR276252-1;1:100 for Immunofluorescent analysis, 1:1000 for Western blotting;Immunofluorescent and Western blotting analysis of intestine in fetal and early stage.http://www.abcam.cn/olfm4-antibody-ab85046.html

anti-Annexin A1;Abcam;ab214486;GR228356-7;1:500;Immunofluorescent analysis of esophagus and small intestine in fetal stage.http://www.abcam.cn/annexin-a1-antibody-epr19342-ab214486.html

anti-CD31;Abcam;ab24590;GR310739-1;1:100;Immunofluorescent analysis of esophagus in fetal stage. http://www.abcam.cn/cd31-antibody-p2b1-ab24590.html

anti-MUC16;Abcam; ab1107;GR290946-2;1:100; Immunofluorescent analysis of esophagus in fetal stage. http://www.abcam.cn/muc16-antibody-x75-ab1107.html

anti-Cytokeratin 6(KRT6A); Abcam; ab52620;GR63452-10;1:100; Immunofluorescent analysis of esophagus in fetal stage. http://www.abcam.cn/cytokeratin-6-antibody-epr1603y-ab52620.html

anti-Cytokeratin 6B; Thermo Scientific; PA5-29134;SF2398911C; 1:200; Immunofluorescent analysis of esophagus in fetal stage. https://www.thermofisher.com/antibody/product/Cytokeratin-6B-Antibody-Polyclonal/PA5-29134

anti-LRIG1;Abcam;ab214102;GR281748-2;1:100;Immunofluorescent analysis of stomachs in fetal stage. http://www.abcam.cn/lrig1-antibody-lrig121-ab214102.html

anti-CD14;Abcam; ab28061; GR248603-3;1:200; Immunofluorescent analysis of intestine in fetal and adult stage. http://www.abcam.cn/cd14-antibody-mem-15-fitc-ab28061.html

anti-CD34;Abcam; ab81289; GR201207-27;1:100; Immunofluorescent analysis of intestine in fetal and adult stage. http://www.abcam.cn/cd34-antibody-ep373y-ab81289.html

anti-GAPDH; Sigma; G9545; 1:10000; Western blotting analysis of intestine in fetal stage. https://www.sigmaaldrich.com/catalog/product/sigma/g9545?lang=zh&region=CN

Secondary antibodies:

Donkey anti-mouse AlexaFluor®568; Thermo Fisher Scientific; 1:1000; A10037. For Immunofluorescent staining. https://www.thermofisher.com/antibody/product/Donkey-anti-Mouse-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A10037

Donkey anti-rabbit AlexaFluor®647; Thermo Fisher Scientific; 1:1000; A31573. For Immunofluorescent staining. https://www.thermofisher.com/antibody/product/Donkey-anti-Rabbit-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-31573

Donkey anti-goat AlexaFluor®488; Thermo Fisher Scientific; 1:1000; A11055. For Immunofluorescent staining. https://www.thermofisher.com/antibody/product/Donkey-anti-Goat-IgG-H-L-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-11055

Goat anti-Rabbit AlexaFluor®488; Thermo Fisher Scientific; 1:1000; A11034. For Immunofluorescent staining. https://www.thermofisher.com/antibody/product/Goat-anti-Rabbit-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-11034

Anti-mouse IgG, HRP-linked Antibody; Cell Signaling Technology; 1:5000; 7076s. For Western blotting. https://www.cellsignal.com/products/secondary-antibodies/anti-mouse-igg-hrp-linked-antibody/7076?N=4294956287&Ntt=7076s&fromPage=plp&_requestid=2784877

Anti-rabbit IgG, HRP-linked Antibody; Cell Signaling Technology; 1:5000; 7074s. For Western blotting. https://www.cellsignal.com/products/secondary-antibodies/anti-rabbit-igg-hrp-linked-antibody/7074?site-search-type=Products&N=4294956287&Ntt=7074s&fromPage=plp&_requestid=2784897

Validation

All the antibodies used in the project were from the commercial antibodies and were validated based on the information on the manufacturers' instructions. The utility of these antibodies is stated on the websites of the suppliers listed in the above table.

Human research participants

Policy information about studies involving human research participants

Population characteristics

The information of the two adults who were used to collect large intestine were list below:

Case 1
She is a 47-year old woman who was diagnosed as sigmoid colon cancer. Laparoscopic left hemicolectomy and lymph node dissection was performed in May 19,2017. Histopathological examination revealed moderately differentiated adenocarcinoma, the AJCC stage of the cancer was stage II, T3N0M0. The tumor harbored Kras exon2 mutation, and showed MSH2, MSH6 deficiency.
Case2
He is a 53-year old man who was diagnosed as sigmoid colon cancer. Laparoscopic left hemicolectomy and lymph node dissection was performed in June 1,2017. Histopathological examination confirmed poorly differentiated adenocarcinoma, the AJCC stage of the cancer was stage III, T3N2M0. The tumor did not show MMR deficiency.

The information of the 15 embryos were list below:
Age: 6W, 2 embryos; 7W, 2 embryo; 8W, 2 embryo; 9W, 1 embryo; 11W, 1 embryo; 14W, 1 embryo; 16W, 1 embryo; 19W, 1 embryo; 21W, 2 embryos (twins); 24W, 1 embryo; 25W, 1 embryo.
Gender: Unknown.
In summary, the 15 embryos were obtained from all couples who sign the informed consent and cannot continue pregnancy due to severe pregnancy disease or severe fetal anomaly.

Method-specific reporting

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ Magnetic resonance imaging