Stem Cell Reports, Volume *17*

Supplemental Information

Acquisition of NOTCH dependence is a hallmark of human intestinal stem cell maturation

Yu-Hwai Tsai, Angeline Wu, Joshua H. Wu, Meghan M. Capeling, Emily M. Holloway, Sha Huang, Michael Czerwinkski, Ian Glass, Peter D.R. Higgins, and Jason R. Spence

Supplemental Information

Acquisition of NOTCH dependence is a hallmark of human intestinal stem cell maturation

Authors: Yu-Hwai Tsai[1,#], Angeline Wu[1,#], Joshua H. Wu[1], Meghan M. Capeling, Emily M. Holloway[2], Sha Huang[1], Michael Czerwinkski[1], Ian Glass[4], Peter D.R. Higgins[1], Jason R. Spence[1,2,3,*]

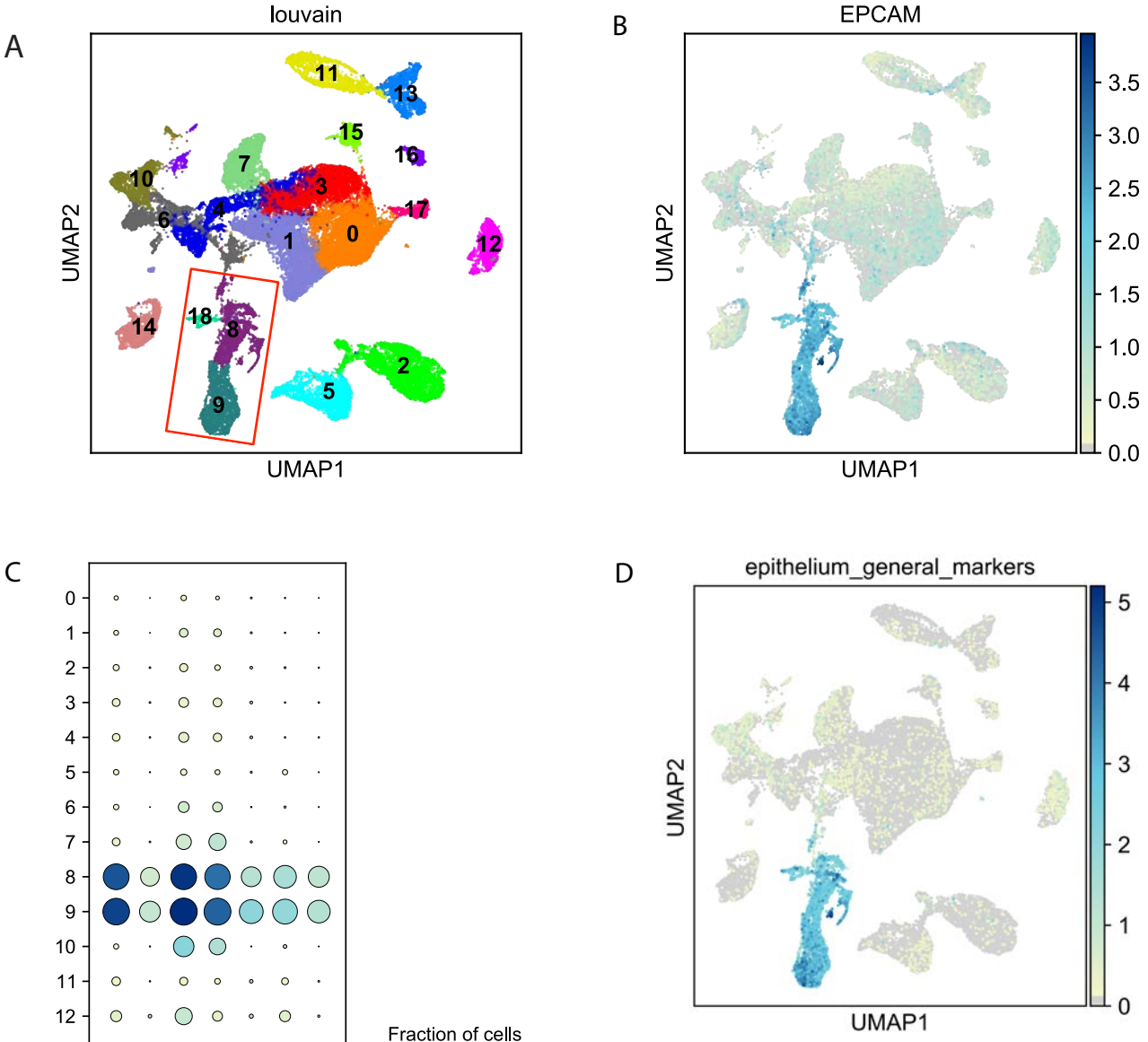
1. Department of Internal Medicine, Gastroenterology, University of Michigan Medical School, Ann Arbor, MI 48109, USA;

2. Department of Cell and Developmental Biology, University of Michigan Medical School, Ann Arbor, MI 48109, USA;

3. Department of Biomedical Engineering, University of Michigan College of Engineering, Ann Arbor, MI 48109, USA;

4. Department of Pediatrics, Genetic Medicine, University of Washington, Seattle, Washington;

Supplemental Figure 1.

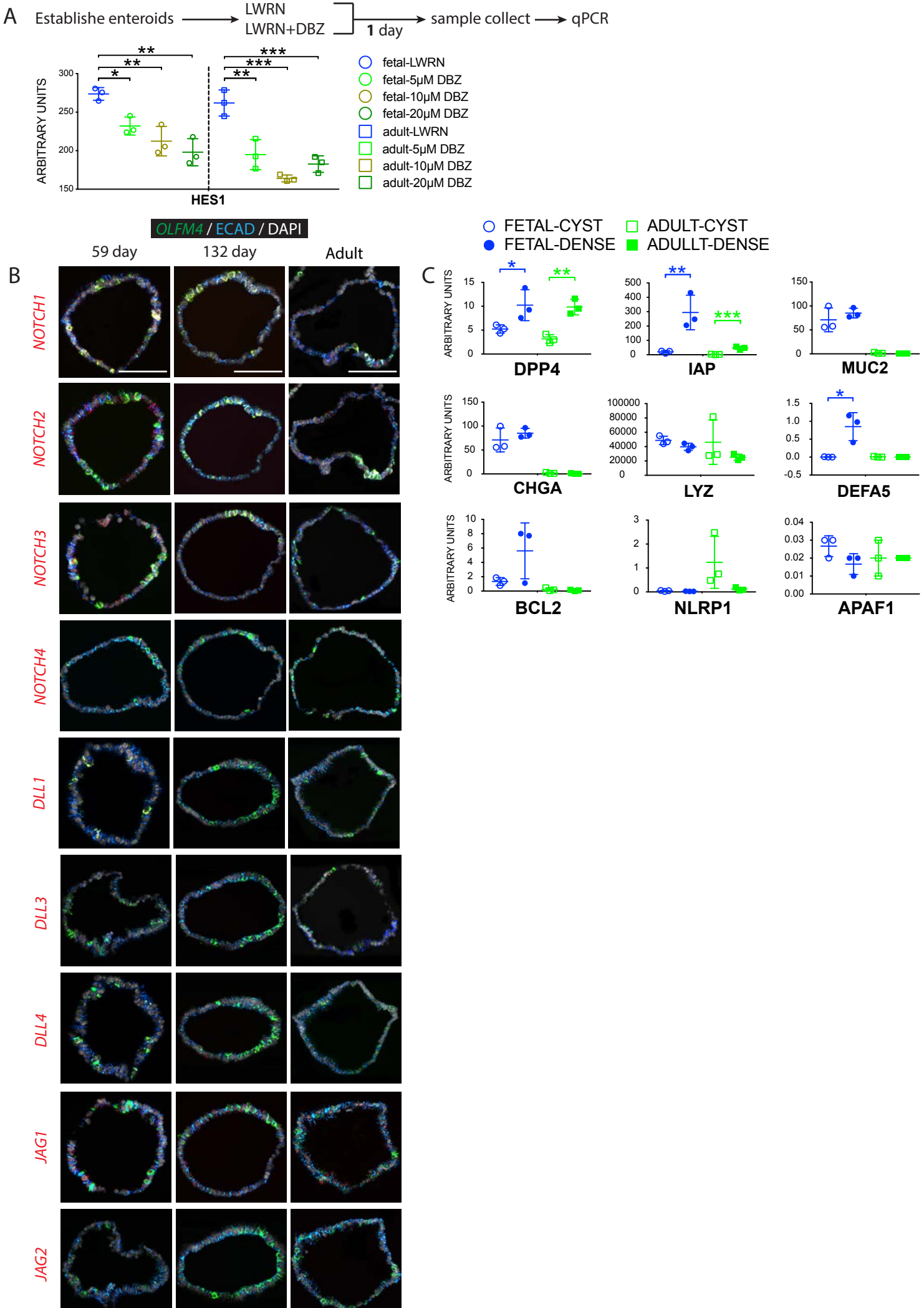


Supplemental Figure 1. **Extracted *EPCAM*$^{+}$ cells by scRNA seq analysis across fetal development, Related to Figure 1.** (A) UMAP visualizations of clusters profiled by canonical gene. *EPCAM*$^{+}$ cells with high expression in clusters 8, 9 and 19, indicated by the red outline. (B) Feature plot *EPCAM*$^{+}$ cells by scRNA seq analysis from fetal tissues ages 47 to 132 days post conception. (C) Dot plot of all epithelium+ markers including *EPCAM, CDH1, KRT8, KRT18, FXYD3, PERP* and *CDX2* according to https://pubmed.ncbi.nlm.nih.gov/34019796/. (D) Feature plot of epithelium+ markers corresponding to (C) by scRNA seq analysis.

Supplemental Figure 2.

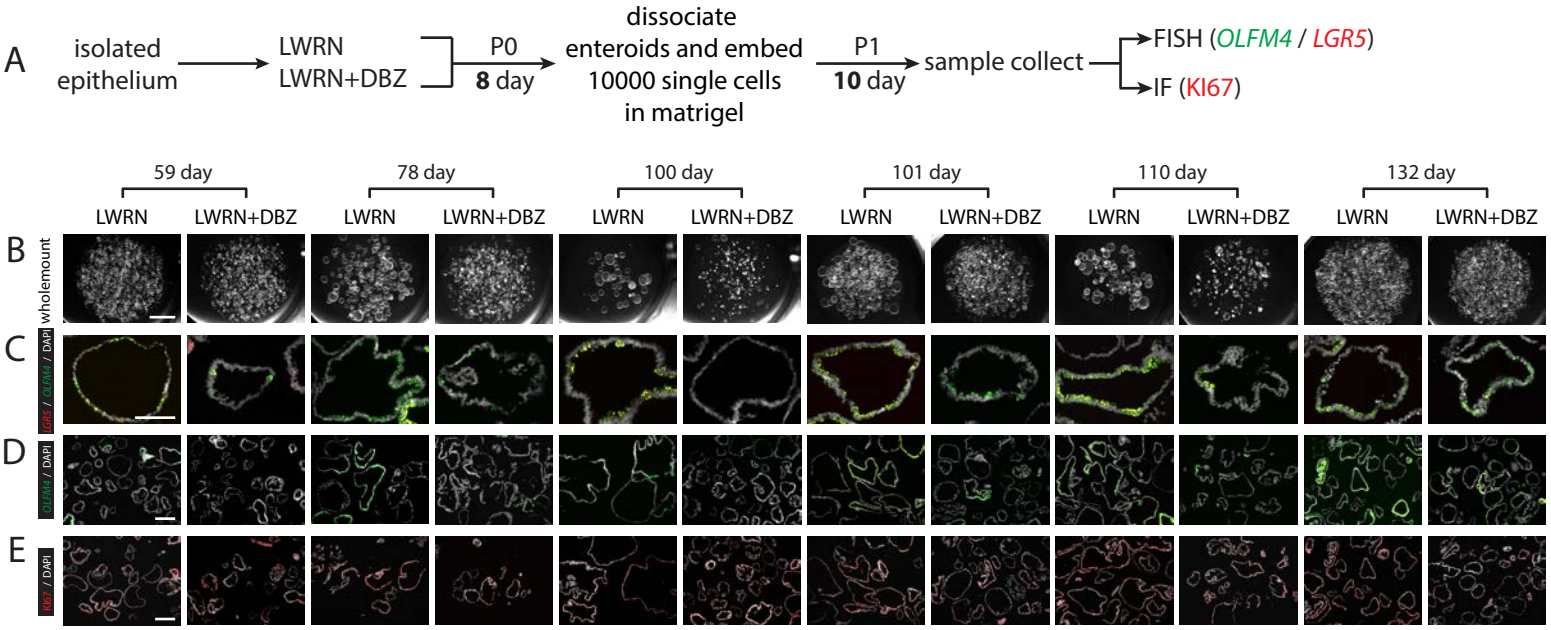
Supplemental Figure 2. ***OLFM4* gene expression increases concurrently with age throughout the intestine, Related to Figure 1.** (A) IF protein staining for OLFM4 (red) and ECAD (blue) with DAPI (grey) on fetal duodenum aged 70, 101, 103, 110, 120, 132 and 145 days post conception. (B) In situ hybridization (ISH) staining for *OLFM4* (DAB) of fetal duodenum, jejunum and ileum from respective timepoints corresponding to panel A (n=1 biological replicate per each timepoint). Scale bars represent 100 μm.

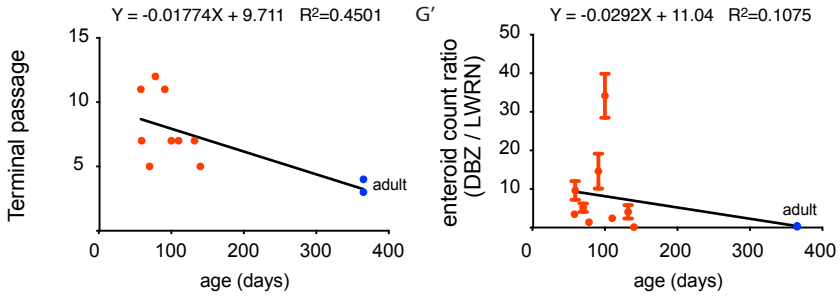
Supplemental Figure 3.

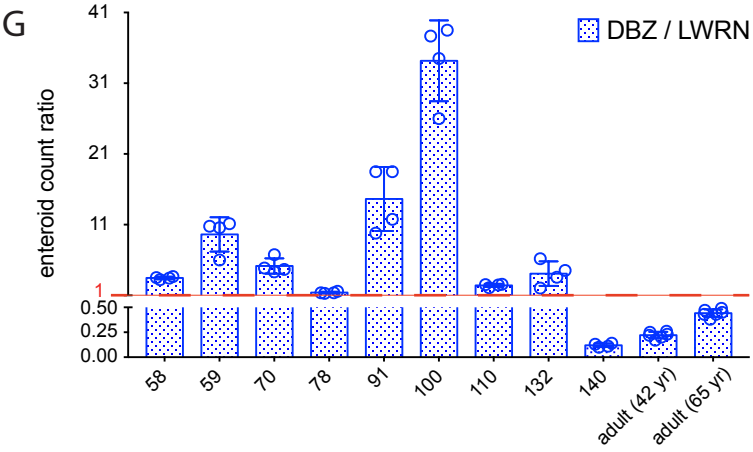


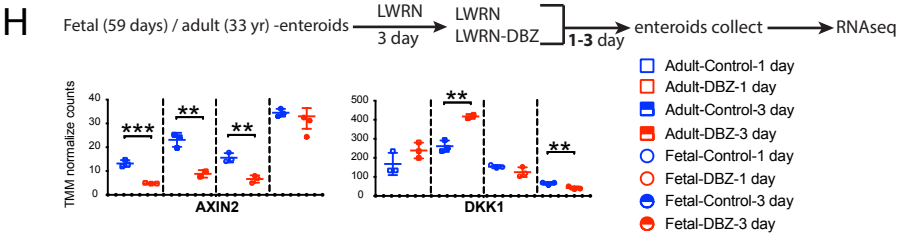
Supplemental Figure 3. **Related to Figures 2, 3 and 4.** (A) Real-time PCR analysis of NOTCH target gene *HES1*, investigating the dosage-dependent response under DBZ, γ-secretase inhibition, of fetal enteroids (142 days post conception) and adult (33 years old) at 5, 10 and 20 μM (n=1 biological replicate, data points represent experimental replicates). (B) FISH staining for *OLFM4* (green), *NOTCH1*, *NOTCH2*, *NOTCH3, NOTCH4, DLL1, DLL3, DLL4, JAG1* and *JAG2* (red) and IF staining for ECAD (blue) with DAPI (grey) on respective timepoints of enteroids corresponding to panel Figure 2B. Scale bars represent 100 μm. (C) Real-time PCR analysis of cell markers, *DPP4*, *IAP*, *MUC2*, *CHGA*, *LYZ*, *DEFA5* and apoptotic markers, *BCL2*, *NLRP1* and *APAF1* (n=1 biological replicate, data points represent experimental replicates). All statistics were analyzed with unpaired t-tests by GraphPad Prism 7.0 and data are presented as the Mean +/- SEM. In all figures, * = $P<0.05$, ** = $P<0.01$, *** = $P<0.001$, **** = $P<0.0001$.

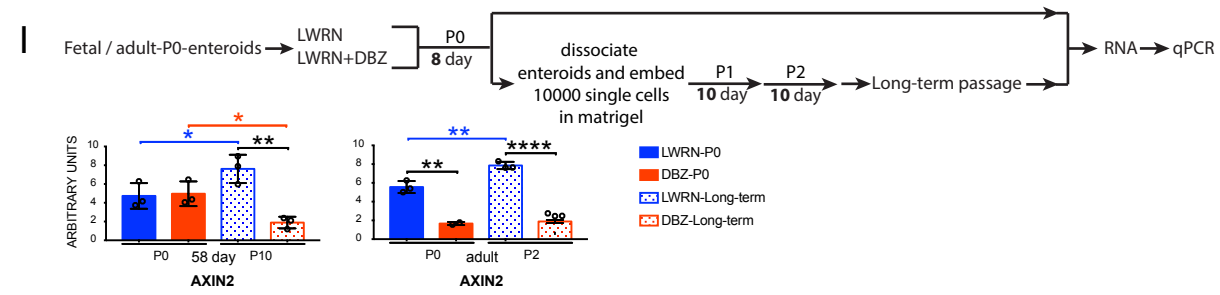
Supplemental Figure 4.











Supplemental Figure 4. **Related to Figures 4, 5 and 6.** (A) Experimental schematic for data presented S4B-E. (B) Stereomicroscope images of P1 enteroids derived from fresh epithelium of fetal duodenum aged 59, 78, 100, 101, 110 and 132 days post conception after LWRN and LWRN+DBZ treatments for 10 days. Scale bar represents 2 mm. (C) FISH staining for *OLFM4* (green) and *LGR5* (red) and IF staining for ECAD (blue) with DAPI (grey) on respective timepoints corresponding to panel B. Scale bar represents 100 μm. (D) FISH staining for *OLFM4* (green) with DAPI (grey) on respective timepoints corresponding to panels B and C. Scale bar represents 200 μm. (E) IF staining for KI67 (red) with DAPI (grey) on respective timepoints corresponding to panels B, C and D (n=1 biological replicate per each timepoint). Scale bar represents 200 μm. (F) Linear regression line plot between fetal age and enteroids passage number (adult age was used as 365 days) corresponding to Figure 6C, (G) Bar chart plot between enteroid count ratio of DBZ treatment to control corresponding to Figure 6D. (G') Linear regression line plot corresponding to (G). All plots were generated by GraphPad Prism Prism 7.0. (H) TMM normalized counts of *AXIN2* and *DKK1* on LWRN-control and LWRN-DBZ for fetal (59 days post conception) and adult (33 years old) enteroids (n=1 biological replicate, data points represent experimental replicates). (I) qPCR analysis of *AXIN2* gene expression on LWRN-control and LWRN-DBZ for fetal (58 days post conception) and adult (65 years old) enteroids (n=1 biological replicate, data points represent experimental replicates). All statistics were analyzed with unpaired t-tests by GraphPad Prism 7.0 and data are presented as the Mean +/- SEM. In all figures, * = $P<0.05$, ** = $P<0.01$, *** = $P<0.001$, **** = $P<0.0001$.

SUPPLEMENTAL METHODS

QUANTIFICATION AND STATISTICAL ANALYSIS

For statistical analysis, data are expressed as the median of each sample set with each data point represented in the plots. As noted in figure legends, unpaired t-tests, were carried out with GraphPad Prism 7.0 and data are presented as the Mean +/- SEM. In all figures, * = P<0.05, ** = P<0.01, *** = P<0.001, **** = P<0.0001. All experiments were conducted on at least 3 independent biological replicates.

Computational analysis of single-cell RNA sequencing data

Overview

To visualize distinct cell populations within the single-cell RNA sequencing dataset, we employed the general workflow outlined by the Scanpy Python package (Wolf et al., 2018). This pipeline includes the following steps: filtering cells for quality control, log normalization of counts per cell, extraction of highly variable genes, regressing out specified variables, scaling, reducing dimensionality with principal component analysis (PCA) and uniform manifold approximation and projection (UMAP) (McInnes et al., 2018), and clustering by the Louvain algorithm (Blondel et al., 2008).

Sequencing data and processing FASTQ reads into gene expression matrices

All single-cell RNA sequencing was performed at the University of Michigan Advanced Genomics Core with an Illumina Novaseq 6000. The 10x Genomics Cell Ranger pipeline was used to process raw Illumina base calls (BCLs) into gene expression matrices. BCL files were demultiplexed to trim adaptor sequences and unique molecular identifiers (UMIs) from reads. Each sample was then aligned to the human reference genome (hg19) to create a filtered feature bar code matrix that contains only the detectable genes for each sample.

Quality control

To ensure quality of the data, all samples were filtered to remove cells expressing fewer than 500 genes or greater than 7000 genes, with high UMI counts greater than 30000, or a fraction of mitochondrial genes greater than 0.1.

Normalization and Scaling

Data matrix read counts per cell were log normalized, and highly variable genes were extracted. Using Scanpy's simple linear regression functionality, the effects of total reads per cell and mitochondrial transcript fraction were removed. The output was then

scaled by a z-transformation. Following these steps, a total of 37484 cells, 2256 genes (Figure S1A-C) were kept for clustering and visualization.

Variable Gene Selection

Highly variable genes were selected by splitting genes into 20 equal-width bins based on log normalized mean expression. Normalized variance-to-mean dispersion values were calculated for each bin. Genes with log normalized mean expression levels between 0.125 and 3 and normalized dispersion values above 0.5 were considered highly variable and extracted for downstream analysis.

Dimension Reduction and Clustering

Principal component analysis (PCA) was conducted on the filtered expression matrix followed. Using the top principal components, a neighborhood graph was calculated for the nearest neighbors (Figure 1A-C- 11 principal components, 15 neighbors; Figure 2 A and C – 11 principal components, 15 neighbors; Figure S1A-C – 20 principal components, 30 neighbors). The UMAP algorithm was then applied for visualization on 2 dimensions. Using the Louvain algorithm, clusters were identified with a resolution of (Figure 1A-C – 0.4; Figure 2 A and C – 0.5; Figure S1A-C – 0.5).

Cluster Annotation

Using canonically expressed gene markers, each cluster's general cell identity was annotated. Markers utilized include epithelium, EPCAM, stem cell markers (OLFM4, LGR5) and NOTCH components (NOTCH1, NOTCH2, NOTCH3, NOTCH4, DLL1, DLL3, DLL4, JAG1, JAG2).

Sub-clustering

After annotating clusters within the UMAP embedding, specific clusters of interest were identified for further sub-clustering and analysis. The corresponding cells were extracted from the original filtered but unnormalized data matrix to include (Figure 1A-C 3665 cells, Figure 2C – 343 cells for C2, 884 cells for C0, 212 cells for C8). The extracted cell matrix then underwent log normalization, variable gene extraction, linear regression, z transformation, and dimension reduction to obtain a 2-dimensional UMAP embedding for visualization.

Supplemental Table 3

Antibody	Source	Dilution
ECAD	BD Transduction Lab, 610181	1:500
ECAD	R&D systems, AF748	1:100
KI67	Thermo Scientific, Sp6 RM-9106-S1	1:400
OLFM4	Abcam, ab85046	1:500
In Situ Probe	**Source**	
HS-DLL1-C1	ACD-RNAscope®Probe, 532631	
HS-DLL3-C1	ACD-RNAscope®Probe, 411338	
HS-DLL4-C1	ACD-RNAscope®Probe, 603001	
HS-JAG1-C1	ACD-RNAscope®Probe, 546181	
HS-JAG2-C1	ACD-RNAscope®Probe, 460491	
HS-LGR5-C1	ACD-RNAscope®Probe, 311021	
HS-NOTCH1-C1	ACD-RNAscope®Probe, 311861	
HS-NOTCH2-C1	ACD-RNAscope®Probe, 488101	
HS-NOTCH3-C1	ACD-RNAscope®Probe, 558991	
HS-NOTCH4-C1	ACD-RNAscope®Probe, 409631	
HS-OLFM4-C1	ACD-RNAscope®Probe, 311041	
HS-OLFM4-C2	ACD-RNAscope®Probe, 311041	