Supplemental information

A cell atlas of microbe-responsive processes in the zebrafish intestine

Reegan J. Willms, Lena O. Jones, Jennifer C. Hocking, and Edan Foley
Table S2. Classification of 16S rRNA gene sequence datasets. Relative abundance is shown after removal of taxa that were <1% abundant. Related to Figures 1-6.

| Phylum | Firmicutes | | | Proteobacteria | |
|---|---|---|---|---|---|
| | Bacilli | Alphaproteobacteria | Betaproteobacteria | Gammaproteobacteria |
| Class | Bacillales | Lactobacillales | Acetobacterales | Rhizobiales | Burkholderiales | Neisseriales | Alteromonadales | Aeromonadales | Pseudomonadales | Vibronales |
| Order | Bacillales | Lactobacillales | Acetobacterales | Rhizobiales | Burkholderiales | Neisseriales | Alteromonadales | Aeromonadales | Pseudomonadales | Vibronales |
| CV1 | 0.78 | 6.20 | 4.91 | 0.25 | 0.58 | 0.26 | 0.46 | 0.72 | 1.64 | 84.21 |
| CV2 | 1.07 | 22.25 | 6.26 | 2.90 | 1.56 | 2.53 | 2.71 | 12.32 | 12.37 | 36.03 |
| CV3 | 0.56 | 35.06 | 2.76 | 1.18 | 1.18 | 3.09 | 2.28 | 3.75 | 6.06 | 43.69 |
Table S3. Description of 16s rRNA gene sequencing datasets used in this study and associated metadata. Related to Figures 1-6.

| Sample Name | # Input Reads | Reads post-filtration | Target | 515F-Primer | 816R-Primer | Barcode-F | Barcode-R |
|-------------|---------------|-----------------------|--------|-------------|-------------|-----------|-----------|
| CV1         | 174987        | 153739                | 16S-V4 | GTGCCAGCM    | GGACTACHVG  | CCAACA    | CGATGT    |
| CV2         | 163143        | 148080                |        | GCCGCGGTAA   | GGTWTCTAAT  | CCAACA    | TGACCA    |
| CV3         | 172703        | 139893                |        | CCAACA      | GCCAAT      | CCAACA    | GCCAAT    |
Figure S1. Isolation of single cells via fluorescence activated cell sorting. (A-B) Gating strategy for isolating dissociated single intestinal cells from CV (A) and GF (B) fish for replicate 1 of the single cell isolation protocol. Forward and side scatter were used to determine the single cell population, and Zombie Aqua viability die was used to select live cells. Related to Figure 1.
Figure S2. Transcriptionally distinct cell populations in CV and GF zebrafish intestines. (A) 2D t-SNE projections of profiled cells, color coded by cluster and annotated numerically. (B) t-SNE plot of overlapping and color-coded CV and GF cells from the integrated dataset. Related to Figure 1.
Figure S3. Identification of tuft-like cells in the zebrafish intestinal epithelium. (A) GO enrichment analysis of tuft-like cells, based on tuft-like cell genetic markers from the conventional single-cell RNA sequencing dataset. Top 8 GO terms are shown. Enrichment score is represented by bar length and p-value is indicated with white circles. (B-C) Expression of mouse orthologues for the top 50 zebrafish tuft-like cell marker genes in a single-cell RNA sequencing dataset of the mouse small intestinal epithelium from Haber et al., 2017, generated using the Broad Single Cell Portal. 32/50 genes had orthologues that were detected in the mouse small intestinal dataset. (B) Heatmap of mouse orthologue expression per mouse intestinal epithelial cell type, colored by scaled mean expression, where the size of the dot indicates proportion of expressing cells per cell type. (C) TSNE plot of mouse small intestinal epithelial cells, showing mean expression (Log2(TPM+1)) of 32 tuft-like marker gene orthologues per cell. Enrichment is evident in the annotated tuft cell clusters.
(D) Transmission electron micrograph of the adult zebrafish posterior intestinal epithelium. Yellow arrowhead points to an apical tuft protruding through the epithelial brush border. Related to Figure 1.
Figure S4. Analysis of zebrafish Best/Otop2 cell expression markers in the human colonic epithelium. Expression of human orthologues for the top 50 zebrafish Best4/Otop2 cell marker genes in a single-cell RNA sequencing dataset of the human colonic epithelium from Smillie et al., 2019, generated using the Broad Single Cell Portal (accession SCP259). 35/50 genes had orthologues that were detected in the human colonic dataset. Heatmap of human orthologue expression per epithelial cell type is shown, colored by scaled mean expression, where the size of the dot indicates proportion of expressing cells per cell type. Related to Figure 1.
Figure S5. Microbes stimulate specialized processes in progenitor-like cell subsets. Heatmaps of differentially expressed genes (GF vs. CV, p<0.05) involved in metabolism (A), transcription (B), immunity (C) and growth (D), in progenitor-like subsets 0-3 (from Figure 1D-E), color coded according to Log₂(FC). All non-zero value expression changes are significant (p<0.05) as determined with a non-parametric Wilcoxon rank sum test. Related to Figure 2.
Figure S6. Immune gene expression across conventional and germ-free cell populations. Heatmap showing relative expression of a representative set of microbial sensors, NF-kB pathway components, cytokines and chemokines in CV and GF cell types. Related to Figure 2.
Figure S7. Secretory cell regional specification and goblet cell characterization. (A) Heatmap showing relative expression of established regional marker genes in each secretory cell type. (B-C) Violin plots for agr2 (B) and muc5.3 (C) expression in goblet and goblet-like clusters. Related to Figure 3.