Science Advances
AAAS

Supplementary Materials for

Single-cell profiling of the amphioxus digestive tract reveals conservation of endocrine cells in chordates

Yichen Dai *et al.*

Corresponding author: Xuming Zhou, zhouxuming@ioz.ac.cn; Guang Li, guangli@xmu.edu.cn

Sci. Adv. **10**, eadq0702 (2024)
DOI: 10.1126/sciadv.adq0702

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

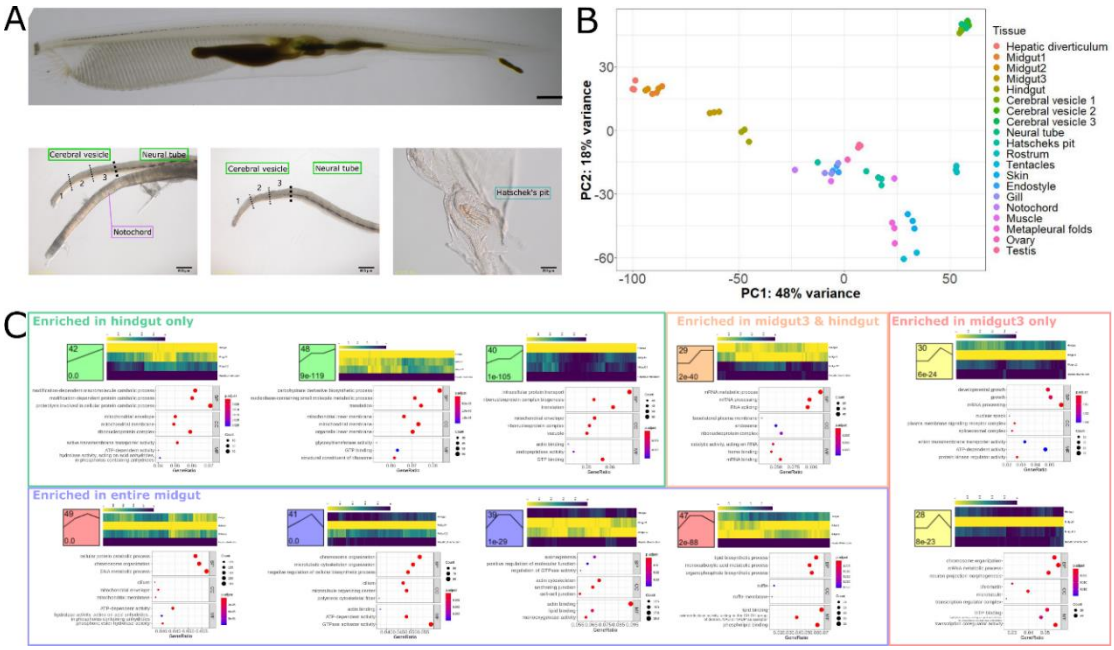


Figure S1 Adult amphioxus tissue types. (**A**) A representative juvenile individual. An adult was not chosen because they have gonad tissue that blocks observation of the digestive tract. Enlarged photos of dissected neural tube along with notochord tissue is shown, with approximate location of incisions made for cerebral vesicle samples 1, 2, and 3 are shown using dotted lines. A dissected Hatscheck's pit sample also shown. Scale bar = 1 mm for lateral view of a juvenile individual. Scale bar = 500 μm for photos of neural tube and notochord. Scale bar = 200 μm for photo of Hatscheck's pit. (**B**) Principal-component analysis (PCA) plot showing distribution of all adult tissue samples. Each type of tissue is depicted using a different color. (**C**) Clustering of gene expression profiles in the amphioxus gut. Amongst all predicted clusters, only the ten clusters with statistical significance are shown. These clusters are assigned into different panels indicated by colored lines, and the tissue that genes grouped into these clusters are enriched in are indicated at the top of the panel. Each cluster is indicated by a colored square, with the approximate trend of gene expression along the anterior-posterior gut indicated by a line drawn horizontally across the panel. The arbitrary number assigned to each cluster is indicated at the top of each square, and the *p*-value is shown at the bottom. The same color is assigned to clusters that show similar expression profiles. Gene expression and GO-term enrichment of all genes in each cluster is shown.

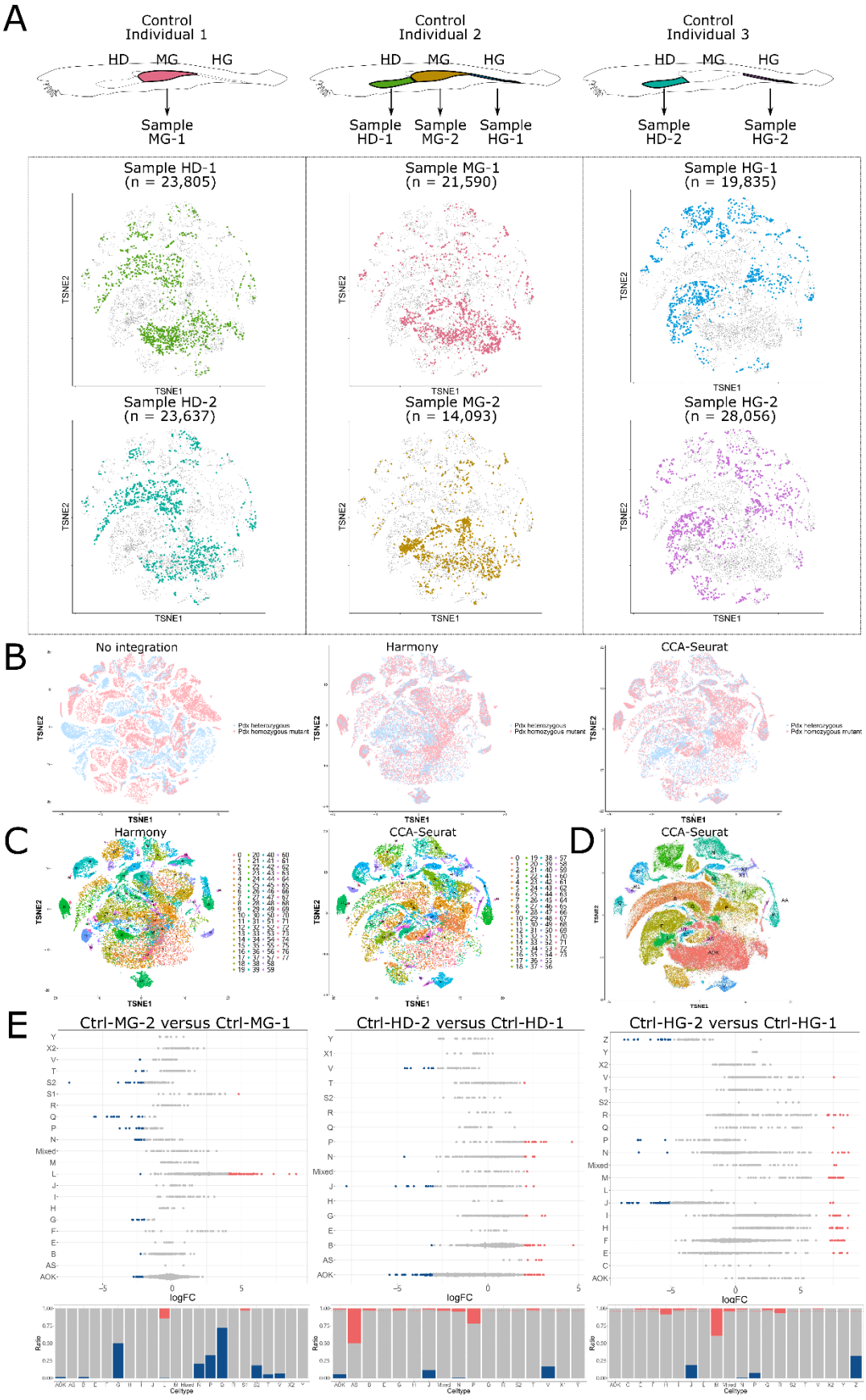

A
Control Individual 1
HD MG HG
Sample MG-1
Control Individual 2
HD MG HG
Sample HD-1
Sample MG-2
Sample HG-1
Control Individual 3
HD MG HG
Sample HD-2
Sample HG-2
Sample HD-1 (n = 23,805)
Sample MG-1 (n = 21,590)
Sample HG-1 (n = 19,835)
Sample HD-2 (n = 23,637)
Sample MG-2 (n = 14,093)
Sample HG-2 (n = 28,056)
TSNE1
TSNE2
B
No integration
Harmony
CCA-Seurat
Pdx heterozygous
Pdx homozygous mutant
C
Harmony
CCA-Seurat
D
CCA-Seurat
E
Ctrl-MG-2 versus Ctrl-MG-1
Ctrl-HD-2 versus Ctrl-HD-1
Ctrl-HG-2 versus Ctrl-HG-1
logFC
Ratio
Celltype

Figure S2 scRNA-seq of six control samples. (**A**) Six control samples were dissected from three adult amphioxus. TSNE plots showing the distribution of all six samples are shown, with each cluster randomly down-sampled to no more than 1,000 cells to save computational space. (**B**) TSNE plots showing distribution of control and *Pdx* mutant cells analyzed with three different integration methods, including no integration, Harmony (*45*), and CCA-Seurat (*61*). (**C**) Result of unsupervised cell clustering in the same dataset integrated using Harmony (*45*) and CCA-Seurat (*61*). (**D**) Distribution of all 26 manually merged cell clusters in the dataset integrated using CCA-Seurat (*61*). (**E**) Difference in cell abundance between control biological repeat samples calculated by Milo (*46*). The ratio of increased (red), decreased (blue), and not changed (grey) cell neighborhoods in each cluster are shown as a barplot. HD: hepatic diverticulum; MG: midgut; HG: hindgut.

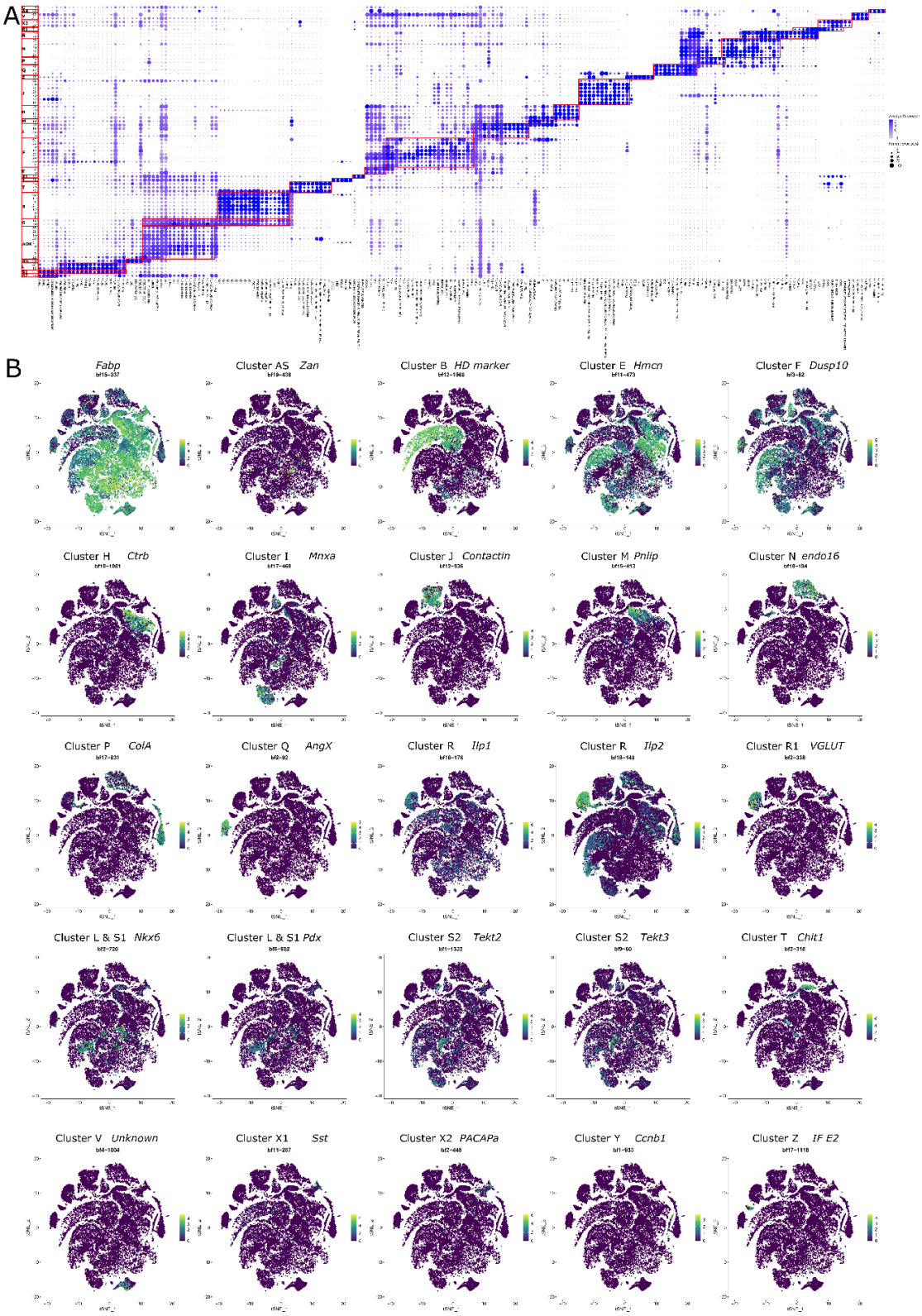


Figure S3 Marker genes in control cell clusters. (**A**) Dot plot showing expression of marker genes in different clusters, with marker gene name or ID shown on the x-axis and

cluster ID shown on the y-axis. (**B**) TSNE plot depicting the expression pattern of marker genes discussed in the main text. HD: hepatic diverticulum; MG: midgut; HG: hindgut.

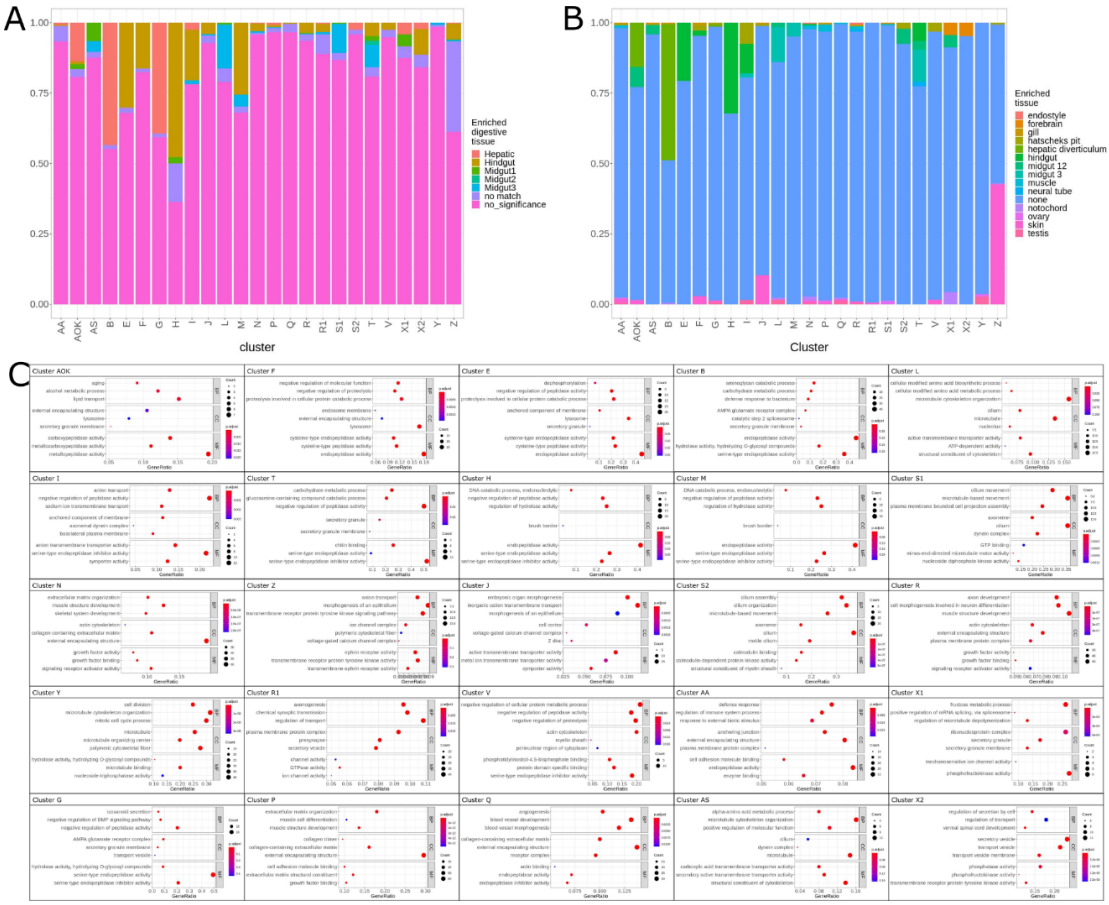


Figure S4 Marker genes for the 26 amphioxus gut cell clusters. (**A**) Percentage of marker genes for each cell cluster that are also marker genes for different sections of the digestive tract. Marker genes for single cell clusters are nominated by the Seurat algorithm while marker genes for bulk-tissue samples are defined as genes which are expressed at least five times higher in one tissue type compared to the average of all other tissue types. “No match” genes are found in single cell data but are recorded as “below 1 TPM” in all bulk digestive tissue samples. “No significance” genes are not marker genes for any type of digestive tissue. (**B**) Percentage of marker genes for each cell cluster that are also marker genes for different types of adult tissue. Gene definitions follow that in (**A**). (**C**) Top three GO terms for marker genes of all 25 cell clusters. The only exception is cluster C, which did not have enough marker genes for analysis.

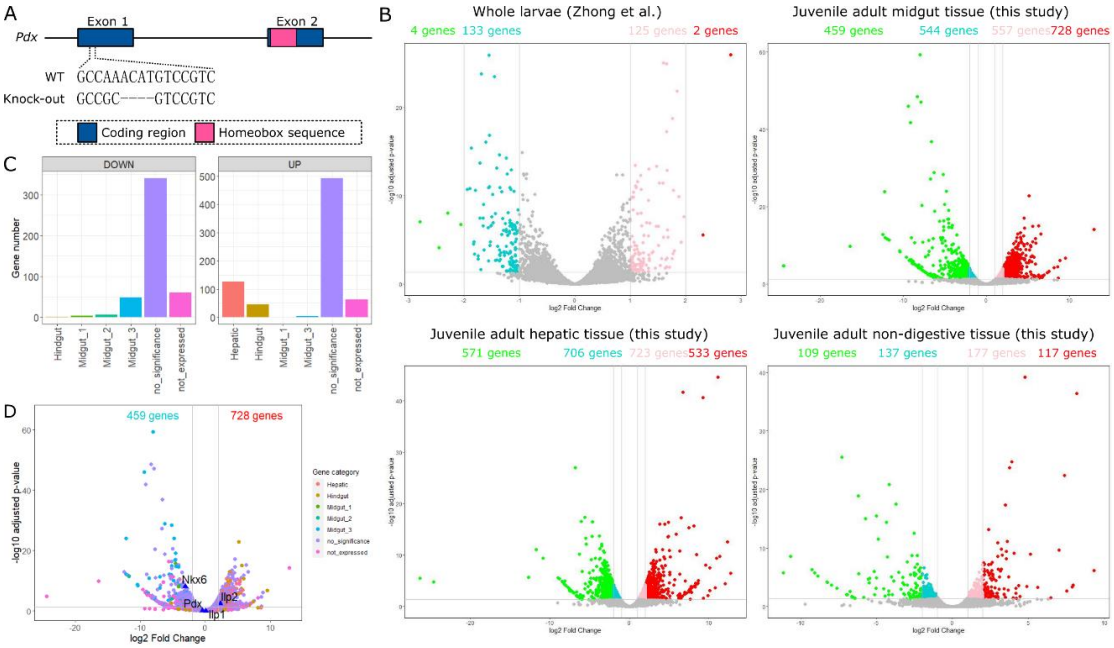
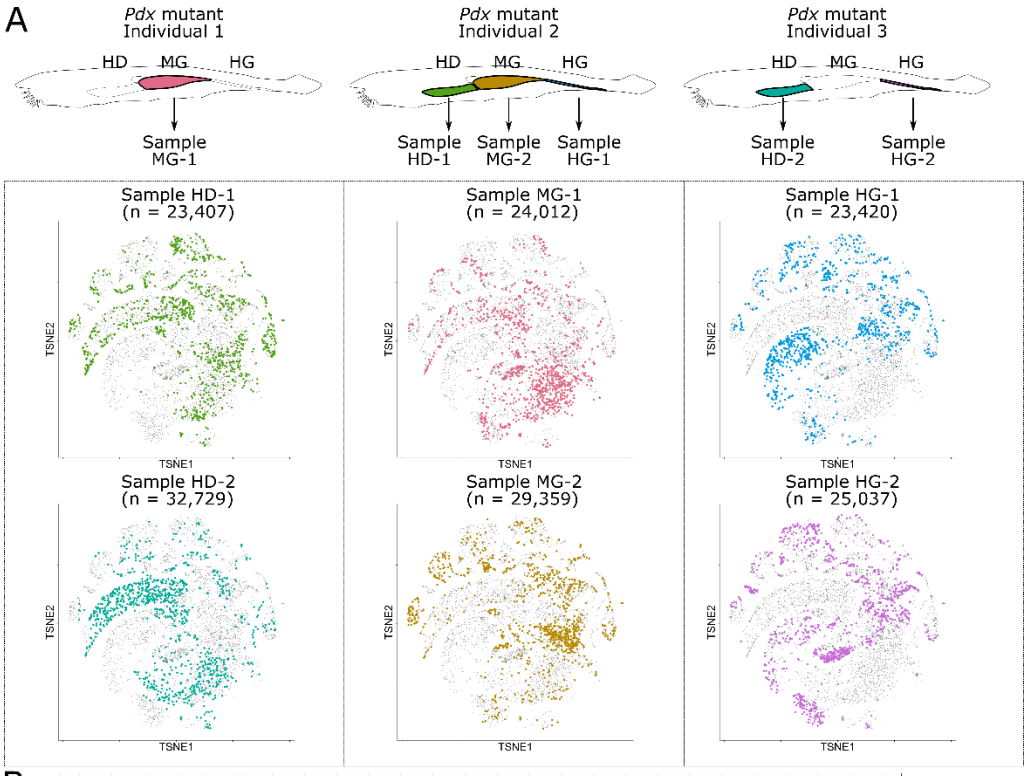
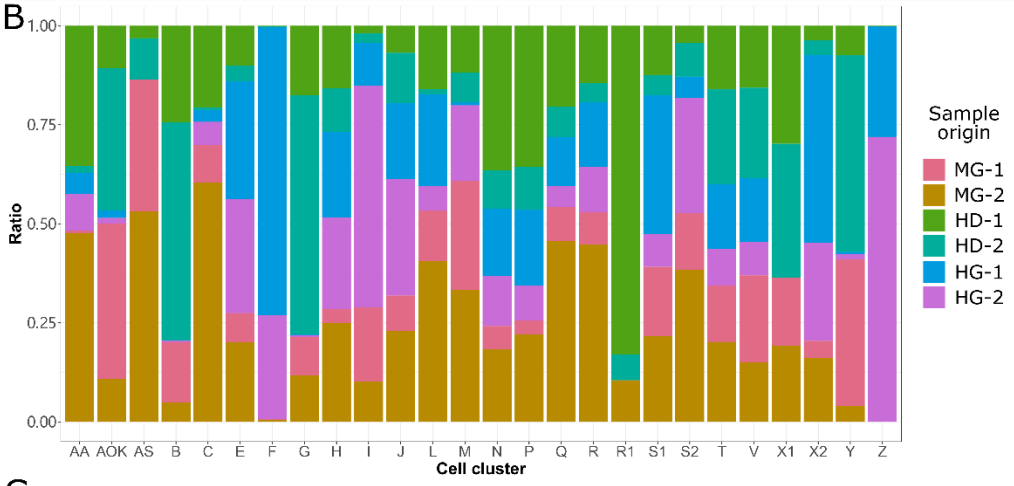


Figure S5 Bulk tissue transcriptome results for control and *Pdx* mutant juvenile tissue. (**A**) *Pdx* gene knock-out is caused by a 4 base-pair (bp) deletion in the first exon, upstream of the homeobox sequence. (**B**) Differential expression analysis of data from a previous study (Zhong et al.) (*43*) compared to data from this study. Each dot represents one gene, and genes with no significant difference in expression between control and mutant samples are shown in grey. Genes significantly down-regulated in *Pdx* mutant tissue with absolute log2(fold change) value between one to two are shown in cyan and genes with absolute log2(fold change) value above two are shown in bright green. Genes significantly up-regulated in *Pdx* mutant tissue with absolute log2(fold change) value between one to two are shown in pink while genes with absolute log2(fold change) value above two are shown in bright red. (**C**) Number of up or down-regulated genes in *Pdx* mutants that are also digestive tissue-specific marker genes. "No significance" genes are not enriched in a specific region of the digestive tract. "Not expressed" genes are not expressed above 1 TPM in any adult digestive tissue but are significantly up or down-regulated in juvenile *Pdx* mutants. (**D**) All differentially expressed genes in *Pdx* mutant midgut tissue compared to control, with genes colored using the same color scheme shown in (**C**). Four genes of interest, *Nkx6*, *Pdx*, *Ilp1*, and *Ilp2*, are highlighted.


A
Pdx mutant
Individual 1
HD
MG
HG
Sample
MG-1
Pdx mutant
Individual 2
HD
MG
HG
Sample
HD-1
Sample
MG-2
Sample
HG-1
Pdx mutant
Individual 3
HD
MG
HG
Sample
HD-2
Sample
HG-2
Sample HD-1
(n = 23,407)
Sample MG-1
(n = 24,012)
Sample HG-1
(n = 23,420)
Sample HD-2
(n = 32,729)
Sample MG-2
(n = 29,359)
Sample HG-2
(n = 25,037)
TSNE1
TSNE2


B
Ratio
1.00
0.75
0.50
0.25
0.00
AA AOK AS B C E F G H I J L M N P Q R R1 S1 S2 T V X1 X2 Y Z
Cell cluster
Sample
origin
MG-1
MG-2
HD-1
HD-2
HG-1
HG-2

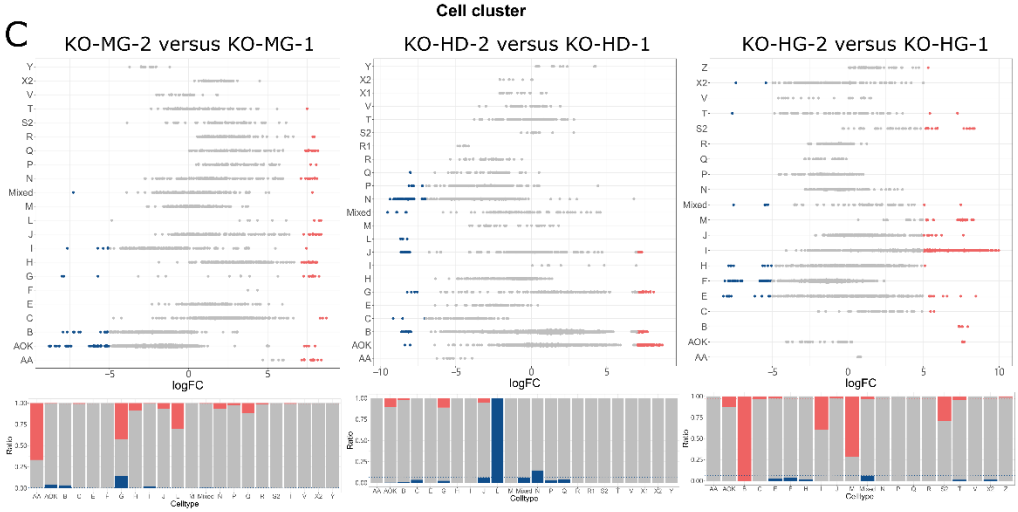

C
KO-MG-2 versus KO-MG-1
KO-HD-2 versus KO-HD-1
KO-HG-2 versus KO-HG-1
logFC
Celltype
Ratio

Figure S6 scRNA-seq of six *Pdx* mutant samples. (**A**) Six mutant samples were dissected from three adults. TSNE plots showing the distribution of all six samples are shown, with each cluster randomly down-sampled to no more than 1,000 cells to save computational space. (**B**) Summary of cell cluster identity and their sample of origin. (**C**) Difference in cell abundance between mutant biological repeat samples calculated by Milo (*46*). The ratio of increased (red), decreased (blue), and not changed (grey) cell neighborhoods in each cluster are shown as a barplot. HD: hepatic diverticulum; MG: midgut; HG: hindgut.

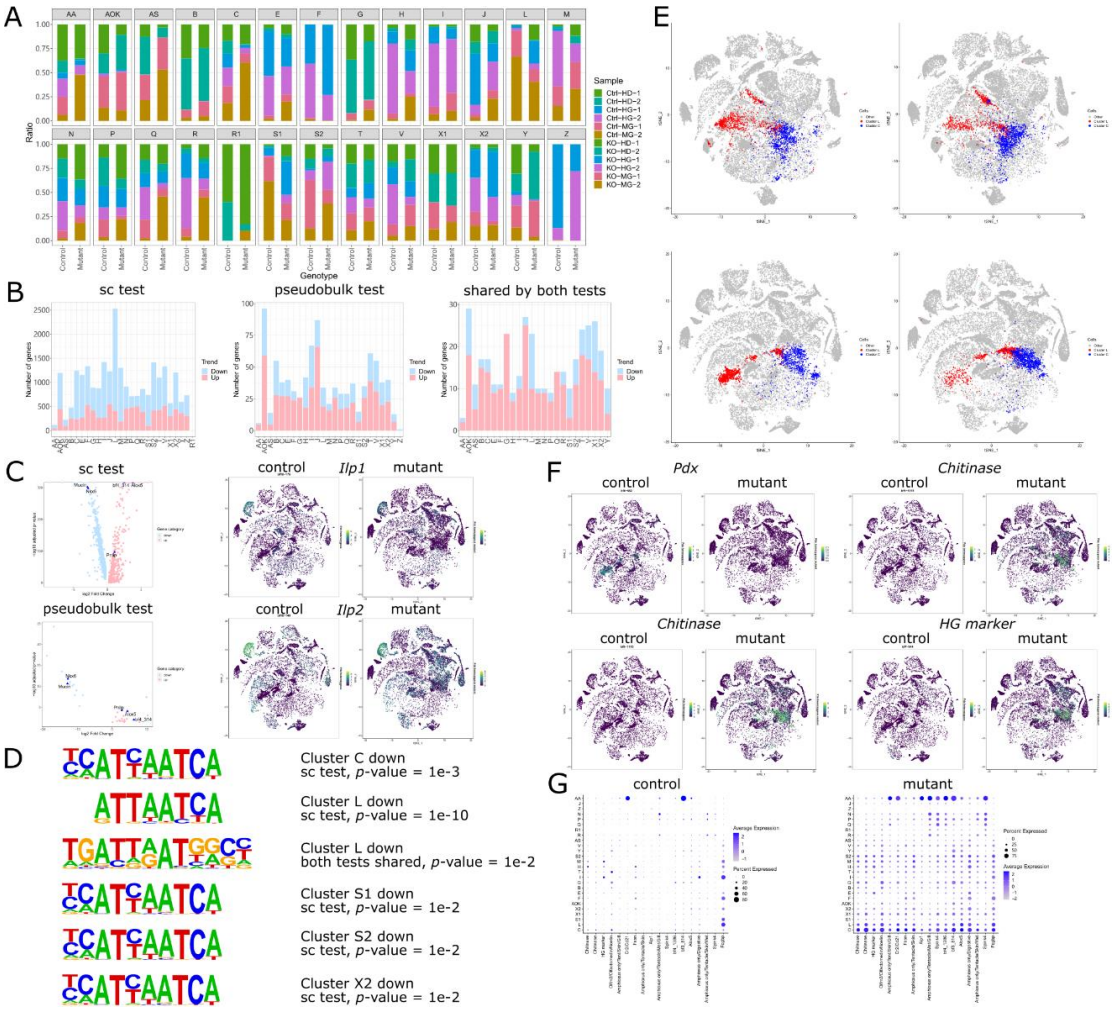


Figure S7 Comparison of control and *Pdx* mutant digestive scRNA-seq results. (**A**) Cell tissue origin ratio in control and *Pdx* mutant samples. Each cell cluster is shown separately, with control cells shown on the left and *Pdx* mutant cells shown on the right. (**B**) Total number of up-and down-regulated genes in each cell cluster in *Pdx* mutants compared to control calculated using two different methods. Down-regulated genes are shown in blue while up-regulated genes are in pink. (**C**) All differentially regulated genes in control and *Pdx* mutant cluster L cells calculated by two different methods. Genes down-regulated in *Pdx* mutant cluster L cells are labeled in blue while genes up-regulated are labeled in pink. Expression of two genes of interest, *Ilp1* and *Ilp2*, are shown in TSNE plots, with control cells on the left and mutant cells on the right. (**D**) HOMER motif prediction (*63*) results for gene down-regulated in clusters C, L, S1, S2, and X2. (**E**) Distribution of cluster C (blue) and cluster L (red) cells in the same dataset integrated using different methods, Harmony (*45*) and CCA-Seurat (*61*). (**F**) TSNE plot depicting

the expression pattern of marker genes discussed in the main text. (**G**) Dot plot of marker genes used to characterize mutant cell cluster C. HD: hepatic diverticulum; MG: midgut; HG: hindgut.

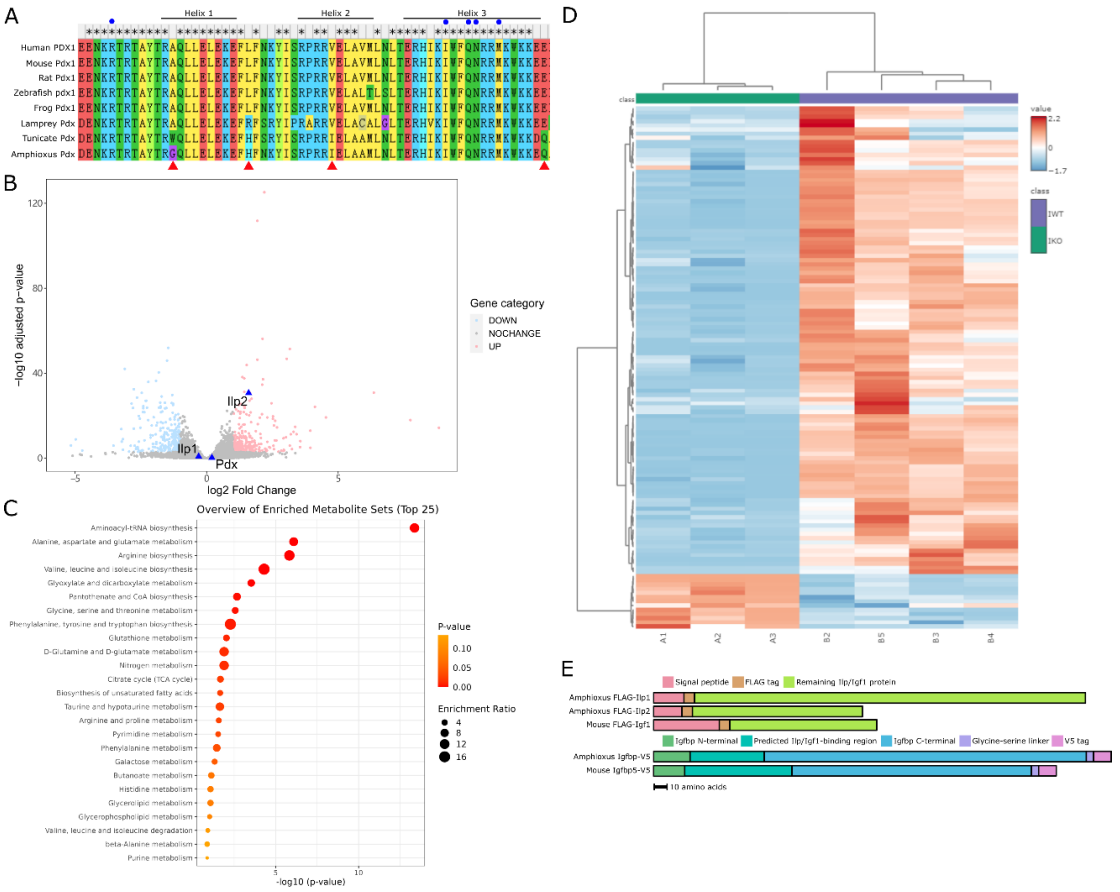


Figure S8 Amphioxus Ilp1 protein function. (**A**) Pdx protein homeodomain sequence across amphioxus, tunicate, and vertebrate species. Red triangles indicate amino acids in amphioxus Pdx that are different from the majority of vertebrate species. Blue dots indicate amino acids that are predicted to directly interact with DNA (*68*). (**B**) Differentially expressed genes in *Ilp1* mutant larvae compared to control. Genes up-regulated in *Ilp1* mutant larvae are indicated in pink, and down-regulated genes are shown in blue. *Pdx*, *Ilp1*, and *Ilp2* genes are highlighted using blue triangles. (**C**) Top 25 metabolic processes relevant to significantly altered metabolites in *Ilp1* mutant larvae compared to control. (**D**) Heatmap of metabolite concentration with sample identity shown on the x-axis. “IWT” refers to control samples while “IKO” refers to *Ilp1* mutant samples. Metabolite concentration is normalized across all samples. (**E**) Five expression constructs, with amphioxus Ilp1 and Ilp2 proteins and mouse Igf1 proteins labeled with a FLAG tag near the N-terminus, and amphioxus Igfbp and mouse Igfbp5 proteins labeled with a V5 tag at the C-terminus. Length of each protein is shown in proportion to predicted sequence length.

Table S1 Metadata of all wild type adult bulk tissue transcriptome samples. Gene IDs for all genes shown in Fig. 1 are also included.

Table S2 Metadata of all control digestive tract single-cell clusters, including read alignment information for all six control samples, cell cluster identity assignment in datasets integrated using different methods, batch information and cell distribution in different samples, gene IDs use for plotting Fig. 2D, marker gene information for all control cell clusters, and marker gene information for all zebrafish digestive cell clusters.

Table S3 Metadata of bulk tissue transcriptome results for control and *Pdx* mutant juvenile midgut, hepatic diverticulum, and non-digestive tissue.

Table S4 Metadata of all mutant digestive tract single-cell clusters, including read alignment information for all six mutant samples, batch information and cell distribution in different samples, a full list of all up and down-regulated genes in *Pdx* mutant cluster L cells compared to control using both calculation methods, and HOMER gene prediction results for down-regulated genes nominated by the sc test method in cluster L.

Table S5 Full list of genes differentially regulated in *Ilp1* mutant larvae, PCR primers used to clone full length gene coding sequences, and *in situ* primers used in this paper. All gene IDs used for gene phylogeny analysis are also included.

Movie S1 Zoom-in of live wild type juvenile amphioxus foregut and midgut. The movie records movement of food particles inside a live, well-fed individual kept in a clear petri dish filled with seawater containing algae.

Movie S2 Zoom-in of live *Pdx* mutant juvenile amphioxus foregut and midgut. The movie records a live, well-fed individual kept in a clear petri dish filled with seawater containing algae. There is clear lack of "food cord" formation compared to the individual shown in Movie S1.

Dataset S1 HOMER motif prediction (*63*) results for all differentially expressed genes nominated in all cell clusters by the sc test.

Dataset S2 All core promoter sequences used for analysis in Fig. 5C. For all genes apart from *B. floridae Ilp1* and *Ilp2*, 500 bp upstream of the predicted transcription start site was extracted from NCBI (please refer to Table S1 for all NCBI gene IDs). For *B. floridae Ilp1* and *Ilp2* promoter sequences, we extracted the 500 bp upstream sequence from an updated chromosome-level assembly (*119*).