PLOS Biology

METHODS AND RESOURCES

Identification of intestinal enteroendocrine cell subtypes and their associated hormones in zebrafish

Margaret Morash[1], Richard G. Kay[2], Erik J. Soderblom[3], Grace H. MacLean[1], Jia Wen[1], Peyton J. Moore[1], Colin R. Lickwar[1], Mujahid Ali Shah[4], Julia Ganz[4], Fiona M. Gribble[2], Frank Reimann[2], Rodger A. Liddle[5], John F. Rawls[1,5]*

1 Department of Molecular Genetics and Microbiology, Duke University School of Medicine, Durham, North Carolina, United States of America, **2** Institute of Metabolic Science, Addenbrooke's Hospital, University of Cambridge, Cambridge, United Kingdom, **3** Proteomics and Metabolomics Shared Resource, Duke University School of Medicine, Durham, North Carolina, United States of America, **4** Department of Integrative Biology, Michigan State University, East Lansing, Michigan, United States of America, **5** Division of Gastroenterology, Department of Medicine, Duke University School of Medicine, Durham, North Carolina, United States of America

* john.rawls@duke.edu

OPEN ACCESS

Citation: Morash M, Kay RG, Soderblom EJ, MacLean GH, Wen J, Moore PJ, et al. (2025) Identification of intestinal enteroendocrine cell subtypes and their associated hormones in zebrafish. PLoS Biol 23(12): e3003522. https://doi.org/10.1371/journal.pbio.3003522

Academic Editor: Mary C. Mullins, University of Pennsylvania School of Medicine, UNITED STATES OF AMERICA

Received: January 1, 2025

Accepted: November 9, 2025

Published: December 18, 2025

Data availability statement: The Underlying data for quantitative panels within the paper can be found in the S1 Data Supporting Information file. The mass spectrometry peptidomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository

Abstract

Enteroendocrine cells (EECs) are rare sensory cells in the intestinal epithelium that coordinate digestive physiology by secreting a diverse repertoire of peptide hormones. These hormones are the main effectors of EEC function, and their characterization requires direct observation by mass spectrometry due to the specialized protein cleavage and posttranslational modifications that yield their mature forms. Based on the distinct subset of hormones they predominantly secrete, EECs can be categorized into subtypes. How each EEC subtype is specified, however, remains poorly understood. Here, we describe EEC subtype differentiation and hormone production in the zebrafish. Using single-cell RNA sequencing data, we identified EEC progenitors and six EEC subtypes in zebrafish and revealed that their expression profiles are consistent across larval and adult stages. Mass spectrometry analysis of isolated zebrafish EECs identified highly processed peptides derived from 19 of 23 hormone-coding genes expressed by EECs, including a previously undescribed zebrafish *secretin* ortholog. We assembled reporters for zebrafish EEC subtypes to test the lineage relationships between EEC subtypes and the EEC progenitor population, which expresses *neurogenin 3 (neurog3)*. Despite its essential role in mammalian EEC differentiation, we found that selective cytotoxic ablation of *neurog3*+ cells in zebrafish only reduced a subset of EEC subtypes and loss of the *neurog3* gene had no impact on EEC numbers. Finally, we discovered that selective ablation of *ghrelin*+ EECs reduced a different subset of EEC subtypes, together suggesting that *neurog3*+ and *ghrelin*+ cells serve as distinct precursors for separate EEC subtypes. We anticipate these observations and resources will facilitate future studies in the

with the dataset identifier PXD058654. The single cell RNA sequencing dataset is derived from previously published datasets (PMID 3430599, 37948182). OurThe subsetted, integrated dataset shown in Fig 1 has been deposited to Zenodo and isare available at DOI https://doi.org/10.5281/zenodo.17342282.

Funding: This work was supported by the National Institutes of Health (R01-DK093399 to JFR, F30-DK135357 to MM, R21-NS123629 to JG), the Gordon and Betty Moore Foundation (44590A to JFR), the Simons Foundation Autism Research Initiative (SPC-001799 to JFR), the Pew Biomedical Scholars Program (00032413 to JFR), the Duke Institute for Brain Sciences (Pilot grant to JFR), the National Science Foundation (CAREER #2143267 to JG), and Duke University (start-up funds to JFR). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations : ANOVA, analysis of variance; DTA, diphtheria toxin subunit A; EECs, enteroendocrine cells; gRNAs, guide RNAs; INSL5, insulin-like 5; PBSTX, phosphate buffer saline-Triton X-100; scRNA-seq, single-cell RNA sequencing.

zebrafish to discern the developmental biology, physiology, and endocrinology of EEC subtypes.

Introduction

Enteroendocrine cells (EECs) are specialized intestinal epithelial cells that help coordinate digestive physiology in response to ingested food, resident microorganisms, and other stimuli. EECs are an ancient and conserved feature of the digestive tract in bilaterian animals [1–3], with evidence of cells with similar function in distantly related placozoa [4,5]. EECs respond to stimuli by releasing secretory vesicles containing at least one of over a dozen different peptide hormones they produce. These EEC peptide hormones govern diverse physiologic functions like nutrient digestion, gut motility, satiety, and insulin signaling [6,7]. Each EEC peptide hormone is produced through the transcription and translation of its preprohormone protein, which is then subjected to successive cleavage and posttranslational modification steps prior to secretion from the cell [8]. These highly specialized processing steps make it largely impossible to predict the final functional peptide's form from the primary amino acid sequence alone, necessitating direct observation of peptides through sensitive peptidomic studies to shed light on their biology.

As the functions of different EEC peptide hormones are highly varied and sometimes antagonistic, ensuring the release of the appropriate hormone is essential for mounting the appropriate physiologic response to a given stimulus. This is largely accomplished through the differentiation of EECs into distinct subtypes that each secrete a single or small subset of hormones [6,7]. EEC differentiation begins with secretory progenitor cells within the intestinal epithelium, which can give rise to any secretory lineage, such as enteroendocrine, goblet, tuft, and Paneth cells [9]. In mouse and humans, transient expression of the transcription factor *Neurogenin 3* (*Neurog3*) commits these secretory progenitors to an EEC fate, and *Neurog3* expression is therefore used as a marker of EEC progenitors [10–14]. Select transcription factors including *Neuronal differentiation 1* (*Neurod1)* are subsequently broadly expressed to aid in the differentiation of *Neurog3*+ EEC progenitors into mature EEC subtypes [15–18]. More recent studies in mammalian organoid models have proposed successive factors expressed only in a subset of EECs that drive EECs towards a single specific subtype fate [19,20]. This work has begun to shed light on how EEC subtype identity—and thus the cell's hormone profile and function—is determined, but further studies and in vivo testing are needed to elucidate this important process.

While EEC subtypes were historically believed to be terminally differentiated and therefore insulated from one another once committed, recent work in organoids indicated that some EEC subtypes can transition into other subtypes as they travel up the crypt-villus axis [21]. This new concept of EEC subtype plasticity raises interesting questions about the lineage relationships between EEC subtypes. While a new concept in the intestine, there are examples of plastic lineage relationships between endocrine subtypes in the stomach and pancreatic islet where ghrelin-producing

cells have been shown to give rise to other endocrine populations [22]. Multiple single-cell RNA sequencing (scRNA-seq) studies in mouse have hinted at a similar role for *Ghrelin* (*Ghrl*) expressing cells in the small intestine as they identified a *Ghrl*-expressing EEC progenitor population [23–25]. RNA velocity and partition-based graph abstraction of one of those datasets predicted a differentiation trajectory through the *Ghrl*+ progenitor cluster to many EEC subtypes [24], but a role for *Ghrl*+ cells in EEC differentiation has never been directly tested. Together, these studies suggest that establishment of EEC subtype identity may be more dynamic than previously thought.

The zebrafish is an ideal vertebrate model to study EEC subtype dynamics. We previously revealed extensive similarities between EEC morphology, sensitivity, and neural connectivity in zebrafish and mammals [26,27]. Additionally, many important aspects of EEC differentiation are conserved between zebrafish, mammals, and insects [28–33], although there are slight differences. Specifically, *neurog3*, the zebrafish ortholog of *Neurog3*, is expressed in some zebrafish EECs, but is thought to play a smaller role in EEC differentiation, although its exact function remains unclear [28,30]. Instead, *neurod1*, the zebrafish ortholog of *Neurod1*, has been shown to be necessary and sufficient for EEC differentiation [26,27,31]. Similar to mammals, *neurod1* is expressed across all EECs in zebrafish and is used as a pan-EEC reporter [26]. Critically, the optical transparency of zebrafish enables in vivo imaging of gut anatomy and physiology through early adulthood. Combined with the ease of generating genetic reporters, this provides a powerful system for investigating EEC subtype development and dynamics in their native context. However, zebrafish EEC subtypes and their processed hormone products have not been previously described.

In this study, we aimed to define EEC subtypes in zebrafish and establish an experimental system in which to study their development and function. To do this, we leveraged scRNA-seq data from larval and adult zebrafish to reveal seven distinct EEC populations, six of which we characterized in vivo across larval development. We also deployed peptidomic analysis to demonstrate that these EECs produce a wide array of highly processed and conserved peptide hormones. To better understand how these EEC subtypes develop, we probed their lineage relationships by genetically restricting certain cell states and assessing the impacts on the remaining subtype populations. In this way, we were able to reveal that both predicted contributors, such as *neurog3*+ cells, as well as untested factors, such as *ghrl*+ cells, play major roles in development of EEC subtypes in zebrafish.

Results

Zebrafish EEC transcriptional signatures identify EEC progenitors and six distinct EEC subtypes

To understand EEC subtype diversity and differentiation in zebrafish, we used established markers of secretory cells [9,28,29,34–47] to identify and subset these populations from published intestinal scRNA-seq datasets derived from larval [48] and adult [49] zebrafish (S1A and S1B Fig). Integration of these secretory cells from larval and adult datasets showed the expected structure of a hub of secretory progenitor cells with spokes of distinct secretory populations (S1C and S1D Fig). With this confirmation, we then subsetted and reintegrated just EECs and secretory progenitors into a new combined larval and adult dataset (Fig 1A and 1B). Within this dataset, we identified clusters 0–6 as EECs based on *neurod1* expression (Fig 1C) and clusters 7–10 as secretory progenitors (S2A Fig). We found that each EEC cluster contained both larval and adult cells with no apparent age dependency for any EEC subtype (Figs 1D, S2B, and S2C), suggesting that EEC subtypes and their underlying gene expression programs are surprisingly consistent between larval and mature adult stages in zebrafish.

To identify the shared features of larval and adult EECs, we assessed the expression of EEC markers and hormone coding genes in the combined dataset (Fig 1E). Hormone coding genes (marked with an asterisk in Fig 1E) are often used to separate EECs into subtypes. In accord, we found that expression of each EEC hormone coding gene is largely partitioned into a single cluster with remarkable consistency across larval and adult stages. All *neurod1*+ clusters expressed EEC hormone coding genes except for cluster 0, which instead expressed *neurog3*. This pattern is consistent with features of EEC progenitors in mammals, which are marked by *Neurog3* expression and have not yet begun to express

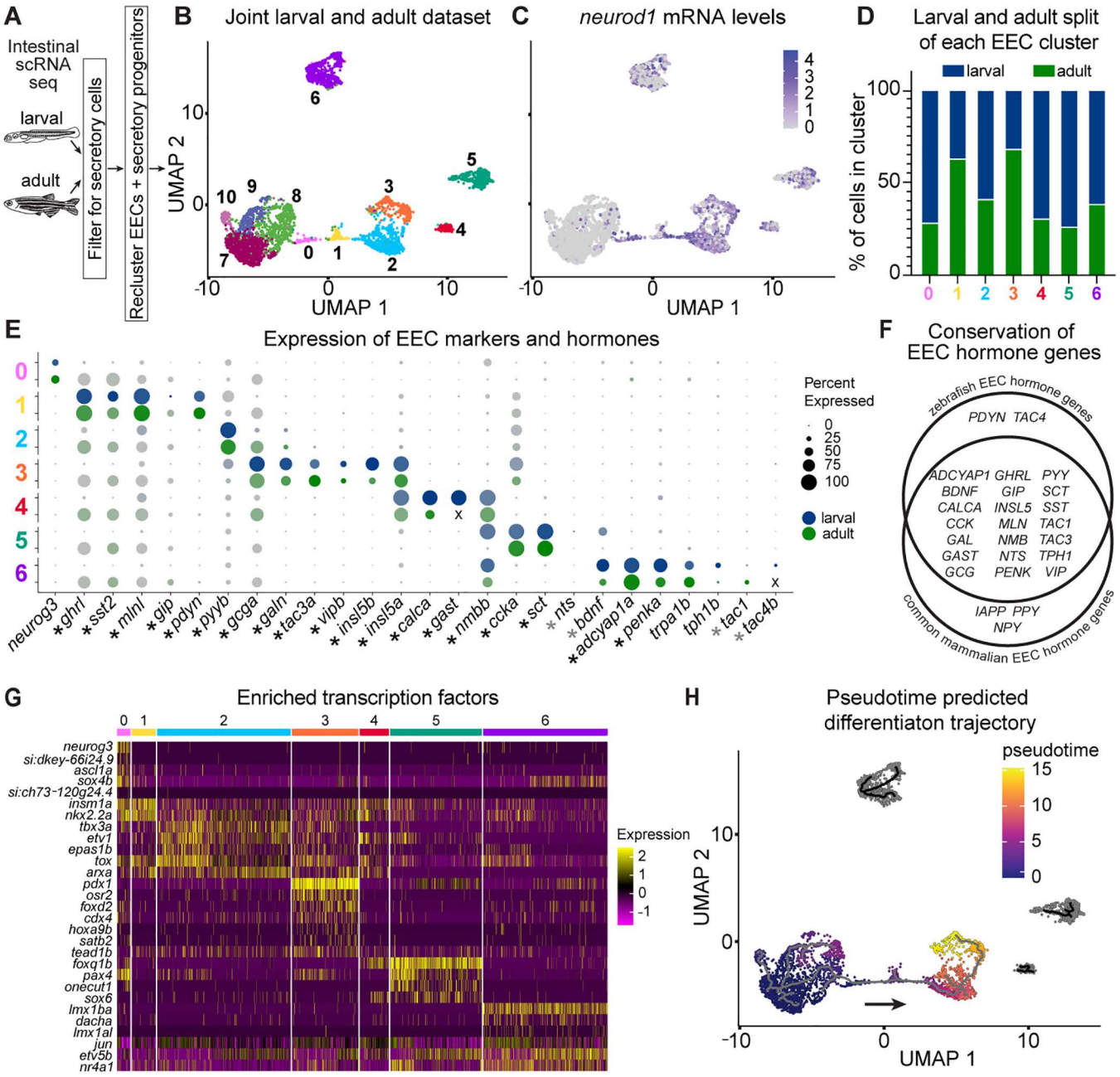


Fig 1. scRNA-seq of larval and adult EECs identifies 7 distinct clusters. (A) Schematic overview of identifying secretory cells from published larval [48] and adult [49] datasets. **(B)** UMAP of secretory progenitors and EECs from larval and adult zebrafish identifies 11 clusters. **(C)** Normalized expression of the zebrafish pan-EEC transcription factor and reporter *neurod1* labels clusters 0–6. **(D)** The percentage of cells derived from either the larval (blue) or adult (green) datasets in each of the 7 EEC clusters. **(E)** Scaled expression dotplot of relevant marker genes of the 7 EEC clusters. Dots represent gene expression in the specified cluster. Dot radius corresponds to the percent of cells expressing the gene and dot transparency corresponds to expression intensity. Expression is reported for both the larval (blue) and adult (green) cells in the cluster, and genes not annotated in a given dataset are noted with an X. EEC hormone coding genes are marked with an asterisk next to their gene name on the *x* axis. Black asterisks indicate genes we identified peptides from and gray asterisks mark those with no detected peptides (Figs 2B and S4). Note, annotations of *gast* (see scRNA-seq methods), *tac4b* (see scRNA-seq methods, S15 Fig), and *sct* (see Results and Figs 2 and S6) were not previously identified and were defined in this paper. **(F)** Venn diagram of EEC hormone coding gene homologs found in zebrafish and mammals. **(G)** Heatmap of expression of transcription factors significantly enriched in each EEC cluster in both larval and adult cells. **(H)** UMAP colored by pseudotime as determined by Monocle 3 when cluster 7 was provided as the origin. Brighter coloring indicates "older" cell age, and the gray line is the predicted differentiation trajectory. Underlying data can be found in S1 Data.

https://doi.org/10.1371/journal.pbio.3003522.g001

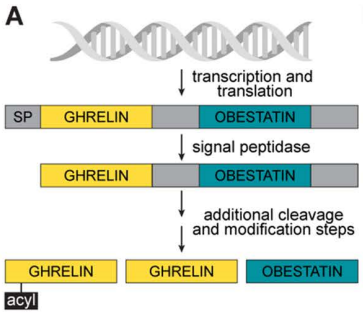
hormone coding genes [10–14,50]. As such, we operationally defined cluster 0 as a putative EEC progenitor population and clusters 1–6 as mature EEC subtypes. We observed substantial overlap between the EEC hormone coding genes expressed in our zebrafish dataset and those reported in published mammalian datasets (Fig 1F) [19,23,24,51–55], highlighting the critical role of these signaling molecules during vertebrate evolution.

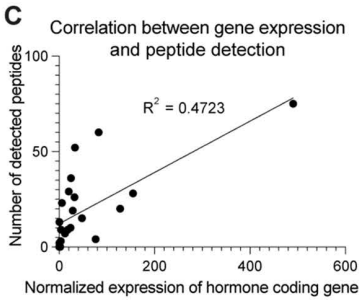
To test if the EEC subtype populations we identified are consistent across studies, we examined the EECs in another published larval zebrafish scRNA-seq dataset (S3A Fig) [56]. After selecting and re-clustering the EECs in that additional dataset, we found that the EEC hormone genes we highlighted in Fig 1E partitioned into clusters remarkably similarly to what we observed in our dataset (S3B Fig). The one exception was that *ghrl* and *pyyb* were expressed in two separate clusters in our combined dataset, but one in the additional dataset. However, the expression of *ghrl* and *pyyb* was non-overlapping in the additional dataset with *ghrl* primarily expressed at the boundary between the *pyyb*+ cluster and a cluster enriched for secretory progenitor markers (S3C Fig). This indicates that the zebrafish EEC subtypes described here are relatively consistent across independent studies.

To explore the broader transcriptional programs driving EEC subtype identity, we identified enriched genes specific to each cluster (S1 Table, S2D Fig). Using existing ontologies [57], we identified transcription factors that were significantly enriched in both larval and adult cells in each cluster (Fig 1G). We also performed pseudotime analysis [58], a computational method for ordering cells along a hypothetical timeline based on transcriptomic similarity that enables the study of cell differentiation dynamics. Using the secretory progenitor cluster 7 as the start point, this analysis predicted a differentiation trajectory from secretory progenitor clusters to EEC subtype clusters (Fig 1H). These complementary approaches highlighted certain patterns in EEC populations. For example, we note that putative EEC progenitor cluster 0 is enriched in *ascl1a* and *sox4b* in addition to *neurog3*. While mammalian *Neurog3* is the most canonical EEC progenitor marker, studies have also shown that *Ascl1*/*ascl1a* and *Sox4*/*sox4b* homologs label endocrine progenitor populations in mammals [19,24,41,59–61] and zebrafish [28,31,62,63]. EEC subtype cluster 1 is enriched for *insm1a* and *nkx2.2a* whose mammalian homologs *Insm1* and *Nkx2.2* are known to play a role in the differentiation of *Neurog3*+ EEC progenitors into mature EEC subtypes [64–67]. Together with the pseudotime data, these findings support a classification of cluster 0 as EEC progenitors and subtype cluster 1 as potentially an early EEC state. Interestingly, subtype cluster 1 also expresses the EEC hormone coding genes *ghrelin* (*ghrl*), *somatostatin* (*sst2*), *motilin-like* (*mlnl*), and *gastric inhibitory peptide* (*gip*), which are canonically thought to mark a mature EEC population [50,68]. EEC subtype clusters 2 and 3 appear to have a close relationship given their adjacent positions in pseudotime and the high expression of subtype cluster 2 enriched transcription factors in subtype cluster 3. EEC subtype cluster 3 is uniquely enriched for the known endocrine transcription factor *pdx1* [69–71] as well as *osr2*, which has been shown in mammals and zebrafish to be more highly expressed in the mid and distal small intestine [23,72], indicating a potential regional distribution of this subtype. Finally, subtype cluster 6, which expresses the rate-limiting enzyme in serotonin synthesis *tph1b* (Fig 1E), is enriched for transcription factors *lmx1ba* and *lmx1al* whose orthologs in mammals have been shown to be important for the differentiation of serotonergic neurons and EECs, respectively [66,73,74]. Serotonergic EECs, termed enterochromaffin cells, separate from other EECs in mammalian scRNA-seq studies [19,24,25,75]. Likewise, in our dataset, we find that subtype cluster 6 most clearly separates from the remaining EEC populations (Fig 1B). Altogether, this identifies cluster 6 as zebrafish enterochromaffin cells and suggests the divergence between enterochromaffin cells and other EECs is an ancestral trait to fishes and mammals. This *in silico* analysis of scRNA-seq data highlights a limited set of transcription factors that might play a role in promoting differentiation from secretory progenitors into discrete EEC subtypes.

Zebrafish EECs produce fully processed and modified peptide hormones in vivo

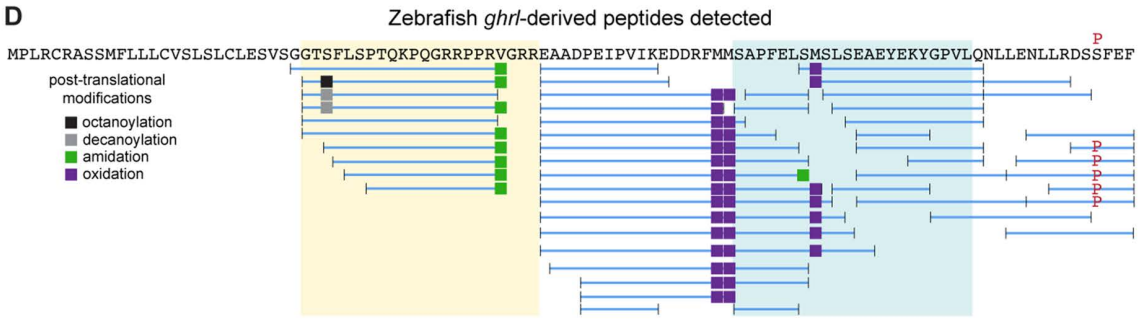
EEC peptide hormones are derived from the initial translated preprohormone through several steps of specialized cleavage and posttranslational modification (Fig 2A). As a point of nomenclature, in this text we refer to the peptide-coding gene and transcript in italics (*GHRELIN* for human, *Ghrl* for mouse, and *ghrl* for zebrafish and other model organisms), the

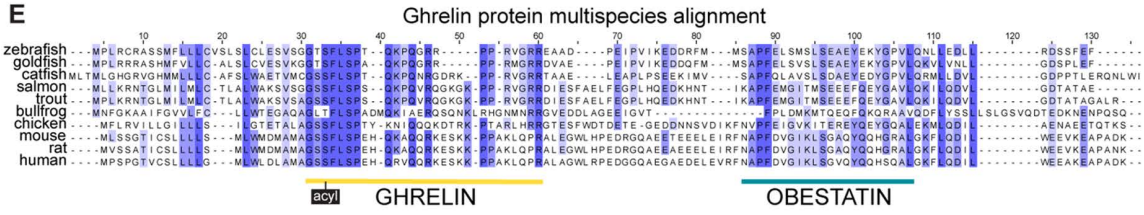




B

Expression rank	EEC hormone coding gene	Count of peptides detected
1	*peptide YY b* (*pyyb*)	75
2	*motilin-like* (*mlnl*)	28
3	*insulin-like 5a* (*insl5a*)	20
4	*proenkephalin a* (*penka*)	60
5	*cholecystokinin a* (*ccka*)	4
6	*secretin* (*sct*)	15
7	*ghrelin* (*ghrl*)	52
8	*adenylate cyclase activating poleypeptide 1a* (*adcyap1a*)	26
9	*neuromedin bb* (*nmbb*)	19
10	*glucagon a* (*gcga*)	36
11	*galanin* (*galn*)	10
12	*vasoactive intestinal peptide b* (*vipb*)	30
13	*gastrin* (*gast*)	9
14	*tachykinin precursor 3a* (*tac3a*)	7
15	*somatostatin 2* (*sst2*)	23
16	*calcitonin-related polypeptide a* (*calca*)	9
17	*gastric inhibitory peptide* (*gip*)	3
18	*tachykinin precursor 4b* (*tac4b*)	0
19	*brain-derived neurotrophic factor* (*bdnf*)	0
20	*insulin-like 5b* (*insl5b*)	13
21	*prodynorphin* (*pdyn*)	2
22	*tachykinin precursor 1* (*tac1*)	0
23	*neurotensin* (*nts*)	0





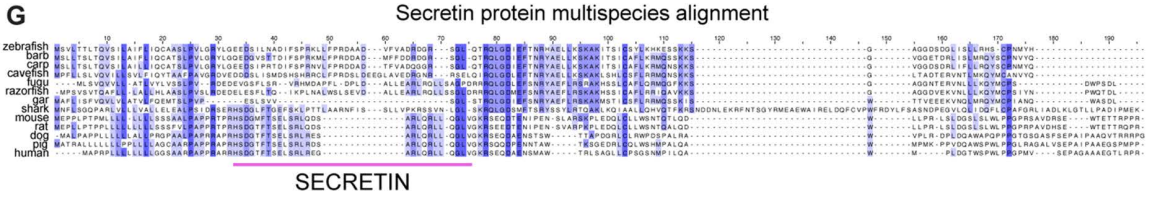


Fig 2. Peptidomic analysis of zebrafish EECs reveals production of conserved and novel peptides. (A) Schematic of peptide hormone processing using the human EEC hormone coding gene *GHRELIN* as an example. **(B)** Number of peptides identified from each zebrafish hormone coding

gene ranked by normalized mRNA expression level as assessed by scRNA-seq (S1 Table). Note, annotations of *gast* (see scRNA-seq methods), *tac4b* (see scRNA-seq methods, S15 Fig), and *sct* (see Results and Figs 2 and S6) were not previously identified and were defined in this paper. **(C)** Correlation between the normalized mRNA expression of zebrafish EEC hormone coding genes and the number of peptides detected. **(D)** Representation of *ghrl*-derived peptides detected in zebrafish EECs. The primary amino acid sequence is shown at the top in black with observed missense variants indicated in red above. Blue horizontal lines below the amino acid sequence represent the individual peptides detected in our study with small black vertical lines denoting the stop and start of each peptide. Abutting peptides share a single vertical line to represent the stop of the preceding peptide and the start of the subsequent one. Different colored squares represent various posttranslational modifications detected. Yellow shading marks the region aligning to the functional ghrelin peptide in humans and teal shading marks the region aligning to the functional obestatin peptide in humans. **(E)** Alignment of the primary amino acid sequence of the Ghrelin protein in zebrafish (Uniprot F1QKX9), goldfish (Uniprot Q8AUU1), catfish (Ensembl ENSIPUT00015057246.1), salmon (Uniprot B3IYK1), trout (Uniprot A0A060WHA4), bullfrog (Uniprot Q90W22), chicken (Uniprot Q8AV73), mouse (Uniprot Q9EQX0), rat (Uniprot Q9QYH7), human (Uniprot Q9UBU3). Amino acids are color-coded based on their percent identity match across all the reported species with darker coloring indicating a more conserved residue. Human functional peptide annotations are labeled below, including the conserved acylation at the serine 3 residue. **(F)** Representation of peptides detected from the zebrafish *sct* gene. Annotation is as described in D, with pink shading marking the region aligning to the functional secretin peptide in humans and gray shading marking the region highly conserved across fishes. **(G)** Alignment of the primary amino acid sequence of the Secretin protein in zebrafish (Ensembl ENSDARG00000099233), Sumatra barb (NCBI LOC122330700), grass carp (NCBI LOC127507777), Pachon cavefish (Ensembl ENSAMXG00005000672), fugu (Ensembl enstrug00000026212), razorfish (NCBI CAJ1056004), gar (NCBI LOC107075768), elephant shark (NCBI LOC121850333), mouse (Uniprot Q08535), rat (Uniprot P11384), dog (Uniprot A0A813PAI0), pig (Uniprot P63298), human (Uniprot P09683). Coloring and annotation as described in **E**. Underlying data can be found in S1 Data.

https://doi.org/10.1371/journal.pbio.3003522.g002

pro-peptide protein with a leading capitalization (Ghrelin), and the secreted peptide in lowercase (ghrelin). To understand what processed peptides are produced by zebrafish EECs, we performed mass spectrometry on sorted EECs from larval and adult samples. Due to technical limitations, adult samples contained roughly twice as many cells as larval samples, and peptides derived from adult samples are thus overrepresented in our peptidomic dataset (see Methods). In S2 Table, we report each peptide and the quantitation (area under the peak value) from each of the 3 larval and 3 adult biological replicate samples. However, we did not perform comparisons between values from larval versus adult samples as the lower cell content in larval samples meant that low or absent signal in larval samples could be due to the limit of detection of the instrument and not reflective of the biology of the system. Additionally, the EEC peptide data presented in Figs 2 and S4 only reports the detection of unique individual peptides and should not be interpreted as reflective of their natural relative abundance in EEC secretory granules.

Of the 23 hormone-coding genes transcriptionally expressed by zebrafish EECs (Fig 1E), we identified peptide fragments aligning to 19 of them (Fig 2B). Genes with no detected peptides included *neurotensin* (*nts*), *tachykinin precursor 1* (*tac1*), and *tachykinin 4b* (*tac4b*), which were very lowly expressed at the transcript level, and *brain-derived neurotrophic factor* (*bdnf*), whose peptide product is too large for optimal detection with our methodology. Of note, while expression of *tph1b* is used as a proxy for serotonin because it catalyzes the rate-limiting step in its formation, serotonin itself is derived from a single amino acid and is not suited to detection with our methodology. The EEC peptides we identified include products of hormone coding genes specifically expressed by EEC subtypes 1–6, suggesting each of those subtypes produce peptide hormones. We found a modest correlation between the expression level of a gene in scRNA-seq data (S1 Table) and the number of unique peptides detected that align to that gene (Fig 2C). We also identified peptides from other common components of the hormone-containing secretory vesicles released by EECs. Namely, we detected peptides from *proprotein convertase subtilisin/kexin type 1* and *2* (*pcsk1* and *pcsk2*), known for their role in specialized cleavage of peptide hormones [76], and *secretogranin 2a, 2b, 3,* and *5*, which help form the secretory vesicles and can be processed into hormones themselves (S4Q–S4X Fig) [77–80]. Together, these observations indicate that our EEC peptidomic approach effectively isolated and defined peptide contents of zebrafish EEC secretory vesicles.

Using this dataset, we generated an atlas of zebrafish EEC peptides including their sequence, posttranslational modifications, and alignment to homologous proteins in humans and other vertebrates. This is provided as a resource in the supplemental materials supporting this article (S4 Fig), and we highlight Ghrelin-derived peptides as an example (Fig 2D).

In mammals, two main peptide hormone products from the *GHRELIN* gene have been described, ghrelin and obestatin [81,82]. Mammalian ghrelin is a 28-amino acid peptide that is often fatty acylated at the serine 3 residue [83]. While both acylated and deacylated ghrelin peptides are produced, fatty acylation improves binding to ghrelin's receptor and thus is critical to its function [81,83]. As such, the serine 3 residue is highly conserved across species with evidence of acylation at this site in each vertebrate examined (Fig 2E) [81,84,85]. Indeed, studies in goldfish, a close relative of zebrafish, identified a functional 17-amino acid ghrelin peptide that is acylated at the serine 3 residue [84,86], and intraperitoneal injection of acylated (but not desacylated) ghrelin stimulated increased feeding in goldfish and grass carp [86,87]. Mammalian obestatin is a 22-amino acid peptide that was identified more recently by bioinformatic prediction [82], and its function is still debated [82,88]. We detected 52 Ghrelin-derived peptides in zebrafish, including 10 that align with the portion of the protein shown to encode a functional ghrelin peptide in humans and goldfish (Fig 2D and 2E) [86,89,90]. As fatty acylation is key to ghrelin's function and not included in the posttranslational modification parameters of our initial search, we performed a second search for just ghrelin peptides and included octanoyl and decanoyl modifications in our parameters. Peptides with both octanoyl- and decanoyl-modified serine 3 residues were detected, demonstrating remarkable conservation of the processing and posttranslational modification of this peptide hormone (Figs 2D, S5A, and S5B). While functional studies remain to be performed, the alignment and modification data suggest these peptides could be endogenous zebrafish ghrelin. In contrast, while we detected peptides aligning to portions of the reported obestatin region, we did not detect any peptide spanning that entire region. Our data suggests obestatin may not be produced in fishes, although co-injection of a synthetic predicted obestatin-like peptide has been shown to inhibit the appetite-stimulating effects of injected ghrelin in grass carp [87].

While our data on Ghrelin-derived peptides confirmed our predictions based on studies in closely related organisms, we were also able to leverage our peptidomic dataset for novel discoveries. For example, we detected several peptides aligned to the largely unannotated gene *si:zfos-2372e4.1* (Fig 2F). Like most of the known peptide-coding genes, *si:zfos-2372e4.1* is highly expressed in a specific EEC cluster, namely cluster 5, which is also enriched for *ccka* expression (Fig 1E). Notably, in other zebrafish scRNA-seq datasets, expression of *ccka* is most highly correlated with expression of *si:zfos-2372e4.1* ($r = 0.74$) [91], demonstrating a consistency of this pattern across laboratories. However, *si:zfos-2372e4.1* had no gene or protein annotations. On our analysis, predictive software Signal P [92,93] assigned a high likelihood (0.81) that Si:zfos-2372e4.1 contains a signal peptide sequence at amino acid 1–23, indicating the protein is secreted. The prediction of a signal peptide is supported by the lack of detected peptides aligning to the N-terminus of the prepropeptide (Fig 2F).

To identify the potential functions of this gene, we performed a TBLASTN search using the full prepropeptide amino acid sequence, and the only similar sequences were in other fish species—largely in unannotated or predicted genes. Intriguingly, however, there was an approximately 20 amino acid sequence in Si:zfos-2372e4.1 that consistently aligned with those proteins in other species, suggesting that portion of the protein may be conserved and therefore perhaps functional. We looked specifically for any homology in gar, a holostean fish whose lineage diverged from teleost before the teleost-specific third round (3R) whole genome duplication (S6A Fig) [94]. Our search identified a gene in gar that, while also unannotated, allowed us to bridge to additional vertebrate species [95], including an elephant shark gene labeled as prosecretin with homology to *SECRETIN* (*SCT*) genes in humans and other tetrapods (S6B Fig) [96].

To further investigate the possibility that zebrafish *si:zfos-2372e4.1* may be a *SCT* ortholog, we performed a syntenic analysis (S6B Fig). We identified genes in the ~250,000 base pair region upstream and downstream of the human *SCT* gene. Over half of those genes we identified had orthologs in zebrafish as annotated by Ensembl [97], and we traced the chromosomal arrangement of those genes in the Chondrichthyes elephant shark (*Callorhinchhus milii*), the Osteichthyes gar (*Lepisosteus oculatus*), and the Osteichthyes zebrafish (*Danio rerio*). While some rearrangements did occur, the immediate neighborhood remained remarkably consistent across all species (S6B Fig, S3 Table). This syntenic conservation across bony fish, cartilaginous fish, and tetrapods supports our identification of *si:zfos-2372e4.1* as *sct* and suggests

that this organization was present in the last common ancestor of all jawed vertebrates (S6A Fig) [98]. Given the evidence of homology from our TBLASTN and synteny assessments, we conclude that *si:zfos-2372e4.1* is the zebrafish *sct* ortholog.

We aligned both known Secretin prepropeptides and unannotated proteins that were returned from our TBLASTN search (Fig 2G). We found the homology within fishes and between fishes and mammals in the region annotated as human secretin was relatively limited. However, we did detect several peptides aligning to this region in zebrafish (shaded in pink in Fig 2F). Moreover, we detected several additional peptides that aligned to a region that showed strong homology in fishes (shaded in gray in Fig 2F). All of these peptides had an N-terminal pyroglutamylated glutamine residue, supporting that these are genuine processed peptides, and not just degradation products. Functional studies are needed to test which, if any, of the zebrafish Secretin-derived peptides have secretin-like or potentially novel function. While we highlight Ghrelin- and Secretin-derived peptides as examples of conservation and potential novelty, we anticipate that the many additional zebrafish EEC peptides included in S4 Fig will serve as a valuable resource for future studies.

Zebrafish EEC subtypes show distinct expression patterns in vivo

To understand how the seven zebrafish EEC populations we identified develop and function in vivo, we assembled reporters for the EEC progenitor population and five of the EEC subtypes (Fig 3A): *neurog3* and its transgene reporter *neurog3:QF2* for cluster 0, *ghrl* and its transgene reporter *ghrl:QF2* for cluster 1, *peptide YYb* (*pyyb*) and anti-PYY antibody for cluster 2, *glucagon a* (*gcga*) and its transgene reporter *gcga:GFP* for cluster 3, *cholecystokinin a* (*ccka*) and anti-CCK antibody for cluster 5, and *transient receptor potential cation channel, subfamily A, member 1b* (*trpa1b*) and its transgene reporter *trpa1b:GFP* for cluster 6. Transgenic reporters *gcga:GFP* and *trpa1b:GFP* and anti-PYY and anti-CCK antibodies were already available [99–102], but we developed new promoter-driven QF2 transgenic lines for *neurog3* (using a 2114 bp *neurog3* promoter) [103] and *ghrl* (using a 647 bp *ghrl* promoter) (S7 Fig and Methods). These reagents leverage the QF2-QUAS binary expression system, enabling us to combine our QF2 reporters with different QUAS tools [104,105]. As expected, our *ghrl* reporter co-labels with the *neurod1* reporter and labels sparse EECs in larval and adult fish (S7E–S7I Fig). Our *ghrl* reporter also co-labels with cells that stain with anti-Ghrelin antibody and cells that label with *ghrl* mRNA using hybridization chain reaction (S7J–S7M Fig). Our *neurog3* reporter likewise co-labels with the *neurod1* reporter and labels a similar number of cells in the intestine as previously reported lines generated with the same promoter (S7S and S7T Fig) [106].

To understand the dynamics of these EEC subtype reporters during intestinal development, we crossed each of them to the pan-EEC reporter *neurod1:RFP* [26] and imaged them daily from 3 to 6 days post fertilization (dpf) (S8 and S9 Figs). Cell counts from each day and representative images at 6 dpf are shown (Fig 3B–3S). For clarity, we refer below to cells labeled by a transgenic reporter or antibody marker as "+" for that respective gene (e.g., *ghrl*+). Subtype reporters showed a variety of prevalence and regional distribution patterns (Fig 3T–3W), which may contribute to their sub-specialized functions. *ghrl*+, PYY+, and CCK+ cells—labeling subtype clusters 1, 2, and 5, respectively—were all concentrated proximally within the intestine. Mammalian orthologs *Ghrl*/*GHRL* and *Cck*/*CCK* are also proximally skewed in the small intestine, but *Pyy*/*PYY* is more prevalent in the distal small intestine and colon in mammals [1,107]. *gcga*+ cells from subtype cluster 3 peaked in the midgut, consistent with the enrichment of the transcription factor *osr2* in this cluster [72]. In accordance with mammalian literature, *trpa1b*+ cells labeling subtype cluster 6 (enterochromaffin cells) were the most prevalent subtype and were relatively evenly distributed along the length of the intestine [108]. To further explore the regional distribution of EEC subtypes, we used Seurat module scores to look at the expression of regionally enriched genes [109]. Consistent with our imaging results, midgut genes were enriched in cluster 3 and anterior gut markers were enriched across many subtype clusters (S10 Fig). Overall, the subtype reporters appeared as early as 3 dpf but became more widely expressed at 4 dpf, in alignment with *neurod1* reporter activity, growth of the intestine, and mRNA in situ hybridization data [29,30]. After 4 dpf, EEC subtype numbers increased with age except for *gcga*+ and *ghrl*+ cells, which showed a trend towards decreasing after their initial increase in cell number at 4 dpf.

A

Cluster	Reporter	Enriched hormone coding genes	Enriched transcription factors	Enriched receptors
0	*neurog3:QF2*	NA	*neurog3, ascl1a, sox4b, si:dkey-66i24.9, si:ch73-120g24.4*	NA
1	*ghrl:QF2*	***ghrl, mlnl, pdyn, (sst2), (gip)***	*insm1a, nkx2.2a*	*adora2aa*
2	PYY ab	***pyyb***	*tbx3a, etv1, arxa, epas1b, tox*	*cxcr4b, alcamb, ptgen3*
3	*gcga:GFP*	***gcga, galn, tac3a, vipb, insl5b, insl5a***	*pdx1, osr2, foxd2, cdx4, hoxa9b, sat2b, tead1b*	*rxfp3.3a1, gfra1a, plxna4, amn, anxa2b*
4	NA	***insl5a, calca, gast, nmbb***	NA	*cntfr, npr3, gpr17, nrxn3b*
5	CCK ab	***ccka, sct***, *(nts)*	*foxq1b, onecut1, sox6, pax4*	*adgrg4b, trhra, sorcs2, oprm1, tlr18*
6	*trpa1b:GFP*	***nmbb, adcyap1a, penka***, *bdnf, tph1b, tac1, (tac4b)*	*lmx1ba, dacha, lmx1al, jun, etv5b, nr4a1*	*trpa1b, adgrg4a, galr1a, cd63, gpr78a, trpv4, tlr5b*

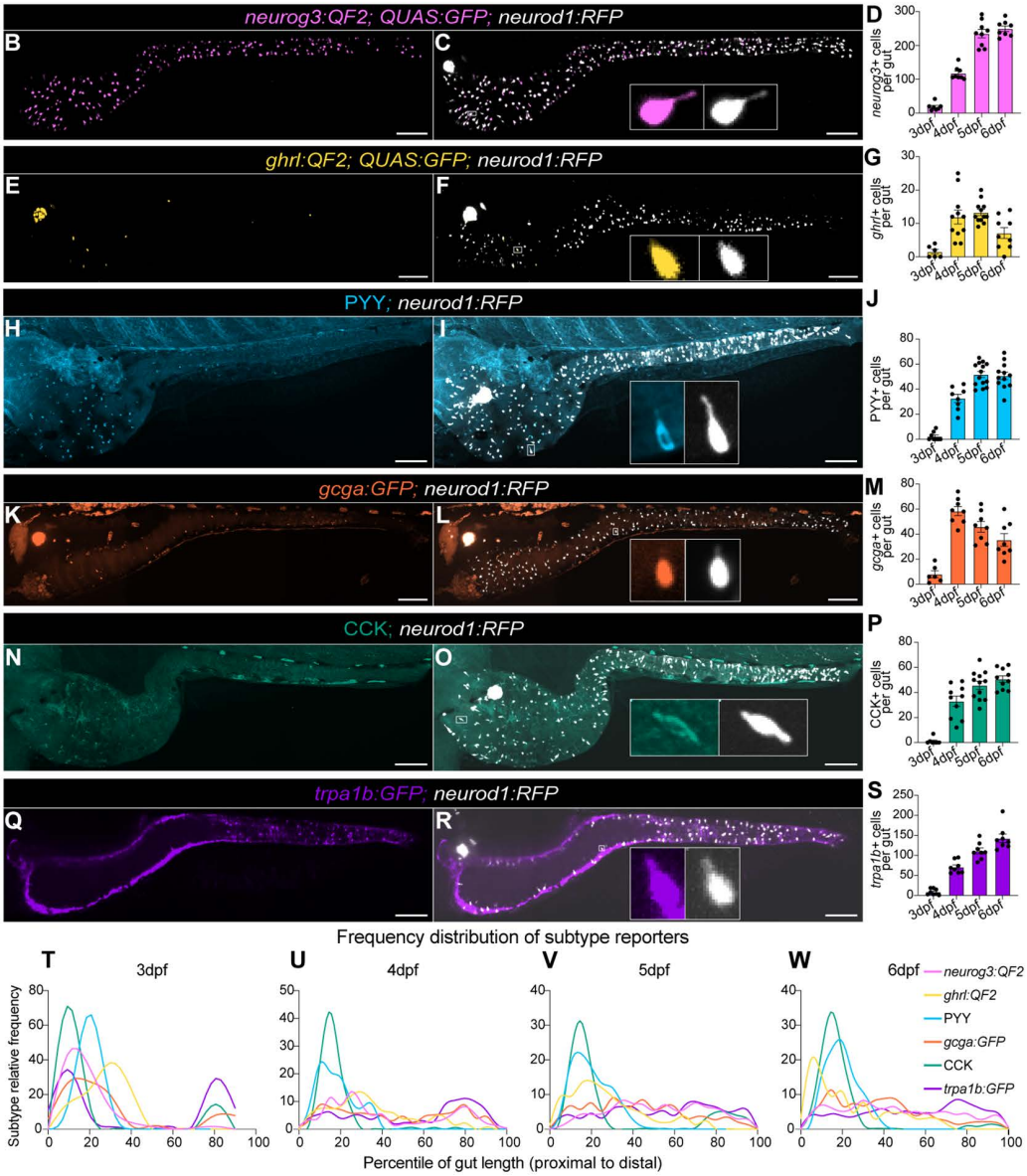


Fig 3. In vivo imaging of EEC subtypes shows distinct spatial and temporal patterns during larval development. (A) Summary of the hormones, transcription factors, and receptors significantly enriched in each EEC cluster and the reporter used to characterize that cluster. Bolding indicates genes from which we detected peptides. Parentheses indicate hormone-coding genes of interest expressed in that cluster that did not meet significance

threshold for enrichment. Note, annotations of *gast* (see scRNA-seq methods), *tac4b* (see scRNA-seq methods, S15 Fig), and *sct* (see Results and Figs 2 and S6) were not previously identified and were defined in this paper. **(B–S)** Representative images from 6 days post fertilization (dpf) fish show subtype distribution and overlap with pan-EEC reporter *neurod1*. Individual cells are shown in inset. Quantitation of EEC subtype cell numbers from 3 to 6 dpf fish are also shown with each dot representing an individual fish. All scale bars are 100 μm. **(T)** 3 dpf, **(U)** 4 dpf, **(V)** 5 dpf, and **(W)** 6 dpf subtype localization data shown as frequency distributions of cells' locations along the length of the gut divided into percentiles from proximal to distal. Distributions are smoothed for visualization. Underlying data can be found in S1 Data.

https://doi.org/10.1371/journal.pbio.3003522.g003

Importantly, while we apply these antibodies and reporters as useful markers of distinct EEC populations, EEC subtype definitions are operational and can vary across modalities and resolutions. For example, while expression of *sst2* and *ghrl* is highest in cluster 1 (S1 Table), cluster 1 contains cells that express only *sst2*, cells that express only *ghrl*, cells that express both *ghrl* and *sst2*, and cells that express neither *ghrl* nor *sst2* (S7N–S7R Fig). With greater clustering resolution and sequencing depth, these populations could be considered distinct subtypes. In this study, we operationally define cluster 1 as an EEC subtype and use the *ghrl* reporter as a label for this subtype.

Ablation of *neurog3*+ cells leads to reduction in EECs limited to select subtypes

We predicted that cells in scRNA-seq cluster 0 were EEC progenitors based on their enriched expression of *neurog3* and other EEC progenitor transcription factors and their early position on the pseudotime-predicted differentiation trajectory (Fig 1G and 1H). As cluster 0 made up a relatively small proportion of total EECs by scRNA-seq (Fig 1B), we were surprised to see our *neurog3* reporter label so many EECs in vivo (Fig 3B–3D). It is unclear if this is due to promiscuity of the *neurog3:QF2* transgene or perdurance of protein expressed from the transgene reporter. The mouse ortholog, *Neurog3*, is known to have transient expression in EEC progenitors [110], thus the perdurance of both QF2 and GFP may contribute to a long-lasting labeling of cells that transiently expressed *neurog3* in their developmental past. Thus, the unexpectedly high number of *neurog3*+ cells at 6 dpf may indicate that a larger portion of EECs transiently express *neurog3* at earlier developmental stages.

To test this hypothesis, we used a combination of reporters to genetically induce Diphtheria toxin subunit A (DTA) expression, and thus cell death, in any *neurog3*+ cell in the intestine (Fig 4A). Specifically, we leveraged a transgene reporter that uses *gata5* regulatory sequences to drive intestinal epithelial expression of a loxP-flanked mCherry upstream of DTA (*gata5:RSD*) [27]. When we combined this reporter with our *neurog3:QF2* and *QUAS:Cre* alleles, the mCherry is recombined out in *neurog3*+ cells, leading to expression of DTA and cell death, which we operationally call "*neurog3*+ cell ablation." This strategy affects cluster 0, which has enriched *neurog3* expression, and effectively prevents the emergence of any further mature subtypes that ever pass through a *neurog3*+ state.

We evaluated the total numbers and regional distribution of *neurod1*+ cells (Fig 4B) in this *neurog3*+ cell ablation model and found that roughly 50% of all EECs are lost with the effect being more pronounced in the proximal gut (Fig 4C). A previous study showed that expression of the *gata5:RSD* transgene is relatively higher in the proximal region of the larval intestine and that combination with *neurod1:Cre* (*neurod1*+ cell ablation) led to ablation of EECs in the proximal and mid intestine while retaining some EECs in the most distal region [27]. To test if this regional activity of the *gata5:RSD* transgene contributed to our observed patterns of regional EEC loss following *neurog3*+ cell ablation, we performed a series of control experiments. First, we quantified the number of *neurog3*+ cells in the *neurog3*-ablated fish and found that while almost no *neurog3*+ cells remained in the first three quarters of the intestine (qrt1–3), levels of *neurog3*+ cells were comparable to unablated control fish in the fourth quarter of the gut (qrt4; S11A and S11B Fig). We then likewise evaluated the effects of *neurod1* ablation on *neurod1*+ cells and observed a similar near complete loss of *neurod1*+ cells in qrt1–3 and preservation of *neurod1*+ cells in qrt4, consistent with previous studies [27]. In comparing the *neurod1*+ cell numbers in control, *neurog3*-ablated, and *neurod1*-ablated fish, there were significantly fewer *neurod1*+ cells in the proximal three-quarters of the gut in the *neurod1*-ablated fish as compared to the *neurog3*-ablated fish (S11C Fig). Similarly,

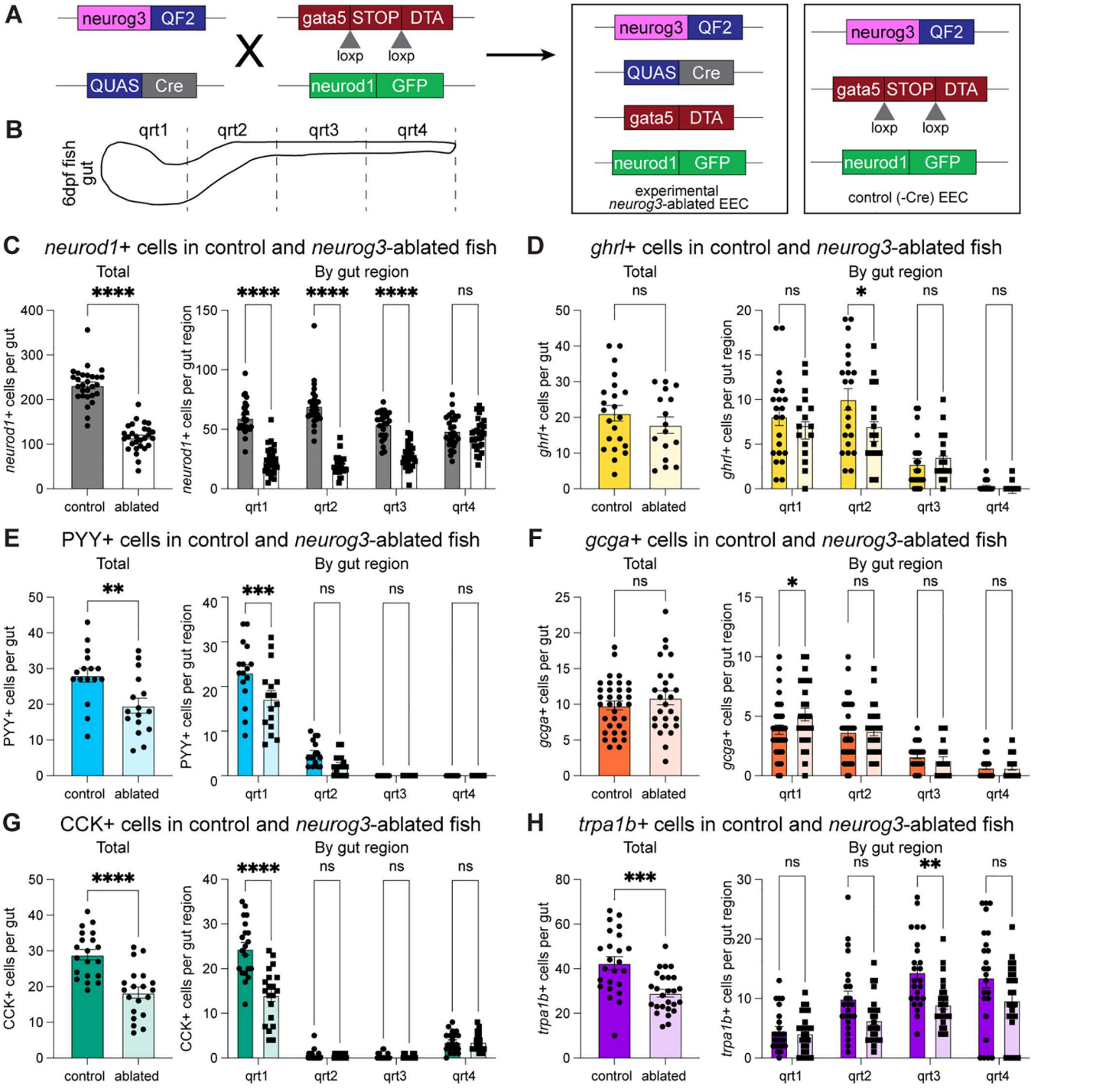


Fig 4. Selective ablation of *neurog3*+ cells in the intestine reduced total EECs and several EEC subtypes. (A) Schematic of *neurog3*+ cell ablation. **(B)** Schematic of regional scoring system. Total and regional counts of **(C)** *neurod1*+ cells, **(D)** *ghrl*+ cells, **(E)** PYY+ cells, **(F)** *gcga*+ cells, **(G)** CCK+ cells, and **(H)** *trpa1b*+ cells in control and *neurog3*+ ablated fish are shown. Each dot represents a 6-day post-fertilization fish. Statistical significance was calculated by unpaired *t* test for total cell numbers and by two-way ANOVA for regional analysis. Significance annotations are as follows: ns ($p > 0.05$), * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$), **** ($p < 0.0001$). Underlying data can be found in S1 Data.

https://doi.org/10.1371/journal.pbio.3003522.g004

trpa1b+ cells, while preserved in qrt4 in all animals, were significantly more reduced in *neurod1*-ablated than *neurog3*-ablated fish (S11D Fig). Together, these experiments show that combining a Cre driver with *gata5:RSD* efficiently kills cells in the first three quarters of the larval gut. They furthermore demonstrate that *neurog3*+ ablation significantly reduces EEC numbers, but not to the same extent as *neurod1* ablation, supporting the conclusion that, in contrast to mammalian systems [10], a *neurog3*+ cell state is not required for all EEC development in zebrafish. These results also demonstrate that ablation using the *gata5:RSD* transgene is inefficient in the final quarter of the gut, likely due to lower *gata5* expression in that region. To account for that lack of ablation in qrt4, we calculated total EEC numbers in our ablation experiments both including qrt4 (Figs 4C–4H and 5B–5F) and excluding qrt4 (S11E and S11F Fig) and observed the effect of ablation was the same regardless of whether qrt4 cells were included or not.

We next combined our cell ablation strategy with EEC subtype reporters to ask if certain EEC subtypes were particularly affected by *neurog3*+ cell ablation. Because we could not combine the *neurog3:QF2* and *ghrl:QF2* reporters, we generated a *neurog3:Cre* transgenic line to test the effects of *neurog3*+ cell ablation on *ghrl*+ cells. We confirmed that *neurog3*+ cell ablation induced with the *neurog3:Cre* reporter similarly reduced *neurod1*+ cell numbers as compared to the *neurog3:QF2; QUAS:Cre* reporter (S12 Fig). When evaluating EEC subtype populations in the context of *neurog3*+ cell ablation, we found that PYY+, CCK+, and *trpa1b*+ cells were all reduced by roughly 30% while *gcga*+ and *ghrl*+ cells were unaffected (Fig 4D–4H). Consistent with our reporter characterization (Fig 3), reductions in EEC subtypes were observed in the region of the gut where they were most prevalent—proximally for PYY+ and CCK+ cells and distally for *trpa1b*+ cells. This suggests that a portion of PYY+, CCK+, and *trpa1b*+ subtypes pass through a *neurog3*+ cell state during their differentiation.

To test if the effects of *neurog3*+ ablation were related to the loss of the *neurog3* gene product, we quantified the number of EECs in *neurog3* mutants [29] and control animals and observed no difference (S13D Fig). This suggests that the *neurog3* gene is not required for zebrafish EEC differentiation, which is consistent with what has been reported in zebrafish pancreatic islet differentiation [29], but contrasts with the known essential role of *Neurog3* in mammalian EEC development [10]. In combination with the results of the *neurog3*+ ablation experiments, we conclude that many zebrafish EECs pass through a *neurog3*+ cell state, but do not require *neurog3* function for their overt development.

Ablation of *ghrl*+ cells leads to reduction in EECs in subtypes distinct from *neurog3*+ ablation

The persistence of *ghrl*+ cells despite *neurog3*+ cell ablation is reminiscent of mammalian data that shows *Ghrl* expression and *Ghrl*+ cell numbers were maintained or increased upon deletion of several transcription factors important for EEC differentiation, such as *Nkx2.2a* and *Pax4* [19,65,66,111–114]. In addition, our pseudotime analysis predicted a differentiation trajectory moving through *ghrl*+ cells, strikingly similar to several mammalian scRNA-seq datasets that suggested a *Ghrl*+ EEC progenitor population [23–25,115]. Older lineage tracing data also supports a role for *Ghrl*+ cells as progenitors of other endocrine populations in the stomach and pancreatic islet [22]. However, the hypothesis that *Ghrl*+ cells in the intestine give rise to other EEC subtype populations remained untested.

Using a similar genetic approach as above, we used *gata5:RSD* to selectively kill any *ghrl*+ cell in the intestinal epithelium, which we operationally define as "*ghrl*+ cell ablation" (Fig 5A). Despite *ghrl*+ cell numbers peaking at approximately 15 cells per animal during larval development (Fig 3G), we found *ghrl*+ cell ablation resulted in a loss of roughly 40 EECs, indicating other EEC subtypes might be affected (Fig 5B). Indeed, we found a 30% reduction in PYY+ and *gcga*+ EECs in *ghrl*-ablated fish while CCK+ and *trpa1b*+ cell numbers remained unaffected (Fig 5C–5F). Notably, this is distinct from the *neurog3*+ ablation result where PYY+, CCK+, and *trpa1b*+ cells were reduced and *ghrl*+ and *gcga*+ cell numbers were unaffected (Fig 4). These data suggest a *neurog3*-independent role for *ghrl*+ cells in the differentiation of PYY+ and *gcga*+ EEC subtypes. This could be explained by PYY+ and *gcga*+ cells passing through a *ghrl*+ cell state or by the loss of a secreted factor from *ghrl*+ cells that promotes differentiation of these subtypes.

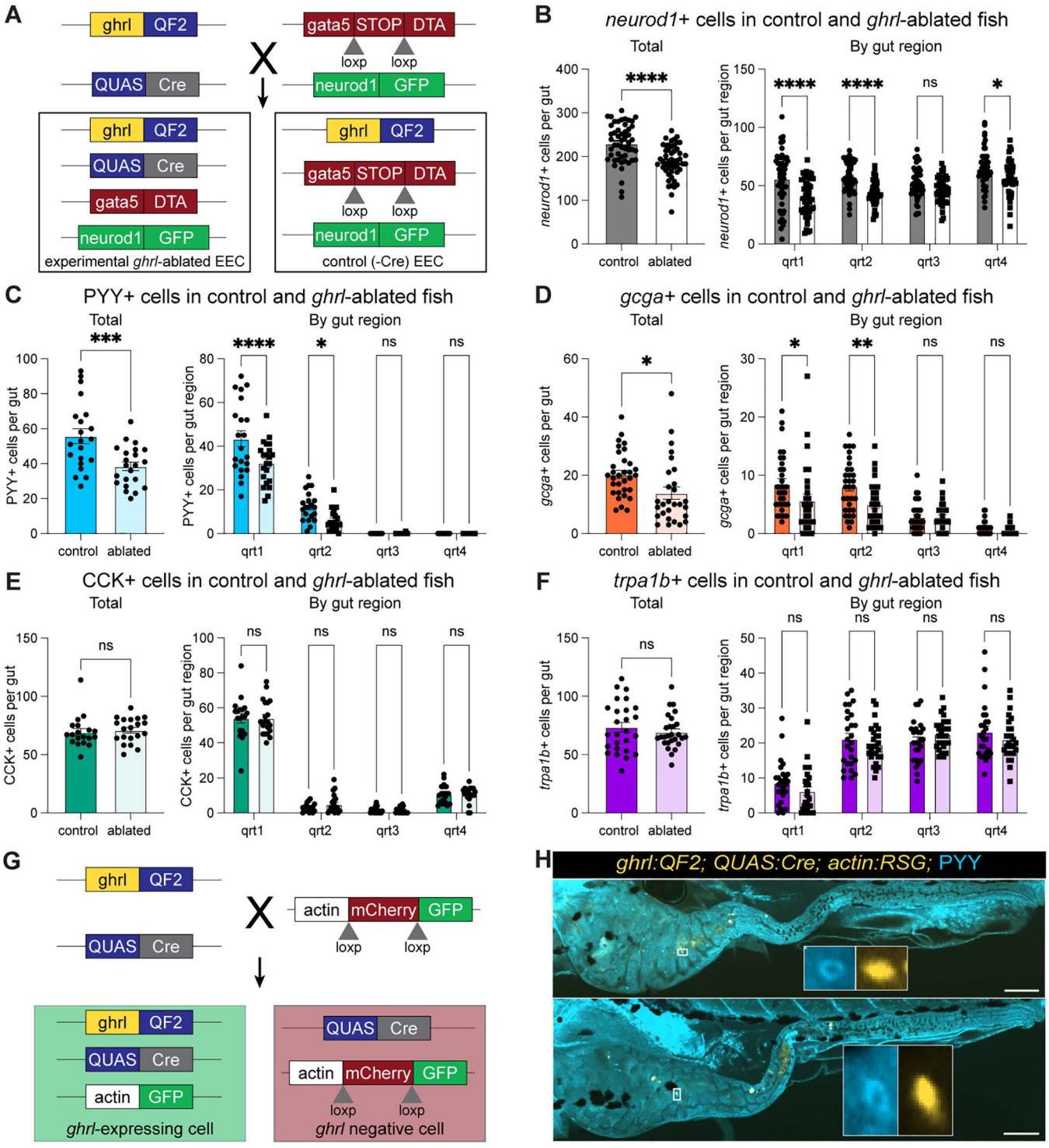


Fig 5. Selective ablation of *ghrl*+ cells suggests lineage relationships between *ghrl*+ cells and other EEC subtypes. (A) Schematic of *ghrl*+ cell ablation. Total and regional counts for **(B)** *neurod1*+ cells, **(C)** PYY+ cells, **(D)** *gcga*+ cells, **(E)** CCK+ cells, and **(F)** *trpa1b*+ cells in control and *ghrl*-ablated fish are shown. Each dot represents a 6-day post-fertilization fish. **(G)** Schematic of *ghrl*+ cell lineage tracing. **(H)** Representative images of *ghrl* lineage tracing with PYY staining showing rare double-positive cells. Scale bars are 100 μm. Statistical significance was calculated by unpaired *t* test for total cell numbers and by two-way ANOVA for regional analysis. Significance annotations are as follows: ns ($p > 0.05$), * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$), **** ($p < 0.0001$). Underlying data can be found in S1 Data.

https://doi.org/10.1371/journal.pbio.3003522.g005

To distinguish these possibilities, we first used lineage tracing tools to indelibly label a cell with GFP if it has ever expressed *ghrl* (Fig 5G). We then performed PYY staining and found rare occurrences of double-positive cells, confirming that some PYY+ cells have indeed passed through a *ghrl*+ cell state (Fig 5H). This suggests that the impact of *ghrl*+ cell ablation on PYY+ cells is at least in part cell autonomous. Due to overlapping fluorophores, we were not able to perform a similar study with *gcga*+ cells.

Given the small number of *ghrl*-derived PYY+ cells in our lineage tracing analysis, we hypothesized that *ghrl*+ cells could also be regulating other EEC subtypes non-cell autonomously through secreted products. As ghrelin peptides were prevalent in our peptidomics dataset (Fig 2D) and have been shown to stimulate growth of intestinal cells [116,117], we tested if loss of the *ghrl* gene impacted EEC development. We used CRISPR/Cas9 to generate *ghrl* mutant zebrafish and evaluated the effects on *neurod1*+ EECs. We found that EEC numbers were equivalent across *ghrl* mutant and wildtype fish (S13 Fig), suggesting that loss of *ghrl* gene function does not overtly impact EEC development.

Ablation of *neurod1*+ cells, but not *ghrl*+ or *neurog3*+ cells, impacts adult survival and growth

To test the ramifications of each of our EEC ablation models on zebrafish growth and survival, we raised ablated fish with their sibling controls to adulthood. We maintained at least five tanks of each control and ablated condition for each experimental replicate and checked survival at several timepoints. At the end point of each study, we also randomly selected fish in each group for measurement of standard length (S14A and S14B Fig) as a metric of animal growth [118]. As we anticipated, *neurod1*+ ablated fish showed the most severe phenotype, with evidence of reduced survival as early as 2 weeks post fertilization (S14C Fig). Additionally, at the 6-week endpoint, *neurod1*+ ablated fish were also observed to be smaller than their sibling controls (S14C Fig). In contrast, neither *neurog3*+ nor *ghrl*+ ablated fish showed survival or growth defects over 6 weeks (S14D, S14F Fig). To further interrogate any possible differences, we raised a second set of *neurog3*+ and *ghrl*+ ablated and control fish to 12 weeks post fertilization, a commonly accepted time point for full maturity [118]. However, we still did not observe any differences in survival or growth (S14E, S14G Fig), suggesting *neurog3*+ and *ghrl*+ ablation have no significant impact on feeding behavior and adult development.

Discussion

Here, we report seven distinct EEC populations in zebrafish and characterized reporters for six of them. Our characterization of EEC subtype reporters is consistent with previously published larval zebrafish in situ hybridization data, building confidence in our proposed EEC subtype identities [28,30]. These studies agree with our data that first onset of these genes is around 3 dpf (72 hpf) with expression increasing by 4 dpf [28,30]. While PYY and CCK antibody staining cells were rare at 3 dpf in our study, *pyyb* and *ccka* in situ hybridization data showed a greater number of positive cells [30], suggesting that protein expression lags slightly behind transcript presence, as might be anticipated. Despite this slight difference, the regional localization of these cells across all timepoints and the number of positive cells at 4 dpf were consistent between our observations and published studies [28,30]. Intriguingly, prior in situ hybridization analysis of genes that we report here as expressed in the same EEC subtype showed the same prevalence and regional patterns, consistent with them being present in the same cells. For example, EEC subtype cluster 1 in our scRNA-seq analysis expressed *ghrl, gip, mlnl,* and *sst2* and was labeled in this study with a *ghrl* transgene reporter. Both our *ghrl* transgene reporter and published in situ data for *ghrl, gip, mlnl,* and *sst2* labeled <10 cells in the proximal gut at 4 dpf. Furthermore, previous studies showed *ghrl* and *mlnl* in situ probes overlap, as would be predicted from our data [30]. Additionally, our identification of EEC cluster 6 as enterochromaffin cells is consistent with previous reports of colocalization of enterochromaffin markers *trpa1b*, *tph1b,* serotonin, and enkephalin (a product of *penka*) in a subset of zebrafish EECs that communicate with the nervous system [27,119]. Finally, our analysis of zebrafish EECs in a separate scRNA-seq dataset showed subtypes with very similar EEC hormone distributions (S3 Fig). This consistency in EEC subtypes across many datasets and modalities builds confidence in our findings.

We show larval and adult EECs have remarkably consistent subtype signatures (Figs 1D, 1E, 1G, and S2D–S2F), suggesting that EEC subtype programing is established in larvae prior to feeding and persists through adulthood. Similar studies across ages in other animals are limited. One scRNA-seq study of the human intestine looked at EECs from adult, pediatric, and fetal tissue samples. Although a comparison across subject age was limited by 90% of the EECs coming from fetal tissue, the cells did not clearly separate by age. Instead, as in our study, clustering was largely driven by hormone expression [75]. Additional studies of mouse and human fetal EECs have likewise shown hormone expression patterns similar to those reported from adult tissue [51,115,120,121]. This suggests that the establishment of durable EEC subtype programs during early development may be a common feature of EECs in fishes as well as mammals.

The repertoire of hormone coding genes expressed in EECs is very similar between zebrafish and mammals (Fig 1F), despite over 400 million years having elapsed since their last common ancestor [122]. We found that orthologs of most commonly reported mammalian EEC hormone coding genes were expressed in our scRNA-seq dataset, with the exceptions of *PPY*, *IAPP*, and *NPY*. *PPY* does not appear to have an ortholog in zebrafish. *IAPP* has an orthologous novel gene that has not been well described in zebrafish, and *NPY* similarly has a zebrafish ortholog, but has largely been characterized in the brain [123,124]. Conversely, two EEC hormone coding genes were found in our zebrafish dataset that have not been commonly reported in mammalian EECs, namely *pdyn* and *tac4b*. *PDYN* and *TAC4* are present in mammals but have not yet been reported in EECs. Despite these small differences, the remarkable conservation of the overall EEC hormone gene repertoire between zebrafish and mammals highlights the critical roles of these signaling molecules during vertebrate evolution.

Additionally, zebrafish orthologs for *GASTRIN* (*GAST*) and *SCT* were previously not identified and were annotated in this study. Despite Bayliss and Starling reporting in 1903 that extracts from salmon intestine have secretin activity [125], studies since then have repeatedly failed to identify secretin orthologs in fishes via sequence homology or synteny analyses [126–130]. Secretin is a member of the PACAP/GCG family of peptides that likely originated from a single exon present over 600 million years ago before the two rounds of whole genome duplication that occurred during the vertebrate radiation [128,129,131–135]. Interestingly, many genes in that family encode two non-overlapping peptides, but mammalian *SCT* is believed to have lost its second C-terminal peptide [135]. Here, we show that there is a region of the Secretin protein C-terminal to the annotated secretin peptide that is highly conserved across fishes and produces highly processed peptides in zebrafish (gray shading in Fig 2F). This suggests that mammals and fishes each maintained only one peptide from the ancestral homolog, with the fish peptide sequence differing enough from the ancestral sequence as to be missed on previous homology searches [126–130]. The recent efforts to sequence genomes of species like gar (*Lepisosteus oculatus*) [95] and elephant shark (*Callorhinchus milii*) [96] allowed us to connect the unannotated zebrafish *sct* gene to more ancient sequences and ultimately identify it as a *sct* ortholog. Intriguingly, *sct* in chicken and other birds has been shown to encode two peptides, both of which were able to bind to and stimulate the chicken secretin receptor SCTR in cultured mammalian cells [136]. The possible in vivo role of these *sct*-derived peptides in birds is unclear, however, as chicken *SCTR* is lowly expressed in the pancreas (where it is known to be enriched in mammals) but more highly expressed in intestine and testes [136], offering other possible sites of action. Together, these data suggest that the second C-terminal Secretin-derived peptide presumed lost in mammals may still be expressed and functional in other vertebrate species. Further studies investigating the roles of both possible peptides encoded by zebrafish *sct* are warranted.

Our peptidomic analysis identified peptides aligning to 19 of the 23 hormone coding genes expressed in our scRNA-seq data, including the newly-identified zebrafish *sct*. Of note, the liquid chromatography and mass spectrometry settings used to generate our EEC peptidomic dataset may have slightly biased the peptides we were able to detect. For example, as in many standard protocols, singly charged peptides were not selected for fragmentation, meaning that we would not identify many Leu and Met enkephalin peptides that might be derived from *proenkephalin a* (*penka*) or *pdyn*. *CCK*-derived peptides have also been shown to be singly charged in human studies, and, for that and several other reasons, have proven exceedingly difficult to identify by LC–MS/MS [137]. Furthermore, the complexity of peptidomic samples, including the high charge states and complex fragmentation, complicate the detection of longer peptides, although peptides up to

60 amino acids long have been reported using similar techniques [138,139]. Additionally, as noted in the Results, the atlas here reports diversity of unique peptides produced from each gene, not their relative natural abundance in EEC secretory granules. Finally, while we analyzed both larval and adult samples, we did not perform quantitative comparison between the two groups. While peptides from adult samples are over-represented in our results (S2 Table), this is likely because the total cell content in adult samples was almost twice that of larval samples (50,000 versus 30,000 cells) due to limitations in sorting and collecting EECs from larvae. As such, for peptides not detected in larval samples, we cannot conclude that they were absent as they may have been below the limit of detection. For this reason, we did not perform direct comparisons between larval and adult samples here. This and other questions raised by our work here would be assisted by the development of antibodies against zebrafish EEC peptides.

We hope the zebrafish EEC peptides reported here will be leveraged for further discovery related to EEC hormone endocrinology, and we refer the reader to the supplemental materials for a complete catalog of the detected peptides (see S4 Fig and S2 Table). There are exciting possibilities for discovery related to the peptides we report here that show homology to known human hormones. Historically, studies of homologous peptide hormones in non-mammalian models have provided critical insights into hormone function and even informed therapeutic design. For example, the first identification of GLP-1 occurred in anglerfish [140], and exendin-4, a peptide hormone produced by gila monster lizards that is analogous to human GLP-1, contributed to the development of long-acting GLP-1 analogs now broadly used for treatment of diabetes and obesity in humans [141–143]. We detected several zebrafish peptides that map to the regions of *gcga* and *gcgb* that align with the GLP-1 encoding portion of human *GCG* (S4B Fig). The arginine 36 residue in GLP-1 is commonly amidated in mammals [144,145], as is seen with one of the *gcgb*-derived peptides detected in zebrafish. However, there are also notable differences between zebrafish *gcga*/*gcgb* and human *GCG* and their corresponding peptides. Firstly, *gcgb* and isoform 2 of *gcga* appear to have lost the GLP-2 coding region, a phenomenon also reported in other fishes [146,147]. Additionally, the GLP-1 region in zebrafish Gcga lacks the highly conserved dibasic PCSK1/2 cleavage site at its C-terminus [148], suggesting alternative processing of this peptide, perhaps at the dibasic RR site immediately upstream of the GLP-2 region. This would result in a 45 amino acid peptide, which would be more challenging to detect and might explain why, although there are many detected peptides aligning to the GLP-1 region in zebrafish, there is not one that perfectly overlaps with the predicted GLP-1 sequence. These nuances in peptide structure and sequence could provide valuable insight into the evolution, function, and regulation of these hormones.

Beyond the few peptides discussed in the Results above, we also detected peptides aligning to regions annotated in humans as coding for peptide hormones like pituitary adenylate cyclase-activating polypeptide (PACAP), glucagon, glicentin-related polypeptide, calcitonin, katacalcin, galanin, insulin-like 5 (INSL5) A and B chain, motilin, neuromedin B, proenkephalin, Leu and Met enkephalins, synenkephalin, neuronostatin, and somatostatin. These could be exciting candidates for future studies into EEC peptide hormone biology. There are also opportunities to potentially uncover novel signaling molecules among the peptides that are abundant in zebrafish but align to a region with no annotated function in humans. For example, recent studies have illustrated how new technologies and datasets such as this can continue to reveal novel peptide hormones [77,78,149,150]. While these are a few examples, we hope the entire zebrafish EEC peptide catalog will be a resource for future discovery related to EEC hormones (S4 Fig, S2 Table).

We also detected peptides derived from *glucagon b, somatostatin 1.2,* and *insulin* (S4B, S4M, and S4P Fig), hormone genes that are commonly thought to be enriched in the pancreatic islet [30,151,152]. While this could be evidence of pancreatic contamination in our sorted EEC samples, those genes do show some limited expression in the joint larval and adult EEC dataset (S5E Fig) and have been reported in zebrafish intestinal bulk RNA-seq datasets before [30,153], suggesting these peptides might be produced in zebrafish EECs.

Despite striking conservation of EEC hormone coding genes and peptides, the partitioning of those genes into EEC subtypes shows notable differences between zebrafish and mammals. Specifically, while *gip* and *nts* are present in zebrafish EECs, *gip* or *nts*-expressing cells do not separate into their own clusters as they often do in mammalian datasets, perhaps

because they are expressed at relatively lower levels in zebrafish [19,52,53]. Additionally, while *PYY* and *GCG* are found together in what have traditionally been called "L cells" in mammals, we see their orthologs *pyyb* and *gcga* separate into clusters 2 and 3, respectively, in our dataset. Of note, cluster 2 and 3 are somewhat intermingled in the UMAP plot and have the most overlapping marker profile of any two clusters (Figs 1G and S2D–S2F), suggesting a close relationship between the two populations. Furthermore, scRNA-seq reports from mouse colon, where *Pyy* is more highly expressed, have separated *Pyy* and *Gcg* into separate clusters [107]. Finally, we also observed high expression of *insulin-like 5* paralogs *insl5a* and *insl5b* in our dataset. The mammalian homolog *INSL5* is mostly restricted to colonic EECs [51,54,107,154]. As there is limited homology between posterior regions of the zebrafish intestine and the mammalian colon [155–157], *insl5a*+ and *insl5b*+ cells likely have a different distribution in zebrafish than they do in mammals. These findings suggest that some differences in EEC subtype identities between zebrafish and mammals may be due to sub-specialization of intestinal regions in mammals.

Our results indicate that *neurog3* has a smaller role in EEC differentiation in zebrafish than in mouse. *Neurog3* has been shown to be necessary and sufficient for committing secretory progenitors to an endocrine fate in both pancreatic islet endocrine cells and EECs in mammals [10–14,158–160]. Although *neurog3* was shown to be dispensable for islet development in zebrafish, its role in EECs had not been formally tested [29]. We observed a more minor role for *neurog3*+ cells in EEC differentiation in zebrafish where only some EEC subtypes—namely *trpa1b*+, CCK+, and PYY+ cells—seem to pass through a *neurog3*+ cell state. This contrasts with our pseudotime data, which predicted that most EEC subtypes, and specifically *ghrl*+, PYY+, and *gcga*+ populations, pass through an EEC progenitor population enriched for *neurog3* expression (Fig 1H). This discrepancy could be because *neurog3* labels only a portion of our putative EEC progenitor population, such that some EEC progenitor cells in cluster 0 are unaffected by *neurog3* ablation and are still able to differentiate into EEC subtypes. Alternatively, the more limited role of *neurog3*+ cells we observed in vivo could be because cell intermediates of alternative differentiation trajectories from secretory progenitors to differentiated EEC subtypes might have been poorly captured in our scRNA-seq data. For example, a few secretory progenitor cluster 8 cells cluster with *ghrl*+ cluster 1 cells, raising the possibility that there could be a path from cluster 8 to cluster 1 that does not pass through *neurog3*+ cells in cluster 0. Finally, cell ablation studies always carry the potential caveat of compensation, in which removing one population makes room for another to at least partially take its place. Despite these methodological limitations, our in vivo studies demonstrate that ablation of *neurog3*+ cells only impacts a subset of EEC subtype populations. This partial effect of *neurog3*+ cell ablation is consistent with published *neurog3* in situ hybridization data [28] and findings that support *neurod1* as a major fate-determining transcription factor in zebrafish EECs [27,31].

Prior studies in mammals raised the intriguing possibility that *Ghrl*-expressing EECs may have a role in the differentiation of EEC progenitors into mature EEC subtypes, but this had not been directly explored. Reports in mice showed *Ghrl* expression and *Ghrl*+ cell numbers were uniquely resistant to deletion of transcription factors important for EEC differentiation, such as *Nkx2.2* [65–67], *Arx* [112,161], *Pax4* [112,114], *Sox4*, *Tox3*, [19] *Myt1* [19], and *Isl1* [113]. While *Ghrl*+ cell numbers were maintained or often increased upon loss of those transcription factors, almost all other EEC subtype markers were reduced. Notably, our zebrafish data showing *ghrl*+ cells were unaffected by *neurog3*+ cell ablation is consistent with this pattern. Altogether, these data support two possible models to explain why *Ghrl*+ cells are uniquely enriched upon loss of various transcription factors important for EEC differentiation: 1) *Ghrl*+ cells are an early-arising mature EEC fate that becomes the default destination for progenitors in the absence of key transcription factors; or 2) *Ghrl*+ cells themselves give rise to other subtypes through the subsequent expression of key transcription factors, and, when those transcription factors are lost, EECs become arrested in a *Ghrl*+ state. Data from different scRNA-seq studies in mammalian EECs are consistent with both models. Computational analysis of time series scRNA-seq data from mouse organoids show that *Ghrl*+ cells are some of the first to arise [19], supporting the first model. However, additional scRNA-seq studies have identified "*Ghrl*+ EEC progenitors." These cells express progenitor markers, including transcription factors important for EEC differentiation, but also express *Ghrl*, thought to be a mature EEC marker [23–25]. As in our study, one of these studies found a predicted pseudotime differentiation trajectory moving through a *Ghrl*+ EEC progenitor state [24],

supporting the second model. Interestingly, studies reporting *Ghrl*+ EEC progenitors also report *Ghrl* expression in later, more mature EEC populations, suggesting the possibility of a hybrid model with two distinct *Ghrl*+ populations, only one of which serves as a progenitor. Despite the preponderance of suggestive evidence, however, a role for *Ghrl*+ cells in EEC differentiation in the intestine has not, to our knowledge, ever been formally tested.

In this study we show that ablation of *ghrl*+ cells impacts other EEC subtypes, namely *gcga*+ and PYY+ cells, and that *ghrl*+ and PYY+ cells have a lineage relationship, in line with model 2 proposed above. Lineage relationship between *ghrl*+ and *gcga*+ cells could not be tested here due to lack of an antibody labeling *gcga*+ cells. While we acknowledge that lineage-based approaches have limitations and can confuse lineage dependency and co-expression, these data are supportive of a role for *ghrl*+ cells in the differentiation of other subtypes. Interestingly, while lineage tracing of *ghrl*+ cells revealed rare instances of overlap with PYY, these events were not frequent enough to fully explain the reduction of PYY cells seen in *ghrl*+ ablated fish. As EECs are known to signal to each other [162–166], it is possible that loss of the signals secreted from *ghrl*+ cells is the primary driver of reductions in other EEC subtypes upon *ghrl*+ cell ablation. We did not see any effect on total EEC number following *ghrl* deletion, however, suggesting that loss of ghrelin signaling is not the cause.

Our findings here together with the existing literature support a new model of zebrafish EEC differentiation (Fig 6). EECs are derived from secretory progenitors and are marked by the expression of *neurod1,* which is required for all EECs in zebrafish [27,31]. Our findings here support that a portion of *trpa1b*+*,* CCK+, and PYY+ EEC subtypes pass through a *neurog3*+ cell state and a portion of *gcga*+ and PYY+ EECs pass through a *ghrl*+ cell state. Both *neurog3*+ and *ghrl*+ cell ablation had only partial impacts on other EEC subtype populations, indicating *neurog3*- and *ghrl*-independent differentiation pathways are also available. Multiple differentiation pathways may contribute to EEC subtype plasticity or compensation, but it is unclear if there are functional differences between cells of the same subtype that differentiate through varying pathways. We also acknowledge that ablation of *ghrl*+ or *neurog3*+ cells may have fundamentally shifted the transcriptional profile of EEC subtype populations. While we evaluated the numbers of cells labeling with each subtype reporter, we do not know if the subtype biology beyond that single reporter is preserved.

An important limitation of this study is its reliance on transgenic reporters to label and ablate EEC subtypes. Most of the transgenic lines used here, most notably those for *neurog3* and *ghrl*, use short promoter sequences from those loci

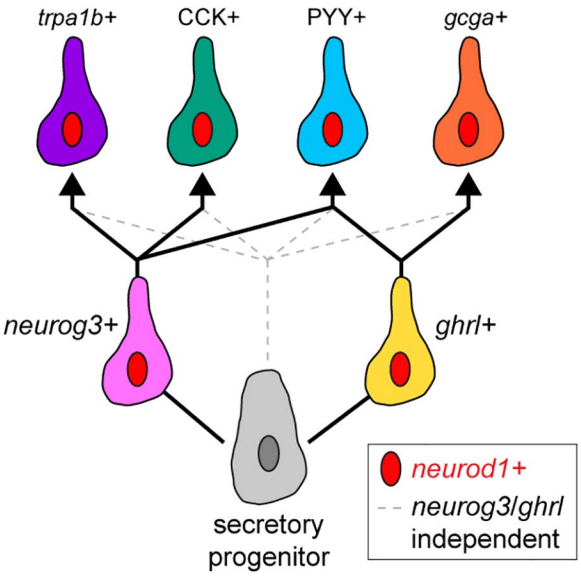


Fig 6. Working model of EEC subtype differentiation in zebrafish.

https://doi.org/10.1371/journal.pbio.3003522.g006

that may not perfectly reflect the expression patterns of the respective endogenous genes. Further, use of the QF2-QUAS system is prone to perdurance of QF2 after the promoter driving its expression is no longer active. As mentioned in the Results, this caveat is particularly apparent in the disparity between the large number of cells labeled by our *neurog3:QF2* transgene and the relatively few cells that express *neurog3* in our scRNA-seq dataset (Figs 1E and 3B–3D). Thus, as with most lineage tracing studies, our findings using these reagents suggest developmental relationships that should be validated in the future using additional technologies.

In conclusion, the tools and datasets reported here can serve as a platform for future hypothesis testing to better understand fundamental EEC subtype and hormone biology. For example, our atlas of endogenous EEC peptides can serve as a starting point for testing their individual endocrine functions and structure-function relationships. The toolkit of EEC subtype reporters also provides a valuable resource for further in vivo studies. The modularity of the QF2-QUAS transgenic tools deployed here will allow future studies to not only label, but selectively activate or inhibit certain subtypes to understand their physiologic roles. Future work could test the subtype-specific impacts of various challenges such as dietary changes, microbial manipulations, or disease models.

Materials and methods

Ethics statement

All zebrafish studies were approved by the Institutional Animal Care and Use Committees of Duke University (protocol A061-22-03) and Michigan State University (protocol PROTO202200367).

Zebrafish lines and husbandry

Zebrafish stocks were maintained on an Ekkwill (EK) background on a 14-hour/10-hour light/dark cycle at 28.5°C in a recirculating aquaculture system (Pentair). From 6 to 14 dpf, larvae were fed Zeigler AP100 <50-micrometer larval diet (Pentair, LD50-AQ) twice daily. Beginning at 14 dpf, larvae were fed *Artemia* (Brine Shrimp Direct, BSEACASE) once daily, supplemented with a daily feed of Skretting Gemma Micro 75 (Bio-Oregon, B5676). From 28 dpf, the Gemma Micro 75 diet was replaced with Gemma Micro 300 (Bio-Oregon, B2809). At the onset of breeding age or sexual maturity, adult fish were given a 50/50 mix of Skretting Gemma Micro 500 (Bio-Oregon, B1473) and Skretting Gemma Wean 0.5 (Bio-Oregon, B2818) and one feeding of *Artemia* daily.

For experiments involving zebrafish larvae, adults were bred naturally in system water and fertilized eggs were transferred to 100 mm petri dishes containing ~60 mL of egg water at approximately 6 hours post-fertilization. The resulting larvae were raised under a 14-hour light/10-hour dark cycle in an air incubator at 28°C at a density of 1 larvae/mL water. All the experiments performed in this study ended at 6 dpf unless specifically indicated. As sex determination in zebrafish occurs after 6 dpf, we do not report numbers of male or female fish in each experiment. The following engineered zebrafish lines were used in this study:

TgBAC(cldn15la:EGFP)^{pd1034} [167], *Tg(neurod1:TagRFP)*^{w69} [168], *TgBAC(trpa1b:EGFP)*^{a129} [102], *TgBAC(gata5:loxp-mcherry-stop-loxp-DTA)*^{pd315} [27], *TgBAC(neurod1:EGFP)*^{nl1} [169], *Tg(gcga:EGFP)*^{ia1} [170], Tg(*β-actin2:loxP-mCherry-loxP-GFP*)^{pd31} [171], *Tg(QUAS:GFP; cryaa:mCherry)*^{pd1199} [172], *neurog3*^{sa211} [29], *Tg(neurog3:QF2; cmlc2:GFP)*^{rdu107} (generated in this study), *Tg(ghrl:QF2; cmlc2:GFP)*^{rdu108} (generated in this study), *Tg(QUAS:cre; cryaa:mCherry)*^{rdu109} (generated in this study), *Tg(neurog3:Cre; cmlc2:GFP)*^{rdu110} (generated in this study), *ghrl* ^{rdu111} (generated in this study), and *Tg(QUAS:TagBFP)*^{rdu113} (generated in this study).

Construction of zebrafish lines

The Gateway Tol2 cloning approach was used to generate *Tg(neurog3:QF2; cmlc2:GFP), Tg(ghrl:QF2; cmlc2:GFP), Tg(QUAS:cre; cryaa:mCherry),* and *Tg(neurog3:Cre; cmlc2:GFP)* plasmids [103]. The 2114 base pair *neurog3* promoter

was previously reported [106] and generously shared by Dr. Olov Andersson (Karolinska Institutet) and cloned into a p5E-MCS vector [103] using TAKARA In-Fusion Snap Assembly master mix (Takara, 638947) with HindIII-linearization of the backbone and amplification of the insert with primers CGGTATCGATAAGCTCCGCGGCCGCCGTACTCGA and ATTCGA-TATCAAGCTCGGCGCGCCCACCCTTTCTGTA. The pME-QF2 [105], p5E-QUAS [173], p3E-polyA [103], pDestTol2CG2 [103], and pDESTtol2pACrymCherry [174] plasmids were previously reported and obtained from Addgene. The pME-Cre plasmid was previously reported [175] and generously shared by Dr. Mark Cronan. The 647 base pair *ghrl* promoter was amplified from purified (Wizard Genomic DNA purification kit, A1120) EK DNA using primers cttaagcttTTCAGAATTCATAT-CAGATCACAGACACT, cttccgcggCACCTTAGTCTTAATTCTTTGCTACATAC. It was then cloned into the p5E-MCS [103] via digestion of insert and backbone with HindIII (NEB, R3104S) and SacII (NEB, R01575S) followed by phosphatase treatment (NEB, M0289S) and ligation (NEB, M0202S). The appropriate p5E, pME, and p3E entry vectors were cloned into destination vectors with either a green heart (*cmlc2:GFP;* pDestTol2CG2) or red eye (*cryaa:mCherry*; pDESTtol-2pACrymCherry) as listed above using an LR-Clonase (ThermoFisher, 12538-120) reaction.

The *ghrl* mutant line was generated using CRISPR-Cas9. The guide RNAs (gRNAs) were designed using the "CRIS-PRscan" tool (www.crisprscan.org/) [176] and purchased fully synthesized from Integrated DNA Technologies (see S4 Table). At the one-cell stage, EK strain zebrafish embryos were injected with 1–2 nL of a cocktail consisting of Cas9 protein (1,000 ng/μL), gRNA (150 ng/μL), and 0.05% phenol red. Injected embryos were screened for mutagenesis with forward primer CACAGTGACAGTTGTAGACTTTAATGCTAAT and reverse primers GTCTCTAAGAAGATTCTC-CAGAAGATTCTGA (mutant), AAACATGCTGCTGGCACGGCA (wildtype). The mutations were further determined through Sanger sequencing of the region encompassing the gRNA targeting sites. Ten F0 injected fish were out-crossed to wild-type fish and progeny were screened to test if mutations were transmissible. F1s from F0 founders were then raised to adulthood and sequenced. Two F1s were found to have identical lesions—the deletion started after the first nucleotide in the first exon and continued through the first 17 nucleotides of the third exon—and were bred to generate a stable line. F2s and F3s from this line were used in the reported experiments and were given the allele name *rdu111*. When testing *ghrl* mutants, genotyping of the *mttp* locus was used as a control to verify there was viable DNA in samples using forward primer AGAGACGGTGTCCAAGCAGG and GCTCAAAGACTTTCTTGC with an expected band of 137 base pairs [177].

mVISTA alignment

To identify conserved regions of the *ghrl* promoter across *Danio* species, we exported the *ghrl* sequence and annotation from ensembl.org [97]. We blasted the *D. rerio* Ghrl amino acid sequence against other *Danio* species and extracted 5 kb upstream and downstream of the aligned region for each species. We then put each of these sequences into the mVISTA alignment [178,179], using the *D. rerio* sequence and annotation as the reference. We identified and amplified a 647 base pair region upstream of the *ghrl* transcriptional start site that was highly conserved across *Danio* species (S7B Fig).

Larval immunofluorescence and live imaging

All larval imaging was performed on Andor Dragonfly Spinning Disc Confocal plus with a 20× lens in the Duke Light Microscopy Core Facility. For live imaging experiments, zebrafish larvae were anesthetized with tricaine and mounted in 1% low-melt agarose. For larvae imaged at 3 and 4 dpf, the yolk was mechanically removed with fine watchmaker forceps prior to mounting.

Whole-mount immunofluorescence staining was performed as previously described [26]. Larvae were euthanized with 5× tricaine and roughly 20 larvae per condition were transferred to a 1.5 mL tube. Excess media was removed and 1 mL of chilled 4% paraformaldehyde was added. Larvae were incubated in fixative overnight (>16 hours) at 4°C. Samples were then washed 2 × 5 min in PBS. For samples of 3 or 4 dpf larvae, yolk was then mechanically removed using fine watch-maker forceps before dehydrating in 50% methanol/50% PBS solution for 5 min. Larvae were then washed 3× in 100% methanol before incubation in 100% methanol at −20°C for at least 2 hours. Larvae were then rehydrated in serial 50%

methanol/50% PBS, 100% PBS washes, and permeabilized at room temperature for at least 20 min with PBS with 0.1% Tween-20, 1% DMSO, and 0.3% Triton X-100. Larvae were then washed and blocked with 4% bovine serum albumin solution for at least 30 min at room temperature before washing and incubating in primary antibody solution for at least 24 hours at 4°C. Primary antibodies in this study were diluted at 1:100 [rabbit anti-PYY (custom antibody generated in [100]), rabbit anti-CCK (custom antibody generated in [99]), rabbit anti-Ghrl (Anaspec, 55529), goat anti-IgG (Jackson ImmunoResearch, 115-005-044)] or 1:500 [chicken anti-GFP (Aves, GFP-1010)]. Larvae were then washed every 15 min for 2 hours before incubation with Alexa Fluor Invitrogen secondary antibodies diluted at 1:250 [488 goat anti-chicken (ThermoFisher, A11039), 647 goat anti-rabbit (ThermoFisher, A21244), 568 goat anti-rabbit (ThermoFisher, A11011)]. Larvae were incubated with secondary antibodies overnight (>16 hours) at 4°C, washed every 15 min for 1 hour, and mounted for imaging.

neurod1 cell analysis in *neurog3* mutants

The *neurog3*^{sa211} mutant allele was obtained from the Zebrafish Mutation Project (Sanger Institute, UK). Adult heterozygous *neurog3*^{sa211} zebrafish were bred naturally in system water, and fertilized eggs were collected, transferred to 100 mm petri dishes containing embryo medium, and allowed to develop at 28.5°C. All steps of the staining were performed on nutator at room temperature except when indicated otherwise. Approximately 30 larvae were fixed in 750 μL of 4% PFA prepared in 1× sweet buffer (8% sucrose, 0.2 M $CaCl_2$, 0.2 M PO_4 buffer, pH 7.3) for 2 hours. Then, larvae were washed twice in 0.5% Phosphate buffer saline-Triton X-100 (PBSTX) for 15 min, and stored in PBSTX at 4°C. Prior to blocking, larvae were washed 5× in double distilled water for a minimum of 60 min each and then they were incubated for 1 hour in blocking solution [0.5% PBS-Triton X-100, 1% DMSO, 2% bovine serum albumin (Sigma-Aldrich, A3059), 5% normal growth serum (ThermoFisher, PCN5000). Next, larvae were incubated for 16–18 hours in anti-DsRed primary antibody (1:300; Takara Bio, 632496) diluted in blocking solution. Larvae were then washed 3× for a minimum of 3 hours each in 0.3% PBS Triton X-100 before incubating for 16–18 hours in Alexa Fluor 568 secondary antibody (1:1,500; ThermoFisher, A11035) diluted in blocking solution. Finally, larvae were washed 3× for a minimum of 3 hours each in dark and stored at 4°C in 0.3% PBS Triton X-100 before mounting.

Before imaging the immunostained intestines, each larva was tail clipped for genotyping and the body stored in 0.3% PBSTX at 4°C. For extraction of genomic DNA, the fin tissue was incubated in 25 μL of 1× lysis buffer [per 1 mL: 100 μL of 10× PCR buffer (BioLabs NEB, B9004S), 50 μL ~20 mg/mL of proteinase K (ThermoFisher, 100005393), and 850 μL of nuclease-free water] at 55°C for 50 min, followed by 98°C for 10 min. PCR amplification was performed using 1 μL of genomic DNA, 19 μL master mix containing 10 μL of 2× GoTaq Green Master Mix, 7 μL of nuclease-free water, and 1 μL each of the forward (5′-CAAGATCCTCGTGCGCTTTG-3′) and reverse (5′-CCTGATCCTGGGCGTGATTT-3′) primers. The PCR protocol was as follows: 95°C for 2 min; 34 cycles of 95°C for 30 s, 66°C for 30 s, and 72°C for 60 s; final extension at 72°C for 5 min; hold at 12°C. PCR products were cleaned using ExoSAP-IT following the manufactures protocol (ThermoFisher, 78201.1.ML), then sent for Sanger sequencing. Larvae were classified as homozygous wild type or homozygous mutant based on the absence or presence of the mutant allele (ZMP, Zebrafish Mutation Project, Zebrafish Mutant Resource) [180,181]. Only intestines of homozygous wild-type or homozygous mutant were mounted and imaged.

For quantification of EECs, intestines were dissected with forceps and mounted on glass slides in 0.3% PBSTX or Fluoromount-G (SouthernBiotech, 0100-01) and covered with a coverslip. Confocal z-stacks images were captured using the 3i spinning disk confocal microscope with a 20× objective, operated through Slidebook6 software (Version 6.0.13).

In situ hybridization chain reaction

Probes for *ghrl* RNA (S4 Table) were designed according to previously published parameters [182]. *cubn* probes were generously shared by Dr. Michel Bagnat. Hairpins and amplification, hybridization, and wash buffers were purchased from Molecular Instruments [183]. Our methods were adapted from previously published procedures in zebrafish [109,182]. 6

dpf larvae were fixed in 4% PFA for 3 hours at room temperature, followed by PBS and ice-cold acetone washes before being incubated in acetone at −20°C for 8 min. Larvae were then washed with PBS and incubated with Molecular Instruments probe hybridization buffer for 30 min rotating at 37°C. Larvae were then incubated in probe solution (8 nM of each probe) for 48 hours rotating at 37°C. To remove excess probes, larvae were washed 4 × 30 min each in wash buffer pre-warmed to 37°C followed by 2 × 10 min washes in 5× SSCT buffer at room temperature. Larvae were then incubated in amplification buffer for 30 min at room temperature while hairpins were prepared. Sixty pmol of each hairpin were heated separately to 95°C for 90 s and then snap cooled in the dark at room temperature for 30 min. Larvae were incubated in hairpin solution rotating in the dark overnight at room temperature. Excess hairpins were removed with 2 × 5 min, 2 × 30 min, and 1 × 5 min washes in SSCT buffer at room temperature. Larvae were then mounted and imaged with spinning disc confocal.

Image analysis with ImageJ

Images were processed and analyzed in ImageJ [184]. Images were automatedly blinded such that the scorer did not know the condition (e.g., ablated or control) of each image. The start and end of the gut was manually marked using the Cell Counter plugin and *x*, *y* coordinates were extracted. *neurod1*+ cells were automatedly counted from maximum intensity projections using the Analyze Particles function with minimum object size of 10 pixels^2. The number and location of cells was automatedly extracted in R and locations were zero-ed using corresponding manual gut markers. To account for overlapping cells being grouped into a single object, the median object size was used to assign a weight to each object. Objects twice as large as the median size given a weight of 2, objects three times as large as the median were given a weight of 3, and so on, with a minimum weight of 1. Weights were summed to count the total number of cells in an image.

neurog3+ cells were counted with a similar automated approach. A composite of *neurod1* and *neurog3* channels was used to identify objects using the Analyze Particles function and the fluorescence of each channel was extracted for each identified object. Objects with a ratio of *neurog3:neurod1* signal >2 were counted as *neurog3*+ only, objects with a ratio of *neurog3:neurod1* signal <0.5 were counted as *neurod1*+ only, and all objects with a ratio in between 2 and 0.5 were counted as double positive. The same adjustment for overlapping cells described above was also applied.

Because of increased background of various sources causing complications with automated thresholding and counting, images from *gcga, trpa1b,* and *ghrl* reporters along with PYY and CCK stains were manually counted using the Cell Counter plugin. Counts for reporter characterization in Fig 3 were taken for each slice of the captured image while maximum intensity projections were used for ablation image analysis for higher throughput. In all cases, the number and location of each manually counted cell were extracted for further analysis.

In addition to total counts, the location of automatedly or manually counted cells was determined relative to the manually annotated start and end points of the gut. In cases where a single object was given a weight >1 and counted as multiple cells, the x, y coordinates of the original object was used for each. Cells were binned into four evenly divided bins along the length of the intestine for visualization of counts per quarter of the gut in bar plots (see Figs 4, 5, S9, and S11) or into 100 evenly divided bins for visualization of density per percentile of the gut in density plots (see Fig 3T–3W).

Adult sections

Adult zebrafish 3–4 months post fertilization of either sex expressing *Tg(neurod1:RFP); Tg(ghrl:QF2; cmlc2:GFP); Tg(QUAS:GFP; cryaa:mCherry)* were retrieved from the circulating aquaculture system before morning feeding and transferred to a separate tank to fast before dissection. Several hours later fish were euthanized in 0.090% 2-Phenoxyethanol and dissected to retrieve the intestine. The intestine was fixed in 4% PFA overnight at 4°C then washed 3× the next day with cold 1× PBS. The intestine was mounted in 4% low-melt agarose in cryomolds and cured at 4°C for 1 hour. 200 μm cross sections of the tissue were sectioned via Vibratome (Leica VT1000S) and mounted on slides with Vectashield containing DAPI (Vectashield H-1200). Z-stack images were collected on a Zeiss 780 inverted confocal microscope at either

20× or 40× magnification. Image processing was done in FIJI including maximum intensity projections of z-stacks from each channel.

Adult Swiss rolls

Adult zebrafish (approximately one year post fertilization) expressing *Tg(neurod1:RFP); Tg(ghrl:QF2; cmlc2:GFP); Tg(QUAS:GFP; cryaa:mCherry)* were retrieved from the recirculating aquaculture system and euthanized with 0.090% 2-Phenoxyethanol. Intestines were dissected and flayed open to discard luminal debris using ice-cold 1× PBS. Once cleaned, a few drops of Carnoy's fixative was used to stiffen the tissue prior to rolling and immediately washed with ice-cold 1× PBS. Intestine was rolled from the proximal to distal end, pinned through with a dissection pin (Fine Science Tools, 26002-10) to maintain shape, then transferred to 4% PFA overnight at 4°C. After overnight fixation, the tissue was transferred to a 30% sucrose solution with 0.02% sodium azide and stored at 4°C until tissue sank in preparation for cryosectioning. Intestines were then mounted in cryomolds (TissueTek, 4566) with FS 22 Clear (Leica, 3801480), frozen in a dry ice bath, and stored as blocks at −80°C until sectioning. Blocks were allowed to equilibrate to internal cryostat temperature for 1 hour prior to sectioning and 10 μm sections were collected via cryostat (Leica, CM1860). Slides were stored at −80°C until imaging. For imaging, sections were mounted with Vectashield containing DAPI (Vectashield, H-1200) and glass coverslips and imaged on an Andor Dragonfly confocal at 20x with z-stack and stitching.

Survival assays

neurod1, neurog3, or *ghrl*-ablated larvae were generated alongside sibling controls, which were identical to their ablated counterparts but lacked the Cre transgene. Groups of 25–30 larvae were placed in housing tanks and put on a recirculating aquaculture system at 6 dpf and were fed and maintained according to standard husbandry methods described above. The number of fish remaining in each tank was counted at three evenly spaced timepoints (2, 4, and 6 weeks post-fertilization for initial experiments and 4, 8, and 12 weeks for extended experiments for *neurog3* and *ghrl* ablated conditions). Any tank that had >50% death rate at the first measurement was excluded from analysis as occasional tank die off is known to occur sporadically. Specifically, excluded tanks were as follows: one control tank in the 12-week *neurog3* ablation experiment and one control and two ablated tanks in the 12-week *ghrl* ablation experiment. At the last time point of each experiment (6 or 12 weeks), manual standard length measurements of roughly representative fish from each tank were performed.

Single-cell RNA sequencing analysis

We imported data from previously published adult [49] and larval [48] zebrafish intestinal scRNA-seq datasets into R (version 4.3.1). Using established markers of intestinal secretory cells, we selected just secretory cells from each dataset (S1 Fig) and integrated them using the SCT method in Seurat version 5.0.3 [185]. We further subsetted this dataset for just secretory progenitors and EECs and re-integrated and clustered it to generate the dataset shown in Fig 1B and analyzed in this manuscript. This final dataset included 2069 cells from the adult dataset and 1891 cells from the larval dataset. Multiple resolutions were evaluated and a resolution of 0.4 was ultimately used (S16 Fig). We used monocle3 (version 1.3.4) to perform pseudotime analysis where use_partition was set to TRUE and root_cells were set to cluster 7 [58]. Module scores were calculated using the AddModuleScore() function. The list of genes included in the regional modules score was derived from the enriched genes of the pharynx, anterior enterocyte, anterior LRE, posterior LRE, and cloaca clusters in [109] and are listed in the Supplemental Materials (S10 Fig, S1 Table). The integrated secretory progenitor and EEC dataset is available on Zenodo at at DOI 10.5281/zenodo.17342282.

Of note, the zebrafish *gastrin* gene is currently named either *LOC100536965* or *CR556712.1* in danRer11 and is annotated in Ensembl as a noncoding lncRNA (ENSDARG00000117418) for unknown reasons. However, multiple phylogeny studies have identified the exact genetic coordinates of *LOC100536965*/*CR556712.1* as zebrafish *gastrin* [186–188]

and peptides we detected derived from this gene align with RefSeq XM_021479754.1, which is a coding prediction. We therefore refer to the gene as *gastrin*. When searching for expression of gastrin in our scRNA-seq dataset, we found expression under the gene name *CR556712.1* in the larval cells, but no expression of either gene name in the adult cells. Searching using either gene name in the complete published adult dataset [49] also did not return any results, suggesting the gene is either not annotated in that dataset, annotated with a different name, or else not expressed. Detection of *gastrin*-derived peptides in our adult peptidomics samples (S2 Table) suggests that *gastrin* is indeed expressed in adult EECs and that incomplete gene annotation may contribute to this discrepancy.

Further, what we annotated as *tac4b* is currently named *si:ch211-131k2.2* in danRer11 (Ensembl ID ENSDARG00000096645). One of the listed protein names for its corresponding Uniprot entry (R4GDR9), however, is Protachykinin-1 isoform X. To evaluate if this gene may be a member of the tachykinin family, we used Ensembl and NCBI to identify orthologs of zebrafish tachykinin genes (*tac1, tac3a, tac3b, tac4, si:ch211-131k2.2*) in common carp (*Cyprinus carpio carpio*), catfish (*Ictalurus punctatus*), rainbow trout (*Oncorhynchus mykiss*), spotted gar (*Lepisosteus oculatus*), elephant shark (*Callorhinchus milii*), pig (*Sus scrofa*), mouse (*Mus musculus*), and human (*Homo sapiens*). We then performed a protein alignment and phylogenetic analysis using MEGA 12 to produce the phylogenetic tree shown in S15 Fig. This tree showed two *tac4* paralogs in all the teleost fish species included in our analysis, but only one *tac4* gene in gar. This pattern is commonly seen due to the third round (3R) of whole genome duplication in the teleost lineage (S6A Fig) [189,190]. Altogether, our analysis suggests that *si:ch211-131k2.2* is a *TAC4* ortholog. While prior studies had suggested zebrafish has two *tac4* paralogs [191], gene annotation has remained obscure. Based on this alignment, we provisionally renamed zebrafish *tac4* as *tac4a* and *si:ch211-131k2.2* as *tac4b*.

To test the generalizability of our EEC subtype annotations, we downloaded scRNA-seq data from [56] available on the Broad Institute Single Cell Portal. Using the same markers described above, we identified an EEC cluster and subsetted and visualized it using UMAP dimensionality reduction to identify EEC subtypes (S3 Fig).

Statistical analysis

For the scRNA-seq analysis, statistical analyses for determination of the cluster-enriched markers were calculated using the FindConservedMarkers() function of the Seurat package in R with a Wilcoxon rank sum test. For all other experiments, statistical analysis was performed using unpaired *t* test, or one-way or two-way analysis of variance (ANOVA) with Tukey's multiple comparisons test with GraphPad Prism. $P < 0.05$ was defined as statistically significant.

EEC FACs sorting

All media were made fresh the morning of the experiment except for the 6M guanidine hydrochloride, for which the same solution was used throughout the two-week collection period. For larval samples, *Tg(neurod1:RFP); TgBAC(cldn15la:GFP)* adult breeders were crossed and embryos were collected at roughly 6 hours post-fertilization and maintained in egg water [192] at approximately 1 embryo/mL density. Unfertilized embryos were removed at 1 dpf and larvae were anesthetized and sorted for presence of both reporters at 5 dpf. The GentleMACS disassociator (Miltenyi, 130-093-235) was placed in the 4°C room to come to temperature for the following day's experiment. At 6 dpf, 75 larvae were euthanized in tricaine and pooled into a single sample with 9 samples being processed concurrently with single positive and non-transgenic controls. Larvae were transferred with minimal media to Gentle MACS C tubes (Miltenyi 130-096-334) containing 2 mL of freshly made disassociation buffer [10 mg/mL cold protease (Sigma, P5380-1G), 10 μM ROCK inhibitor (VWR, S1049), 2.5 mg/mL DNase (Sigma, DN25) in Dulbecco's PBS]. Samples were processed 5 times with protocol C_01 on GentleMACS disassociator with 10 min of incubation on shaker at 4°C after each round. Ten mL of FACS media [0.1% BSA (Fisher, BP1600-100), 10 μM ROCK inhibitor (VWR, S1049) in HBSS (Sigma, H9394-1L)] was added and the contents were poured through a 30 micrometer strainer (Miltenyi 130-098-458) into a fresh 50 mL conical tube. The strainer was washed with an additional 10 mL of FACS media and samples were spun down at 250*g* for 5 min at 4°C, decanted, and resuspended in 1 mL of FACS

media. Samples and controls were then transferred to FACS tubes (Corning, 352052) with 5 μL of 7AAD (Sigma, A9400-1MG) to stain dead cells. Cells were then immediately subjected to FACS at the Duke Cancer Institute Flow Cytometry Shared Resource. *neurod1:RFP; cldn15la:GFP* double positive, 7AAD negative cells from each sample were sorted into a single 1.5 mL LoBind tube (Sigma, Z666505) with 500 μL of 6M guanidine hydrochloride for a total of roughly 30,000 cells. This constituted a single sample for downstream peptidomic analysis. Nontransgenic and single transgenic controls (pools of 50 fish per genotype) were prepared as above and used for gating and compensation. A total of three larval peptidomic samples were collected on separate days, all within 1 week of each other.

For adult samples, procedures were largely similar. Male and female *Tg(neurod1:RFP); TgBAC(cldn15la:GFP)* adults were euthanized at 18 months post fertilization and their intestines were dissected, cut along the longitudinal axis to open the gut, and placed in chilled epithelial buffer with 10 μM ROCK inhibitor (5.6 mM Na_2HPO_4, 8 mM KH_2HPO_4, 96.2 mM NaCl, 1.6 mM KCl, 43.4 mM sucrose, 54.9 mM d-Sorbitol, 10 μM ROCK inhibitor). Samples rocked gently at 4°C for 15 min before being transferred to fresh epithelial buffer with ROCK inhibitor. Samples were shaken at roughly 150 shakes per minute to remove mucus and fecal debris. Intestines were then transferred to gentle MACS C tubes with 2 mL of disassociation buffer, as above. Intestines were digested with 3 rounds of protocol C_01 with 10 min of rocking at 4°C after each round. Identical to above, 10 mL of FACS media was then added, and contents were passed through a 30 micrometer strainer into a fresh 50 mL conical tube. The strainer was then washed with an additional 10 mL of FACS media and the contents were spun down, decanted, and resuspended in 1 mL of FACs media for immediate sorting. *neurod1:RFP; cldn15la:GFP* double positive, 7AAD negative cells from each of nine samples run in parallel were sorted into a single 1.5 mL LoBind tube (Sigma Z666505) with 500 μL of 6M guanidine hydrochloride for a total of roughly 50,000 cells. This constituted a single sample for downstream peptidomic analysis. Nontransgenic and single transgenic controls (1 dissected intestine per genotype) were prepared as above and used for gating and compensation. A total of three adult peptidomic samples were collected on separate days, all within 1 week of each other. All samples were stored at −80°C until they were transferred to the Duke Proteomics and Metabolomics Core Facility and processed for downstream peptidomics as described below.

LC–MS/MS analysis of peptidome samples

Samples sorted into 500 μL of 6M guanidine hydrochloride were subjected to five rounds of 30-s bath sonication (Branson) at 30% power with cooling in between. Samples were then diluted with 500 μL of 0.1% formic acid and centrifuged at 12,000 rpm to remove cellular debris. Lysates were loaded directly onto a 10 mg Oasis HLB SPE cartridge (Waters), washed with 2 × 1 mL of 2% acetonitrile/0.1% formic acid, and eluted with 250 μL of 50% acetonitrile/0.1% formic acid. Eluents were speed-vacuumed to dryness. Samples were then resuspended in 50 μL of 50 mM ammonium bicarbonate with 10 mM dithiolthreitol and heated at 70°C for 30 min, alkylated with 25 mM iodoacetamide for 30 min at room temperature and then spiked with 2 fmol/μL of pre-digested yeast alcohol dehydrogenase (Waters MassPrep standard).

LC/MS/MS was performed using an EvoSep One UPLC coupled to a Thermo Orbitrap Astral high-resolution accurate mass tandem mass spectrometer (Thermo). Briefly, each sample loaded EvoTip was eluted onto a 1.5 μm EvoSep 150 μm ID × 15 cm performance (EvoSep) column using the SPD30 gradient at 55°C. Data collection on the Orbitrap Astral mass spectrometer was performed in a data-dependent acquisition mode of acquisition with a $r = 120{,}000$ (@ m/z 200) full MS scan from m/z 300–2,000 in the OT with a target AGC value of 300% and max accumulation time of 50 ms. Data-dependent MS/MS scans in the Astral were performed on charge states 2–7 from m/z 110–2,000 at a target AGC value of 200% and max accumulation time of 20 ms. HCD collision energy setting of 30% was used for all MS2 scans. Data were imported into PEAKS studio. The software was set to 10 ppm mass accuracy on MS1 and 0.02 Da on MS2 data with no enzyme rules selected. De novo assisted searching was allowed. The MS/MS data was searched against a custom *Danio rerio* database along with a common contaminant/spiked protein database (bovine albumin, bovine casein, yeast ADH, etc.), and an equal number of reversed-sequence "decoys" for false discovery rate determination. Database search

parameters included fixed modification on Cys (carbamidomethyl), variable modification on Met (oxidation), C-terminal amide, and N-terminal acetylation and pyroglutamate. We queried our raw data against a custom database of zebrafish proteins from TrEMBL [193] and Ensembl [97]. To account for genomic polymorphisms within the Ekwill zebrafish strain used here, we supplemented the combined TrEMBL/Ensembl database with protein sequences predicted with customProDB [194] from genetic variants observed in previous RNA sequencing from the same strain [195]. Spectral annotation was set at a maximum 1% peptide false discovery rate based on q-value calculations. Manual verification of spectra of interest were confirmed using targeted extraction/selected ion chromatograms with Skyline (University of Washington, MacCoss Laboratory) [196] including considerations for signal to noise, peak shape, mass errors across isotopologues, and retention time relative to database identifications. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD058654 [197]. In addition, we have shared screenshots of Skyline data for an example of a manually reviewed peptide in S5C and S5D Fig.

Ghrelin acylation search

To identify acylated ghrelin peptides, we performed a PEAKS search on a custom database just containing the F1QKX9 ghrelin sequences from Uniprot and selected octanoylation and decanoylation as variable posttranslational modifications in addition to the modifications in our original search. Spectral matches and chromatogram peaks of the identified peptides are included in the supplemental materials (S5A and S5B Fig).

Zebrafish *secretin* peptide search

As the putative zebrafish *secretin* homolog we identified (*si:zfos-2372e4.1*) did not have a TrEMBL entry, it was not listed in our reference dataset, and, as such, was not identified in our initial annotation of our peptidomics dataset. We first ran a preliminary search of the data with a custom dataset that just contained the protein sequences available on Ensembl for *si:zfos-2372e4.1.* We saw several hits, indicating peptides from that gene were indeed present in our sample. As searching with a drastically smaller reference dataset can inflate the possibility of identifying significant peptides, we added those two *si:zfos-2372e4.1* protein entries to our pre-existing custom zebrafish reference dataset, and re-ran the larger search. In our final annotation (Fig 2F), we only included those peptides that met significance in this more stringent second search.

Protein alignments

Protein sequences for *Danio rerio, Carassius auratus*, *Ictalurus punctatus, Salmo salar, Oncorhynchus mykiss, Aquarana catesbeiana, Gallus gallus, Rattus norvegicus, Mus musculus, Homo sapiens, Cyprinus carpio carpio, Tetraodon nigroviridis, Xenopus tropicalis, Gasterosteus aculeatus aculeatus, Dicentrarchus labrax, Myripristis murdjan, Clupea harengus, Rana temporaria, Oreochromis niloticus* were downloaded from UniProt [193]. When more than one sequence was available, the sequence designated as canonical in UniProt was used. If there were multiple relevant entries, a representative one was used for ease of visualization. Sequences were aligned using MUSCLE [198] and visualized with JalView [199]. Percent identity settings were used to color residues.

For tachykinin analysis (S15 Fig), Molecular Evolutionary Genetics Analysis (MEGA) version 12 [200] was used to align protein sequences (method: MUSCLE [198]) and generate a phylogenetic tree shown in S15 Fig (method: Neighbor-Joining [201] with bootstrapping).

Supporting information

S1 Data. Underlying data for quantitative panels.
(XLSX)

S1 Table. EEC scRNA-seq markers.
(XLSX)

S2 Table. Zebrafish EEC peptidomics.
(XLSX)

S3 Table. *SECRETIN* synteny genome coordinates.
(XLSX)

S4 Table. Reagent sequences.
(XLSX)

S1 Fig. Creation of joint larval and adult intestinal secretory cell scRNA-seq dataset. (A) scRNA-seq data of the adult zebrafish intestine from [49] was clustered and evaluated for expression of secretory cell markers. Clusters showing strong expression of these markers and intestinal epithelial markers but minimal pancreatic endocrine cell markers were selected for subsequent analysis and are outlined with red boxes. **(B)** We similarly processed scRNA-seq data of the larval zebrafish intestine from [48] and selected clusters outlined in red for subsequent analysis. **(C)** UMAP of the joint adult and larval dataset generated by integrating and re-clustering cells identified in panels A and B. Clusters circled with a dashed line were identified as EECs and secretory progenitors based on expression shown in **(D)** and were selected for subsequent integration and re-clustering to form the dataset shown in Fig 1.
(TIF)

S2 Fig. Additional characterization of joint adult and larval secretory progenitor and EEC scRNA-seq dataset. (A) Expression dotplot showing secretory progenitor markers are enriched in clusters 7–10. **(B)** UMAP colored by dataset of origin for each cell with adult cells in green and larval cells in blue. **(C)** Stacked bar plot showing the percentage of cells in each EEC cluster from the adult, larval, or combined dataset. **(D)** Heatmap showing the z-scored expression of the 10 most highly enriched markers for each EEC cluster. **(E)** Heatmap showing the z-scored expression of genes annotated as receptors by [202] that were significantly enriched in both larval and adult cells in each cluster. **(F)** Heatmap showing the z-scored expression of genes annotated as ligands by [202] that were significantly enriched in both larval and adult cells in each cluster. **(G)** Table of major differences between the samples that generated the larval and adult scRNA-seq datasets. Underlying data can be found in S1 Data.
(TIF)

S3 Fig. Clustering of EEC subtypes in additional scRNA-seq dataset. (A) scRNA-seq data of the larval zebrafish digestive tract from [56] was clustered and evaluated for expression of enteroendocrine cell (EEC) markers. The cluster showing strong expression of these markers and intestinal epithelial markers, but minimal pancreatic endocrine cell markers was selected for subsequent analysis and is outlined with a red box on the dotplot and red circle on the UMAP. **(B)** Those selected EECs were then re-clustered to identify EEC subtypes by examining EEC marker expression. **(C)** Seurat Feature Plot with simultaneous visualization of *ghrl* (red) and *pyyb* (blue) expression using the blend function, where overlapping expression would be shown in pink according to the relative expression scale shown. These plots show that, while both genes are expressed in the same cluster in this analysis, they label distinct populations of cells within that cluster. Underlying data can be found in S1 Data.
(TIF)

S4 Fig. Zebrafish EEC peptide atlas. In all panels, the primary amino acid sequence of the gene of interest is shown at the top in black with any identified missense variants indicated in red above. Blue horizontal lines below the amino acid sequence represent the unique peptides detected in our study with small black vertical lines denoting the stop and start of each peptide. Abutting peptides share a single vertical line to represent the stop of the preceding peptide and the start of

the subsequent one. Different colored squares represent various posttranslational modification detected. Shading labels regions aligning to Uniprot-annotated peptides in humans. In cases where multiple peptides are known to be processed from the same sequence, dashed lines indicate different cleavage sites. Below the detected peptides is a multispecies protein alignment where amino acids are color-coded based on their percent identity match across all the reported species with darker coloring indicating a more conserved residue. Human processed peptide annotations taken from Uniprot are labeled below with horizontal lines. The color of these lines corresponds to the color of the shading of the aligned zebrafish amino acids above. Note that many of the peptide cleavage sites occur at dibasic residues (i.e., RR/KR/KK), consistent with cleavage by prohormone convertase enzymes [148]. More detailed information about the peptides shown in this figure are available in S2 Table. **(A)** Zebrafish *adcyap1a*-derived peptides detected in EECs. Alignment of the primary amino acid sequence of Pituitary adenylate cyclase-activating polypeptide protein includes zebrafish (Uniprot Q98TU3), trout (Uniprot A0A8C7QBD4), goldfish (Uniprot A0A6P6RC36), catfish (Uniprot Q90XZ4), chicken (Uniprot Q58FG9), mouse (Uniprot O70176), rat (Uniprot A6KFB2), human (Uniprot P18509). **(B)** Zebrafish *gcga*- and *gcgb*-derived peptides detected in EECs. As *gcga* and *gcgb* are both orthologs of human *GCG*, they are reported together. Peptides aligning to multiple isoforms of *gcga* were detected. Alignment of the primary amino acid sequence of Pro-glucagon protein includes zebrafish *gcga* isoform 2 (Uniprot F1RD10), zebrafish *gcga* isoform 1 (Uniprot A0A0R4IS85), seabass *gcga* (Uniprot A0A8C4IA77), soldierfish *gcga* (Uniprot A0A667XLC6), herring *gcga* (Uniprot A0A6P3WDK7), zebrafish *gcgb* (Uniprot B0R1C3), seabass *gcgb* (Uniprot A0A8P4K489), soldierfish *gcgb* (Uniprot A0A668AYA2), herring *gcgb* (Uniprot A0A6P3W7W6), mouse *Gcg* (Uniprot A2AS86), rat *Gcg* (Uniprot A6HLV7), human (Uniprot P01275). Of note, the arginine 36 residue in GLP-1 is commonly amidated [144,145], as seen in the *gcgb*-derived peptide reported here. **(C)** Zebrafish *calca*-derived peptides detected in EECs. Alignment of the primary amino acid sequence of Calcitonin protein includes zebrafish (Uniprot F1QIK6), catfish (Uniprot A0A2D0RV9), goldfish (Uniprot A0A6P6MFE7), salmon (Uniprot A0A1S3KMJ8), chicken (Uniprot P07660), mouse (Uniprot P70160), rat (Uniprot P01257), and human (Uniprot P01258). **(D)** Zebrafish *ccka*- and *gast*-derived peptides detected in EECs. As Cholecystokinin (CCK) and Gastrin are closely related, they were aligned together, but are displayed separately for ease of visualization. Of note, the zebrafish *gastrin* gene is currently named as *CR556712.1* or *LOC100536965*, but due to evidence from published phylogenetic studies [186–188] and protein sequence alignment, we refer to it as *gastrin* (see Methods). Alignment of the primary amino acid sequence of CCK and Gastrin includes zebrafish *ccka* (Uniprot E9QEB3), carp *ccka* (Uniprot A0A9R1SKH2), tetraodon *ccka* (Uniprot Q8AXP6), xenopus *cck* (Uniprot A0A803J912), chicken *cck* (Uniprot Q9PU41), mouse *Cck* (Uniprot P09240), human *CCK* (Uniprot P06307), zebrafish *gast* (Uniprot A0A8M6Z182), carp *gast* (Uniprot A0A8C1B8C9), tetraodon *gast* (Uniprot Q8AXP5), stickleback *gast* (Ensembl ENSGACT00000024275.2), xenopus *gast* (Uniprot F6W277), chicken *gast* (Ensembl ENSGALT00000043859.2), mouse *Gast* (Uniprot P48757), human *GAST* (Uniprot P01350). **(E)** Zebrafish *galn*-derived peptides detected in EECs. Alignment of the primary amino acid sequence of Galanin includes zebrafish (Uniprot E7EZ53), goldfish (Uniprot Q7ZT91), catfish (Uniprot A0A2D0RH49), salmon (Uniprot A0A1S3KNE4), trout (Uniprot A0A060WZ82), chicken (Uniprot A0A8V0ZZ93), mouse (Uniprot P47212), rat (Uniprot A6HYK9), human (Uniprot P22466). **(F)** Zebrafish *gip*-derived peptides detected in EECs. Alignment of the primary amino acid sequence of Gastric inhibitory polypeptide includes zebrafish (Uniprot A1DPK4), trout (Uniprot A0A060W6J5), salmon (Uniprot A0A1S3NHD2), chicken (Uniprot A0A8V1AAS2), bullfrog (Uniprot A0A2G9RZ64), mouse (Uniprot P48756), rat (Uniprot A6HIC4), human (Uniprot P09681). **(G)** Zebrafish *insl5a*- and *insl5b*-derived peptides detected in EECs. As *insl5a* and *insl5b* are both orthologs of human *INSL5*, they are reported together. Alignment of the primary amino acid sequence of Insulin-like peptide 5 includes zebrafish *insl5a* (Uniprot Q2VT44), goldfish *insl5a* (Uniprot 0A6P6KC48), carp *insl5a* (Uniprot A0A8C1HD46), zebrafish *insl5b* (Uniprot A0ZYT5), goldfish *insl5b* (Uniprot A0A6P6JU13), carp *insl5b* (Uniprot A0A9J8CMM8), xenopus *insl5* (Uniprot A0A6I8SQK1), mouse *Insl5* (Uniprot Q9WUG6), human *INSL5* (Uniprot Q9Y5Q6). **(H)** Zebrafish *mlnl*-derived peptides detected in EECs. Alignment of the primary amino acid sequence of Promotilin includes zebrafish (Uniprot E9QFU1), trout (Uniprot A0A060WHA4), chicken (Uniprot A0A8V1AED3), rat (Uniprot A8IRI0), human (Uniprot P12872). **(I)** Zebrafish *nmbb*-derived peptides

detected in EECs. Alignment of the primary amino acid sequence of Neuromedin B includes zebrafish (Uniprot B3DFU2), goldfish (Uniprot A0A6P6PMY7), trout (Uniprot A0A8K9XPV7), chicken (Uniprot A0A8V0ZQC9), mouse (Uniprot Q9CR53), rat (Uniprot A6JCD9), human (Uniprot P08949). **(J)** Zebrafish *pdyn*-derived peptides detected in EECs. Alignment of the primary amino acid sequence of Proenkephalin B includes zebrafish (Uniprot Q6JT77), goldfish (Uniprot A0A6P6NSK2), catfish (Uniprot E3TFZ0), salmon (Uniprot B5X928), trout (Uniprot A0A060WJX8), chicken (Uniprot A0A8E7KML8), mouse (Uniprot O35852), rat (Uniprot F1M7S3), human (Uniprot P01213). **(K)** Zebrafish *penka*-derived peptides detected in EECs. Alignment of the primary amino acid sequence of Proenkephalin A includes zebrafish (Uniprot A8E7S2), goldfish (Uniprot A0A6P6MMG9), trout (Uniprot A0A8C7NZE4), salmon (Uniprot B5X739), chicken (Uniprot E1C652), mouse (Uniprot P22005), rat (Uniprot A6JFM8), human (Uniprot P01210). **(L)** Zebrafish *pyyb*-derived peptides detected in EECs. Alignment of the primary amino acid sequence of Peptide YY (PYY) includes zebrafish (Uniprot E7F0L6), goldfish (Uniprot A0A6P-6PFJ3), catfish (Uniprot W5UQI6), trout (Uniprot A0A060XNM3), salmon (Uniprot A0A1S3S6S3), chicken (Uniprot A0A0S-3UPI6), bullfrog (Uniprot A0A2G9P7P3), mouse (Uniprot H3BK86), rat (Uniprot F1LSR6), human (Uniprot P10082). **(M)** Zebrafish *sst2*- and *sst1.2*-derived peptides detected in EECs. As both *sst2* and *sst1.2* are orthologs of human *SST*, they are reported together. *sst2* is considered to be part of SST family 4 and *sst1.2* is considered to be part of SST family 3 [203]. Alignment of the primary amino acid sequence of Somatostatin includes zebrafish *sst2* (Uniprot Q9DDE4), goldfish *sst2* (Uniprot A0A6P6P309), carp *sst2* (Uniprot A0A8C1E251), zebrafish *sst1.2* (Uniprot E7FFY9), carp *sst1.2* (Uniprot A0A8C1BWP9), catfish *sst1.2* (Uniprot W5UDS0), frog *sst* (RefSeq XP_040205526.1), chicken *sst* (Uniprot P33094), mouse *Sst* (Uniprot P60041), human *SST* (Uniprot P61278). **(N)** Zebrafish *tac3a*-derived peptides detected in EECs. Alignment of the primary amino acid sequence of Tachykinin 3 includes zebrafish (Uniprot E9QE01), goldfish (Uniprot T2CZA2), catfish (Uniprot I4IY93), trout (Uniprot A0A8C7RBE8), salmon (Uniprot I4IY94), mouse (Uniprot P55099), rat (Uniprot A6HQX0), human (Uniprot Q9UHF0). **(O)** Zebrafish *vipb*-derived peptides detected in EECs. Peptides aligning to multiple isoforms of *vipb* were detected. Alignment of the primary amino acid sequence of Vasoactive intestinal peptides protein include zebrafish isoform 1 (Uniprot A0A8M9Q6Y9), zebrafish isoform 2 (Uniprot B0LF71), goldfish (Uniprot A0A6P-6NBR0), catfish (Uniprot A0A2D0RS13), salmon (Uniprot A0A1S3PEV), trout (Uniprot A0A060X5U6), chicken (Uniprot P48143), mouse (Uniprot A0A0R4J003), rat (Uniprot A0A090AX27), human (Uniprot P01282). **(P)** Zebrafish *ins*-derived peptides detected in EECs. Alignment of the primary amino acid sequences of Insulin protein include zebrafish (Uniprot B2GSI0), stickleback (Ensembl ENSGACT00000066960.1), tilapia (Uniprot I3IUZ1), herring (Ensembl ENSCHAT00020053570.1), xenopus (Uniprot F6QRS1), chicken (Uniprot P67970), mouse (Uniprot P01326), rat (Uniprot P01323), human (Uniprot P01308). **(Q)** Zebrafish *pcsk1nl*-derived peptides detected in EECs. **(R)** Zebrafish *pcsk1*-derived peptides detected in EECs. **(S)** Zebrafish *pcsk2*-derived peptides detected in EECs. **(T)** Zebrafish *scg2a*-derived peptides detected in EECs. **(U)** Zebrafish *scg2b*-derived peptides detected in EECs. **(V)** Zebrafish *scg3*-derived peptides detected in EECs. **(W)** Zebrafish *scg5*-derived peptides detected in EECs. **(X)** Zebrafish *scgn*-derived peptides detected in EECs.
(PDF)

S5 Fig. Example raw peptidomic data of acylated ghrelin and manually curated peptides. (A) Fragmentation spectrum identifying octanolylated ghrelin peptide. **(B)** Fragmentation spectrum identifying decanolylated ghrelin peptide. **(C)** Chromatogram data showing overlapping peaks for double, triple, and quadruple charged YGGFLRKFGPK peptide manually identified from the Pdyn protein. **(D)** Chromatogram data showing overlapping peaks for isotopologues M, M + 1, and M + 2 of the +3 charged peptide. Peak areas are quantified below and mass error for each is shown. **(E)** Expression in the joint larval and adult EEC dataset of hormone genes commonly enriched in the pancreatic islet. Underlying data can be found in S1 Data.
(TIF)

S6 Fig. Synteny of *SECRETIN* loci across vertebrates. (A) Two rounds of whole genome duplication (labeled 1R and 2R) preceded the divergence of vertebrates [204–207]. The most ancient split within vertebrates is between the jawless

Cyclostomi, such as lamprey, and the remaining Gnathostomata. The jawed vertebrates can be divided into those with primarily cartilaginous (Chondrichthyes) or bony (Osteichthyes) skeletons. A final round of whole genome duplication (labeled 3R) occurred in teleost fishes [189,190]. **(B)** The arrangement of genes along human chromosome 11, elephant shark scaffold 17, gar Chromosome 21, and zebrafish chromosome 25 are shown with genes ordered by ascending start location. Genes were identified by searching the ~250,000 base pairs upstream and downstream of human *SCT* and cross-checking with orthologs in zebrafish. Any gap equal to or greater than 0.5 Mega bases is annotated. The location of *SCT* and its direct neighbors are traced between each species with lines. Exact location of each gene is shown in S3 Table.
(TIF)

S7 Fig. Creation of *ghrl:QF2* and *neurog3:QF2* lines. (A) Schematic representation of cloning approach to generating *neurog3:QF2* and *ghrl:QF2* reporters. **(B)** mVISTA [178,179] alignment of the *ghrl* locus in six closely related *Danio* species. *Danio rerio ghrl* annotation is shown at the top and is used as the reference. Dashed lines mark the highly conserved 647 base pair region upstream of the *ghrl* transcriptional start site that was cloned for the *ghrl:QF2* reporter. **(C)** Schematic of where representative cross sections were taken to evaluate the *ghrl:QF2* reporter in adult intestines in panels E and F. **(D)** Schematic of Swiss roll preparation and sectioning of full-length intestine from *ghrl:QF2* reporter adults in panel G. **(E)** Representative image of a proximal intestinal section showing *ghrl*+; *neurod1*+ cells with classical EEC morphology. **(F)** Representative image of *ghrl*+; *neurod1*+ cell with classical EEC morphology in a distal intestinal section. **(G)** Section of Swiss-rolled adult *ghrl:QF2* reporter intestine where the proximal gut is closest to the center. Several *ghrl*+; *neurod1*+ cells are highlighted. **(H)** Quantification of *ghrl*+ cells overlap with the pan-EEC reporter *neurod1:RFP* across larval development. Each dot represents an individual fish. **(I)** A representative image of a 5 dpf *ghrl:QF2* fish with a highlighted example of a *ghrl*+; *neurod1*- cell. **(J)** Staining of 6 dpf larvae with anti-Ghrelin primary antibody and appropriate secondary antibody co-labels cells with *ghrl:QF2* reporter in the pancreatic islet and intestine. While staining is not very strong, it is clearly stronger than negative controls run with **(K)** non-specific anti-IgG primary antibody and appropriate secondary antibody or **(L)** anti-Ghrelin primary antibody and no secondary antibody. **(M)** *ghrl:QF2* reporter co-labels with hybridization chain reaction (HCR) probes targeting *ghrl* mRNA in the islet and intestine. HCR probes targeting *cubn,* a gene enriched in the lysosome-rich enterocytes (LREs) in the distal intestine, serve as a positive control for the HCR technique. **(N)** The *ghrl:QF2* reporter and *sst2:DsRed* reporter both label EECs in the intestine. While they overlap in many cells, some cells are only *ghrl* reporter+ and some are only *sst2* reporter+, particularly in the distal intestine. **(O)** UMAP of secretory progenitors and EECs from larval and adult zebrafish shown in Fig 1. **(P)** Seurat Feature Plot with simultaneous visualization of *ghrl* (red) and *sst2* (blue) expression using the blend function, where overlapping expression is shown in pink according to the relative expression scale shown. **(Q)** Seurat Feature Plot of only *ghrl* expression, derived from the blend-plot described in P. **(R)** Seurat Feature Plot of only *sst2* expression, derived from the blend-plot described in P. **(S)** A representative image of a 6 dpf *neurog3:QF2* reporter fish with a highlighted example of both a *neurog3*+; *neurod1*- and *neurog3*+; *neurod1*+ cell. **(T)** Quantification of *neurog3*:*QF2* reporter overlap with *neurod1:RFP* reporter across larval development. Each dot represents an individual fish. All scale bars are 100 μm. Underlying data can be found in S1 Data.
(TIF)

S8 Fig. EEC subtype reporter imaging across larval development. Representative images showing subtype distribution and overlap with pan-EEC *neurod1* reporter at 3, 4, and 5 dpf for **(A)** *neurog3* reporter, **(B)** *ghrl* reporter, **(C)** anti-PYY antibody, **(D)** *gcga* reporter, **(E)** anti-CCK antibody, **(F)** *trpa1b* reporter. Scale bars are 100 μm.
(TIF)

S9 Fig. EEC subtype reporter cell counts across larval development. The percent of *neurod1*+ cells that label with the subtype reporter and the raw counts of subtype numbers per each equal-length quarter of the gut is shown for 3, 4,

5, and 6 days post fertilization (dpf) fish in the **(A)** *neurog3* reporter, **(B)** *ghrl* reporter, **(C)** anti-PYY antibody, **(D)** *gcga* reporter, **(E)** anti-CCK antibody, and **(F)** *trpa1b* reporter. Each dot represents a fish. Underlying data can be found in S1 Data.
(TIF)

S10 Fig. Regional patterns in EEC subtypes. (A) Cartoon of larval zebrafish intestine. Region-specific genes along the intestinal epithelium were determined using the scRNA-seq dataset from [109]. Horizontal bars below the cartoon represent the approximated anteroposterior location of a given regional cell type and are marked with the panel letter that examines expression of genes specific to that region. Module scores in our combined larval and adult scRNA-seq dataset of the expression of the 25 most enriched genes in the **(B)** pharynx, **(C)** anterior enterocytes, **(D)** ileocytes, **(E)** anterior lysosome-rich enterocytes (LREs), **(F)** posterior LREs, and **(G)** cloaca are shown. Underlying data can be found in S1 Data.
(TIF)

S11 Fig. Ablation validation and alterations in regional distribution. (A) Schematic of *neurog3*+ ablation combined with *neurog3*+, *neurod1*+ cell labeling to determine *neurog3*+ ablation efficiency. **(B)** Total and regional counts of *neurog3*+ cells in *neurog3*+ ablation. **(C)** Total and regional counts of *neurod1*+ cells in control, *neurog3*+, and *neurod1*+ ablation. **(D)** Total and regional counts of *trpa1b*+ cells in control, *neurog3*+, and *neurod1*+ ablation. **(E)** Total EEC and subtype counts in control and *neurog3*+ ablated fish when qrt4 cells are excluded. Counts that include qrt4 are shown in Fig 4C–4H. **(F)** Total EEC and subtype counts in control and *ghrl*-ablated fish when qrt4 cells are excluded. Counts that include qrt4 are shown in Fig 5B–5F. Underlying data can be found in S1 Data.
(TIF)

S12 Fig. *neurog3:Cre* induced ablation leads to similar loss of *neurod1*+ cells as *neurog3:QF2; QUAS:Cre.* (A) Schematic of *neurog3:Cre* induced *neurog3*+ cell ablation. **(B)** Total counts and regional counts of *neurod1*+ cells in the *neurog3:Cre*-ablated and control fish. Each dot represents a 6-day post-fertilization fish. Statistical significance was calculated by unpaired *t* test for total cell numbers and by two-way ANOVA for regional analysis. Significance annotations are as follows: ns ($p > 0.05$), * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$), **** ($p < 0.0001$). Underlying data can be found in S1 Data.
(TIF)

S13 Fig. Loss of *ghrl* or *neurog3* does not impact *neurod1*+ EEC numbers. (A) Schematic of the *ghrl* locus with arrowheads marking the sites targeted with CRISPR guide RNAs. **(B)** Genotyping and control gels run with 1 kb+ NEB ladder and 3 *ghrl* mutant and 3 *ghrl* wildtype samples. Amplification with genotyping primers (red arrows) across the 804 base pair region results in a 300 base pair band due to deletion in mutants (lanes 1–3). Wildtype samples (lanes 4–6) do not amplify due to the highly repetitive nature of intron 2, but all samples amplified with control primers amplifying at a non-affected locus (*mttp* gene). **(C)** Total counts and regional counts of *neurod1*+ cells in *ghrl* wildtype (heterozygous) and *ghrl* mutant (homozygous) fish. **(D)** Total counts of *neurod1*+ cells in *neurog3* wildtype (homozygous) and *neurog3* mutant (homozygous) fish. Each dot represents a 5–6-day post-fertilization fish. Statistical significance was calculated by unpaired *t* test for total cell numbers and by two-way ANOVA for regional analysis. Underlying data can be found in S1 Data.
(TIF)

S14 Fig. Impact of ablation on adult growth and survival. (A) Schematic of adult experiments demonstrating when survival counts and standard length measurements were taken. **(B)** Schematic of standard length measurements performed at the endpoint of each experiment. **(C)** *neurod1*-ablated versus sibling control survival and growth data from two independent experiments, each carried out over 6 weeks. Survival curves are significantly different by Mantel-Cox log-rank test

(*P* = 0.0006). **(D)** *neurog3*-ablated versus sibling controls followed for 6 weeks showed no significant difference in survival or growth. **(E)** *neurog3*-ablated versus sibling controls followed for 12 weeks again showed no significant differences in survival or growth. **(F)** *ghrl*-ablated versus sibling controls followed for 6 weeks showed no significant difference in survival or growth. **(G)** *ghrl*-ablated versus sibling controls followed for 12 weeks again showed no significant differences in survival or growth. Survival analyses were performed on the pooled results across the number of tanks specified in each panel. Statistical significance was calculated by Mantel-Cox log-rank test for survival curves and by unpaired *t* test for standard length. Underlying data can be found in S1 Data.
(TIF)

S15 Fig. Phylogeny of tachykinin family. The analytical procedure encompassed 33 amino acid sequences of tachykinin proteins identified from zebrafish (*Danio* rerio), common carp (*Cyprinus carpio carpio*), catfish (*Ictalurus punctatus*), rainbow trout (*Oncorhynchus mykiss*), spotted gar (*Lepisosteus oculatus*), elephant shark (*Callorhinchus milii*), pig (*Sus scrofa*), mouse (*Mus musculus*), and human (*Homo sapiens*). The gene name and accession number are shown for each branch. Proposed tachykinin family names are shown for those genes without one. The evolutionary history was inferred using the Neighbor-Joining method [201]. The optimal tree with the sum of branch length = 11.694 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches [208]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method [209] and are in the units of the number of amino acid substitutions per site. The pairwise deletion option was applied to all ambiguous positions for each sequence pair resulting in a final data set comprising 226 positions. Evolutionary analyses were conducted in MEGA12 [200,210] utilizing up to 4 parallel computing threads.
(TIF)

S16 Fig. Resolution of combined larval and adult scRNA-seq dataset. Clustering for the scRNA-seq dataset presented in Fig 1 is shown here at resolution 0.1–0.8. The resolution of 0.4 was used for this study.
(TIF)

S1 Raw Images. Raw images of gels used in publication.
(TIF)

Acknowledgments

The authors acknowledge Duke University School of Medicine Zebrafish Core Facilities (Z-Core) including Lawrence Frauen and Elizabeth Reich for their help in maintaining the zebrafish lines used in this study. We also want to thank Kristen Lounsbury, Anthony Turner, and Samantha Lampman for excellent fish care at Michigan State University, as well as Newton PenkoffLidbeck for contributing to this work. The authors also thank Mark Cronan, Olov Andersson, and Michel Bagnat for sharing reagents.

Author contributions

Conceptualization: Margaret Morash, Julia Ganz, Fiona M. Gribble, Frank Reimann, Rodger A. Liddle, John F. Rawls.

Data curation: Margaret Morash, Erik J. Soderblom.

Formal analysis: Margaret Morash, Richard G. Kay.

Funding acquisition: Margaret Morash, Julia Ganz, John F. Rawls.

Investigation: Margaret Morash, Grace H. MacLean, Jia Wen, Peyton J. Moore, Mujahid Ali Shah.

Methodology: Richard G. Kay, Erik J. Soderblom, Colin R. Lickwar.

Supervision: John F. Rawls.

Writing – original draft: Margaret Morash.

Writing – review & editing: Margaret Morash, Richard G. Kay, Erik J. Soderblom, Jia Wen, Peyton J. Moore, Colin R. Lickwar, Mujahid Ali Shah , Julia Ganz , Fiona M. Gribble, Rodger A. Liddle, John F. Rawls.

References

1. Gribble FM, Reimann F. Enteroendocrine cells: chemosensors in the intestinal epithelium. Annu Rev Physiol. 2016;78:277–99. https://doi.org/10.1146/annurev-physiol-021115-105439 PMID: 26442437
2. Annunziata R, Andrikou C, Perillo M, Cuomo C, Arnone MI. Development and evolution of gut structures: from molecules to function. Cell Tissue Res [Internet]. 2019 [cited 2024 Dec 28];377(3):445–58. Available from: https://pubmed.ncbi.nlm.nih.gov/31446445/
3. Hartenstein V, Takashima S, Hartenstein P, Asanad S, Asanad K. bHLH proneural genes as cell fate determinants of entero-endocrine cells, an evolutionarily conserved lineage sharing a common root with sensory neurons. Dev Biol [Internet]. 2017 [cited 2024 Dec 28];431(1):36–47. Available from: https://pubmed.ncbi.nlm.nih.gov/28751238/
4. Smith CL, Mayorova TD. Insights into the evolution of digestive systems from studies of *Trichoplax adhaerens*. Cell Tissue Res. 2019;377(3):353–67. https://doi.org/10.1007/s00441-019-03057-z PMID: 31270610
5. Mayorova TD, Hammar K, Winters CA, Reese TS, Smith CL. The ventral epithelium of trichoplax adhaerens deploys in distinct patterns cells that secrete digestive enzymes, mucus or diverse neuropeptides. Biol Open [Internet]. 2019 [cited 2023 Nov 15];8(8). Available from: https:///pmc/articles/PMC6737977/
6. Bany Bakar R, Reimann F, Gribble FM. The intestine as an endocrine organ and the role of gut hormones in metabolic regulation. Nat Rev Gastroenterol Hepatol [Internet]. 2023 [cited 2023 Nov 29];20(12). Available from: https://pubmed.ncbi.nlm.nih.gov/37626258/
7. Lu VB, Gribble FM, Reimann F. Nutrient-induced cellular mechanisms of gut hormone secretion. Nutrients [Internet]. 2021 [cited 2022 Mar 15];13(3):1–36. Available from: https://pubmed.ncbi.nlm.nih.gov/33803183/
8. Gariepy CE, Dickinson CJ. Translation and posttranslational processing of gastrointestinal peptides. In: Barrett KE, Ghishan FK, Merchant JL, Said HM, Wood JD, Johnson LR, editors. Physiology of the gastrointestinal tract. 4th ed. Elsevier Academic Press; 2006. p. 31–62.
9. Yang Q, Bermingham NA, Finegold MJ, Zoghbi HY. Requirement of Math1 for secretory cell lineage commitment in the mouse intestine. Science. 2001;294(5549):2155–8. https://doi.org/10.1126/science.1065718 PMID: 11739954
10. Jenny M, Uhl C, Roche C, Duluc I, Guillermin V, Guillemot F, et al. Neurogenin3 is differentially required for endocrine cell fate specification in the intestinal and gastric epithelium. EMBO J. 2002;21(23):6338–47. https://doi.org/10.1093/emboj/cdf649 PMID: 12456641
11. López-Díaz L, Jain RN, Keeley TM, VanDussen KL, Brunkan CS, Gumucio DL, et al. Intestinal neurogenin 3 directs differentiation of a bipotential secretory progenitor to endocrine cell rather than goblet cell fate. Dev Biol. 2007;309(2):298–305. https://doi.org/10.1016/j.ydbio.2007.07.014
12. Schonhoff SE, Giel-Moloney M, Leiter AB. Neurogenin 3-expressing progenitor cells in the gastrointestinal tract differentiate into both endocrine and non-endocrine cell types. Dev Biol. 2004;270(2):443–54.
13. Bjerknes M, Cheng H. Neurogenin 3 and the enteroendocrine cell lineage in the adult mouse small intestinal epithelium. Dev Biol. 2006;300(2):722–35.
14. Mellitzer G, Beucher A, Lobstein V, Michel P, Robine S, Kedinger M, et al. Loss of enteroendocrine cells in mice alters lipid absorption and glucose homeostasis and impairs postnatal survival. J Clin Invest. 2010;120(5):1708–21. https://doi.org/10.1172/JCI40794 PMID: 20364088
15. Ratineau C, Petry MW, Mutoh H, Leiter AB. Cyclin D1 represses the basic helix-loop-helix transcription factor, BETA2/NeuroD. J Biol Chem. 2002;277(11):8847–53. https://doi.org/10.1074/jbc.M110747200 PMID: 11788592
16. Li HJ, Ray SK, Pan N, Haigh J, Fritzsch B, Leiter AB. Intestinal Neurod1 expression impairs paneth cell differentiation and promotes enteroendocrine lineage specification. Sci Rep. 2019;9(1):19489. https://doi.org/10.1038/s41598-019-55292-7 PMID: 31862906
17. Naya FJ, Huang HP, Qiu Y, Mutoh H, DeMayo FJ, Leiter AB, et al. Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/neuroD-deficient mice. Genes Dev. 1997;11(18):2323–34. https://doi.org/10.1101/gad.11.18.2323 PMID: 9308961
18. Huang HP, Liu M, El-Hodiri HM, Chu K, Jamrich M, Tsai MJ. Regulation of the pancreatic islet-specific gene BETA2 (neuroD) by neurogenin 3. Mol Cell Biol. 2000;20(9):3292–307. https://doi.org/10.1128/MCB.20.9.3292-3307.2000 PMID: 10757813
19. Gehart H, van Es JH, Hamer K, Beumer J, Kretzschmar K, Dekkers JF, et al. Identification of enteroendocrine regulators by real-time single-cell differentiation mapping. Cell. 2019;176(5):1158–73.e16. https://doi.org/10.1016/j.cell.2019.01.014
20. Singh PNP, Gu W, Madha S, Lynch AW, Cejas P, He R, et al. Transcription factor dynamics, oscillation, and functions in human enteroendocrine cell differentiation. bioRxiv [Internet]. 2024 [cited 2024 Jan 15];2024.01.09.574746. Available from: https://www.biorxiv.org/content/10.1101/2024.01.09.574746v1
21. Beumer J, Artegiani B, Post Y, Reimann F, Gribble F, Nguyen TN, et al. Enteroendocrine cells switch hormone expression along the crypt-to-villus BMP signalling gradient. Nat Cell Biol. 2018;20(8):909–16. https://doi.org/10.1038/s41556-018-0143-y PMID: 30038251
22. Arnes L, Hill JT, Gross S, Magnuson MA, Sussel L. Ghrelin expression in the mouse pancreas defines a unique multipotent progenitor population. PLoS One [Internet]. 2012 [cited 2022 Apr 16];7(12). Available from: https://pmc/articles/PMC3520898/

23. Haber AL, Biton M, Rogel N, Herbst RH, Shekhar K, Smillie C, et al. A single-cell survey of the small intestinal epithelium. Nature. 2017;551(7680):333–9. https://doi.org/10.1038/nature24489 PMID: 29144463
24. Aliluev A, Tritschler S, Sterr M, Oppenländer L, Hinterdobler J, Greisle T, et al. Diet-induced alterations in ISC function underlies obesity and pre-diabetes. Nat Metab, in revision [Internet]. 2021 [cited 2021 Sep 21];3(9):1202–16. Available from: https://www.nature.com/articles/s42255-021-00458-9
25. Piccand J, Vagne C, Blot F, Meunier A, Beucher A, Strasser P, et al. Rfx6 promotes the differentiation of peptide-secreting enteroendocrine cells while repressing genetic programs controlling serotonin production. Mol Metab [Internet]. 2019 [cited 2024 Nov 2];29:24–39. Available from: https://pubmed.ncbi.nlm.nih.gov/31668390/
26. Ye L, Mueller O, Bagwell J, Bagnat M, Liddle RA, Rawls JF. High fat diet induces microbiota-dependent silencing of enteroendocrine cells. Elife [Internet]. 2019 [cited 2021 Oct 17];8. Available from: https://elifesciences.org/articles/48479
27. Ye L, Bae M, Cassilly CD, Jabba SV, Thorpe DW, Martin AM, et al. Enteroendocrine cells sense bacterial tryptophan catabolites to activate enteric and vagal neuronal pathways. Cell Host Microbe [Internet]. 2021 [cited 2021 Oct 17];29(2):179-196.e9. Available from: http://www.cell.com/article/S1931312820306338/fulltext
28. Flasse LC, Stern DG, Pirson JL, Manfroid I, Peers B, Voz ML. The bHLH transcription factor Ascl1a is essential for the specification of the intestinal secretory cells and mediates Notch signaling in the zebrafish intestine. Dev Biol. 2013;376(2):187–97. https://doi.org/10.1016/j.ydbio.2013.01.014
29. Flasse LC, Pirson JL, Stern DG, Von Berg V, Manfroid I, Peers B, et al. Ascl1b and Neurod1, instead of Neurog3, control pancreatic endocrine cell fate in zebrafish. BMC Biol. 2013;11:78. https://doi.org/10.1186/1741-7007-11-78 PMID: 23835295
30. Lavergne A, Tarifeño-Saldivia E, Pirson J, Reuter AS, Flasse L, Manfroid I, et al. Pancreatic and intestinal endocrine cells in zebrafish share common transcriptomic signatures and regulatory programmes. BMC Biol [Internet]. 2020 [cited 2023 Mar 14];18(1). Available from: https://pubmed.ncbi.nlm.nih.gov/32867764/
31. Reuter AS, Stern D, Bernard A, Goossens C, Lavergne A, Flasse L, et al. Identification of an evolutionarily conserved domain in Neurod1 favouring enteroendocrine versus goblet cell fate. PLoS Genet. 2022;18(3):e1010109. https://doi.org/10.1371/journal.pgen.1010109 PMID: 35286299
32. Ferguson M, Foley E. Microbial recognition regulates intestinal epithelial growth in homeostasis and disease. FEBS J. 2022;289(13):3666–91. https://doi.org/10.1111/febs.15910 PMID: 33977656
33. Ye L, Rawls JF. Microbial influences on gut development and gut-brain communication. Development. 2021;148(21):dev194936. https://doi.org/10.1242/dev.194936 PMID: 34758081
34. Gerbe F, Sidot E, Smyth DJ, Ohmoto M, Matsumoto I, Dardalhon V, et al. Intestinal epithelial tuft cells initiate type 2 mucosal immunity to helminth parasites. Nature. 2016;529(7585):226–30. https://doi.org/10.1038/nature16527 PMID: 26762460
35. Szczepanski AP, Tsuboyama N, Watanabe J, Hashizume R, Zhao Z, Wang L. POU2AF2/C11orf53 functions as a coactivator of POU2F3 by maintaining chromatin accessibility and enhancer activity. Sci Adv. 2022;8(40):eabq2403. https://doi.org/10.1126/sciadv.abq2403 PMID: 36197978
36. Wu XS, He X-Y, Ipsaro JJ, Huang Y-H, Preall JB, Ng D, et al. OCA-T1 and OCA-T2 are coactivators of POU2F3 in the tuft cell lineage. Nature. 2022;607(7917):169–75. https://doi.org/10.1038/s41586-022-04842-7 PMID: 35576971
37. Schumacher MA, Hsieh JJ, Liu CY, Appel KL, Waddell A, Almohazey D, et al. Sprouty2 limits intestinal tuft and goblet cell numbers through GSK3β-mediated restriction of epithelial IL-33. Nat Commun. 2021;12(1):836. https://doi.org/10.1038/s41467-021-21113-7 PMID: 33547321
38. Rios D, Wood MB, Li J, Chassaing B, Gewirtz AT, Williams IR. Antigen sampling by intestinal M cells is the principal pathway initiating mucosal IgA production to commensal enteric bacteria. Mucosal Immunol. 2016;9(4):907–16. https://doi.org/10.1038/mi.2015.121 PMID: 26601902
39. Pellegrinet L, Rodilla V, Liu Z, Chen S, Koch U, Espinosa L, et al. Dll1- and Dll4-mediated notch signaling are required for homeostasis of intestinal stem cells. Gastroenterology [Internet]. 2011 [cited 2024 Nov 9];140(4):1230–40.e7. Available from: http://www.gastrojournal.org/article/S0016508511000394/fulltext
40. Crosnier C, Vargesson N, Gschmeissner S, Ariza-McNaughton L, Morrison A, Lewis J. Delta-Notch signalling controls commitment to a secretory fate in the zebrafish intestine. Development. 2005;132(5):1093–104. https://doi.org/10.1242/dev.01644 PMID: 15689380
41. Gracz AD, Samsa LA, Fordham MJ, Trotier DC, Zwarycz B, Lo YH, et al. Sox4 promotes Atoh1-independent intestinal secretory differentiation toward tuft and enteroendocrine fates. Gastroenterology. 2018;155(5):1508–23.e10.
42. Roach G, Heath Wallace R, Cameron A, Emrah Ozel R, Hongay CF, Baral R, et al. Loss of ascl1a prevents secretory cell differentiation within the zebrafish intestinal epithelium resulting in a loss of distal intestinal motility. Dev Biol [Internet]. 2013 [cited 2024 Nov 9];376(2):171–86. Available from: https://pubmed.ncbi.nlm.nih.gov/23353550/
43. Thambyrajah R, Ucanok D, Jalali M, Hough Y, Wilkinson RN, McMahon K, et al. A gene trap transposon eliminates haematopoietic expression of zebrafish Gfi1aa, but does not interfere with haematopoiesis. Dev Biol. 2016;417(1):25–39. https://doi.org/10.1016/j.ydbio.2016.07.010 PMID: 27432513
44. Gregorieff A, Stange DE, Kujala P, Begthel H, van den Born M, Korving J, et al. The Ets-domain transcription factor Spdef promotes maturation of goblet and paneth cells in the intestinal epithelium. Gastroenterology [Internet]. 2009 [cited 2024 Dec 29];137(4):1333–45.e3. Available from: http://www.gastrojournal.org/article/S0016508509010014/fulltext
45. Chang SK, Dohrman AF, Basbaum CB, Ho SB, Tsuda T, Toribara NW, et al. Localization of mucin (MUC2 and MUC3) messenger RNA and peptide expression in human normal intestine and colon cancer. Gastroenterology. 1994;107(1):28–36. https://doi.org/10.1016/0016-5085(94)90057-4 PMID: 8020672

46. Chen Y-C, Lu Y-F, Li I-C, Hwang S-PL. Zebrafish Agr2 is required for terminal differentiation of intestinal goblet cells. PLoS One. 2012;7(4):e34408. https://doi.org/10.1371/journal.pone.0034408 PMID: 22514630
47. Park S-W, Zhen G, Verhaeghe C, Nakagami Y, Nguyenvu LT, Barczak AJ, et al. The protein disulfide isomerase AGR2 is essential for production of intestinal mucus. Proc Natl Acad Sci U S A. 2009;106(17):6950–5. https://doi.org/10.1073/pnas.0808722106 PMID: 19359471
48. Wen J, Mercado GP, Volland A, Doden HL, Lickwar CR, Crooks T, et al. Fxr signaling and microbial metabolism of bile salts in the zebrafish intestine. Sci Adv. 2021;7(30):eabg1371. https://doi.org/10.1126/sciadv.abg1371 PMID: 34301599
49. Jones LO, Willms RJ, Xu X, Graham RDV, Eklund M, Shin M, et al. Single-cell resolution of the adult zebrafish intestine under conventional conditions and in response to an acute *Vibrio cholerae* infection. Cell Rep [Internet]. 2023 [cited 2024 Oct 20];42(11). Available from: http://www.cell.com/article/S2211124723014195/fulltext
50. Li HJ, Ray SK, Singh NK, Johnston B, Leiter AB. Basic helix-loop-helix transcription factors and enteroendocrine cell differentiation. Diabetes Obes Metab. 2011;13 Suppl 1(0 1):5–12. https://doi.org/10.1111/j.1463-1326.2011.01438.x PMID: 21824251
51. Burclaff J, Bliton RJ, Breau KA, Ok MT, Gomez-Martinez I, Ranek JS, et al. A proximal-to-distal survey of healthy adult human small intestine and colon epithelium by single-cell transcriptomics. Cell Mol Gastroenterol Hepatol. 2022;13(5):1554–89. https://doi.org/10.1016/j.jcmgh.2022.02.007 PMID: 35176508
52. Hayashi M, Kaye JA, Douglas ER, Joshi NR, Gribble FM, Reimann F, et al. Enteroendocrine cell lineages that differentially control feeding and gut motility. Elife. 2023;12:e78512. https://doi.org/10.7554/eLife.78512 PMID: 36810133
53. Bai L, Sivakumar N, Yu S, Mesgarzadeh S, Ding T, Ly T, et al. Enteroendocrine cell types that drive food reward and aversion. Elife [Internet]. 2022 [cited 2022 Aug 7];11. Available from: https://elifesciences.org/articles/74964
54. Smith CA, O'Flaherty EAA, Guccio N, Punnoose A, Darwish T, Lewis JE, et al. Single-cell transcriptomic atlas of enteroendocrine cells along the murine gastrointestinal tract. PLoS One. 2024;19(10):e0308942. https://doi.org/10.1371/journal.pone.0308942 PMID: 39378212
55. Beumer J, Geurts MH, Geurts V, Andersson-Rolf A, Akkerman N, Völllmy F, et al. Description and functional validation of human enteroendocrine cell sensors. Science. 2024;386(6719):341–8. https://doi.org/10.1126/science.adl1460 PMID: 39418382
56. Willms RJ, Jones LO, Hocking JC, Foley E. A cell atlas of microbe-responsive processes in the zebrafish intestine. Cell Rep [Internet]. 2022 [cited 2025 Apr 3];38(5). Available from: https://www.cell.com/action/showFullText?pii=S2211124722000237
57. Shen WK, Chen SY, Gan ZQ, Zhang YZ, Yue T, Chen MM, et al. AnimalTFDB 4.0: a comprehensive animal transcription factor database updated with variation and expression annotations. Nucleic Acids Res [Internet]. 2023 [cited 2024 Oct 20];51(D1):D39–45. Available from: https://pubmed.ncbi.nlm.nih.gov/36268869/
58. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 2014;32(4):381–6. https://doi.org/10.1038/nbt.2859 PMID: 24658644
59. Hibdon ES, Keeley TM, Merchant JL, Samuelson LC. The bHLH transcription factor ASCL1 promotes differentiation of endocrine cells in the stomach and is regulated by Notch signaling. Am J Physiol Gastrointest Liver Physiol. 2023;325(5):G458–70. https://doi.org/10.1152/ajpgi.00043.2023 PMID: 37698169
60. Kokubu H, Ohtsuka T, Kageyama R. Mash1 is required for neuroendocrine cell development in the glandular stomach. Genes Cells. 2008;13(1):41–51. https://doi.org/10.1111/j.1365-2443.2007.01146.x PMID: 18173746
61. Singh A, Chaube B, Canfrán-Duque A, Zhang X, Price N, Aryal B, et al. Liver-specific suppression of ANGPTL4 improves obesity-associated diabetes and mitigates atherosclerosis in mice. J Clin Investig [Internet]. 2020 [cited 2021 Jul 13]. Available from: http://www.jci.org/articles/view/140989
62. Mavropoulos A, Devos N, Biemar F, Zecchin E, Argenton F, Edlund H, et al. sox4b is a key player of pancreatic alpha cell differentiation in zebrafish. Dev Biol [Internet]. 2005 [cited 2024 Nov 9];285(1):211–23. Available from: https://pubmed.ncbi.nlm.nih.gov/16055112/
63. Soyer J, Flasse L, Raffelsberger W, Beucher A, Orvain C, Peers B, et al. Rfx6 is an Ngn3-dependent winged helix transcription factor required for pancreatic islet cell development. Development [Internet]. 2010 [cited 2024 Nov 9];137(2):203–12. Available from: https://pubmed.ncbi.nlm.nih.gov/20040487/
64. Gierl MS, Karoulias N, Wende H, Strehle M, Birchmeier C. The zinc-finger factor Insm1 (IA-1) is essential for the development of pancreatic beta cells and intestinal endocrine cells. Genes Dev. 2006;20(17):2465–78. https://doi.org/10.1101/gad.381806 PMID: 16951258
65. Desai S, Loomis Z, Pugh-Bernard A, Schrunk J, Doyle MJ, Minic A, et al. Nkx2.2 regulates cell fate choice in the enteroendocrine cell lineages of the intestine. Dev Biol. 2008;313(1):58–66. https://doi.org/10.1016/j.ydbio.2007.09.047 PMID: 18022152
66. Gross S, Garofalo DC, Balderes DA, Mastracci TL, Dias JM, Perlmann T, et al. The novel enterochromaffin marker Lmx1a regulates serotonin biosynthesis in enteroendocrine cell lineages downstream of Nkx2.2. Development. 2016;143(14):2616–28. https://doi.org/10.1242/dev.130682 PMID: 27287799
67. Wang Y-C, Gallego-Arteche E, Iezza G, Yuan X, Matli MR, Choo S-P, et al. Homeodomain transcription factor NKX2.2 functions in immature cells to control enteroendocrine differentiation and is expressed in gastrointestinal neuroendocrine tumors. Endocr Relat Cancer. 2009;16(1):267–79. https://doi.org/10.1677/ERC-08-0127 PMID: 18987169
68. Mutoh H, Ratineau C, Ray S, Leiter AB. Review article: transcriptional events controlling the terminal differentiation of intestinal endocrine cells. Aliment Pharmacol Ther. 2000;14 Suppl 1:170–5. https://doi.org/10.1046/j.1365-2036.2000.014s1170.x PMID: 10807420
69. Fujita Y, Chui JWY, King DS, Zhang T, Seufert J, Pownall S, et al. Pax6 and Pdx1 are required for production of glucose-dependent insulinotropic polypeptide in proglucagon-expressing L cells. Am J Physiol Endocrinol Metab [Internet]. 2008 [cited 2024 Nov 2];295(3). Available from: https://pubmed.ncbi.nlm.nih.gov/18593849/

70. Offield MF, Jetton TL, Labosky PA, Ray M, Stein RW, Magnuson MA, et al. PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. Development. 1996;122(3):983–95. https://doi.org/10.1242/dev.122.3.983 PMID: 8631275

71. Chen C, Fang R, Davis C, Maravelias C, Sibley E. Pdx1 inactivation restricted to the intestinal epithelium in mice alters duodenal gene expression in enterocytes and enteroendocrine cells. Am J Physiol Gastrointest Liver Physiol [Internet]. 2009 [cited 2024 Nov 2];297(6). Available from: https://pubmed.ncbi.nlm.nih.gov/19808654/

72. Yang W, Liu X, He Z, Zhang Y, Tan X, Liu C. Odd skipped-related 2 as a novel mark for labeling the proximal convoluted tubule within the zebrafish kidney. Heliyon. 2024;10(6):e27582. https://doi.org/10.1016/j.heliyon.2024.e27582 PMID: 38496848

73. Cheng L, Chen C-L, Luo P, Tan M, Qiu M, Johnson R, et al. Lmx1b, Pet-1, and Nkx2.2 coordinately specify serotonergic neurotransmitter phenotype. J Neurosci. 2003;23(31):9961–7. https://doi.org/10.1523/JNEUROSCI.23-31-09961.2003 PMID: 14602809

74. Ding Y-Q, Marklund U, Yuan W, Yin J, Wegman L, Ericson J, et al. Lmx1b is essential for the development of serotonergic neurons. Nat Neurosci. 2003;6(9):933–8. https://doi.org/10.1038/nn1104 PMID: 12897786

75. Elmentaite R, Kumasaka N, Roberts K, Fleming A, Dann E, King HW, et al. Cells of the human intestinal tract mapped across space and time. Nature. 2021;597(7875):250–5. https://doi.org/10.1038/s41586-021-03852-1 PMID: 34497389

76. Muller L, Lindberg I. The cell biology of the prohormone convertases PC1 and PC2. Prog Nucleic Acid Res Mol Biol [Internet]. 1999 [cited 2024 Nov 14];63(C):69–108. Available from: https://pubmed.ncbi.nlm.nih.gov/10506829/

77. Mitchell K, Mikwar M, Da Fonte D, Lu C, Tao BB, Peng D, et al. Secretoneurin is a secretogranin-2 derived hormonal peptide in vertebrate neuroendocrine systems. Gen Comp Endocrinol [Internet]. 2020 [cited 2024 Nov 14];299. Available from: https://pubmed.ncbi.nlm.nih.gov/32828813/

78. Mitchell K, Zhang WS, Lu C, Tao B, Chen L, Hu W, et al. Targeted mutation of secretogranin-2 disrupts sexual behavior and reproduction in zebrafish. Proc Natl Acad Sci U S A. 2020;117(23):12772–83. https://doi.org/10.1073/pnas.2002004117 PMID: 32467166

79. Fischer-Colbrie R, Laslop A, Kirchmair R. Secretogranin II: molecular properties, regulation of biosynthesis and processing to the neuropeptide secretoneurin. Prog Neurobiol. 1995;46(1):49–70. https://doi.org/10.1016/0301-0082(94)00060-u PMID: 7568909

80. Taupenot L, Harper KL, O'Connor DT. The chromogranin-secretogranin family. N Engl J Med. 2003;348(12):1134–49. https://doi.org/10.1056/NEJMra021405 PMID: 12646671

81. Kojima M, Hosoda H, Date Y, Nakazato M, Matsuo H, Kangawa K. Ghrelin is a growth-hormone-releasing acylated peptide from stomach. Nature. 1999;402(6762):656–60. https://doi.org/10.1038/45230 PMID: 10604470

82. Zhang JV, Ren PG, Avsian-Kretchmer O, Luo CW, Rauch R, Klein C, et al. Obestatin, a peptide encoded by the ghrelin gene, opposes ghrelin's effects on food intake. Science [Internet]. 2005 [cited 2024 Oct 27];310(5750):996–9. Available from: https://pubmed.ncbi.nlm.nih.gov/16284174/

83. Kojima M, Kangawa K. Ghrelin: structure and function. Physiol Rev. 2005;85(2):495–522. https://doi.org/10.1152/physrev.00012.2004 PMID: 15788704

84. Kaiya H, Miyazato M, Kangawa K. Recent advances in the phylogenetic study of ghrelin. Peptides (NY) [Internet]. 2011 [cited 2023 Nov 16];32:2155–74. Available from: http://www.ensembl.org/index.html

85. Kojima M, Ida T, Sato T. Structure of mammalian and nonmammalian ghrelins. Vitam Horm [Internet]. 2008 [cited 2024 Nov 9];77:31–46. Available from: https://pubmed.ncbi.nlm.nih.gov/17983852/

86. Miura T, Maruyama K, Kaiya H, Miyazato M, Kangawa K, Uchiyama M, et al. Purification and properties of ghrelin from the intestine of the goldfish, *Carassius auratus*. Peptides. 2009;30(4):758–65. https://doi.org/10.1016/j.peptides.2008.12.016 PMID: 19150635

87. Yuan X, Cai W, Liang X-F, Su H, Yuan Y, Li A, et al. Obestatin partially suppresses ghrelin stimulation of appetite in "high-responders" grass carp, *Ctenopharyngodon idellus*. Comp Biochem Physiol A Mol Integr Physiol. 2015;184:144–9. https://doi.org/10.1016/j.cbpa.2015.02.019 PMID: 25737031

88. Cowan E, Burch KJ, Green BD, Grieve DJ. Obestatin as a key regulator of metabolism and cardiovascular function with emerging therapeutic potential for diabetes. Br J Pharmacol. 2016;173(14):2165–81. https://doi.org/10.1111/bph.13502 PMID: 27111465

89. Unniappan S, Lin X, Cervini L, Rivier J, Kaiya H, Kangawa K, et al. Goldfish ghrelin: molecular characterization of the complementary deoxyribonucleic acid, partial gene structure and evidence for its stimulatory role in food intake. Endocrinology. 2002;143(10):4143–6. https://doi.org/10.1210/en.2002-220644 PMID: 12239128

90. Wren AM, Seal LJ, Cohen MA, Brynes AE, Frost GS, Murphy KG, et al. Ghrelin enhances appetite and increases food intake in humans. J Clin Endocrinol Metab. 2001;86(12):5992. https://doi.org/10.1210/jcem.86.12.8111 PMID: 11739476

91. Sur A, Wang Y, Capar P, Margolin G, Prochaska MK, Farrell JA. Single-cell analysis of shared signatures and transcriptional diversity during zebrafish development. Dev Cell [Internet]. 2023 [cited 2024 Nov 4];58(24):3028–47.e12. Available from: https://pubmed.ncbi.nlm.nih.gov/37995681/

92. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol. 2022;40(7):1023–5. https://doi.org/10.1038/s41587-021-01156-3 PMID: 34980915

93. Nielsen H, Engelbrecht J, Brunak S, von Heijne G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. 1997;10(1):1–6. https://doi.org/10.1093/protein/10.1.1 PMID: 9051728

94. Hughes LC, Ortí G, Huang Y, Sun Y, Baldwin CC, Thompson AW, et al. Comprehensive phylogeny of ray-finned fishes (*Actinopterygii*) based on transcriptomic and genomic data. Proc Natl Acad Sci U S A. 2018;115(24):6249–54. https://doi.org/10.1073/pnas.1719358115 PMID: 29760103

95. Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, et al. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. Nat Genet. 2016;48(4):427–37. https://doi.org/10.1038/ng.3526 PMID: 26950095

96. Venkatesh B, Lee AP, Ravi V, Maurya AK, Lian MM, Swann JB, et al. Elephant shark genome provides unique insights into gnathostome evolution. Nature. 2014;505(7482):174–9. https://doi.org/10.1038/nature12826 PMID: 24402279

97. Harrison PW, Amode MR, Austine-Orimoloye O, Azov AG, Barba M, Barnes I, et al. Ensembl 2024. Nucleic Acids Res. 2024;52(D1):D891–9. https://doi.org/10.1093/nar/gkad1049 PMID: 37953337

98. Brazeau MD, Friedman M. The origin and early phylogenetic history of jawed vertebrates. Nature [Internet]. 2015 [cited 2025 Apr 3];520(7548):490. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC4648279/

99. Chandra R, Samsa LA, Vigna SR, Liddle RA. Pseudopod-like basal cell processes in intestinal cholecystokinin cells. Cell Tissue Res. 2010;341(2):289–97. https://doi.org/10.1007/s00441-010-0997-1 PMID: 20582553

100. Chandra R, Hiniker A, Kuo YM, Nussbaum RL, Liddle RA. α-Synuclein in gut endocrine cells and its implications for Parkinson's disease. JCI Insight [Internet]. 2017 [cited 2024 Nov 10];2(12). Available from: https://pubmed.ncbi.nlm.nih.gov/28614796

101. Zecchin E, Filippi A, Biemar F, Tiso N, Pauls S, Ellertsdottir E, et al. Distinct delta and jagged genes control sequential segregation of pancreatic cell types from precursor pools in zebrafish. Dev Biol. 2007;301(1):192–204. https://doi.org/10.1016/j.ydbio.2006.09.041 PMID: 17059815

102. Pan YA, Choy M, Prober DA, Schier AF. Robo2 determines subtype-specific axonal projections of trigeminal sensory neurons. Development. 2012;139(3):591–600. https://doi.org/10.1242/dev.076588 PMID: 22190641

103. Kwan KM, Fujimoto E, Grabher C, Mangum BD, Hardy ME, Campbell DS, et al. The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. Dev Dyn. 2007;236(11):3088–99. https://doi.org/10.1002/dvdy.21343 PMID: 17937395

104. Burgess J, Burrows JT, Sadhak R, Chiang S, Weiss A, D'Amata C, et al. An optimized QF-binary expression system for use in zebrafish. Dev Biol. 2020;465(2):144–56. https://doi.org/10.1016/j.ydbio.2020.07.007 PMID: 32697972

105. Ghosh A, Halpern ME. Transcriptional regulation using the Q system in transgenic zebrafish. Methods Cell Biol. 2016;135:205–18.

106. Chu L, Terasaki M, Mattsson CL, Teinturier R, Charbord J, Dirice E, et al. In vivo drug discovery for increasing incretin-expressing cells identifies DYRK inhibitors that reinforce the enteroendocrine system. Cell Chem Biol. 2022;29(9):1368–80. https://doi.org/10.1016/j.chembiol.2022.06.007

107. Billing LJ, Larraufie P, Lewis J, Leiter A, Li J, Lam B, et al. Single cell transcriptomic profiling of large intestinal enteroendocrine cells in mice—identification of selective stimuli for insulin-like peptide-5 and glucagon-like peptide-1 co-expressing cells. Mol Metab. 2019;29:158–69. https://doi.org/10.1016/j.molmet.2019.09.001 PMID: 31668387

108. Diwakarla S, Fothergill LJ, Fakhry J, Callaghan B, Furness JB. Heterogeneity of enterochromaffin cells within the gastrointestinal tract. Neurogastroenterol Motil. 2017;29(6):10.1111/nmo.13101. https://doi.org/10.1111/nmo.13101 PMID: 28485065

109. Childers L, Park J, Wang S, Liu R, Barry R, Watts SA, et al. Protein absorption in the zebrafish gut is regulated by interactions between lysosome rich enterocytes and the microbiome. 2025.

110. Schwitzgebel VM, Scheel DW, Conners JR, Kalamaras J, Lee JE, Anderson DJ, et al. Expression of neurogenin3 reveals an islet cell precursor population in the pancreas. Development. 2000;127(16):3533–42. https://doi.org/10.1242/dev.127.16.3533 PMID: 10903178

111. Wang Y, Chen K, Yao Q, Zheng X, Yang Z. Phylogenetic analysis of zebrafish basic helix-loop-helix transcription factors. J Mol Evol. 2009;68(6):629–40. https://doi.org/10.1007/s00239-009-9232-7 PMID: 19449054

112. Beucher A, Gjernes E, Collin C, Courtney M, Meunier A, Collombat P, et al. The homeodomain-containing transcription factors Arx and Pax4 control enteroendocrine subtype specification in mice. PLoS One [Internet]. 2012 [cited 2021 Jun 15];7(5). Available from: https://pubmed.ncbi.nlm.nih.gov/22570716/

113. Terry NA, Walp ER, Lee RA, Kaestner KH, May CL. Impaired enteroendocrine development in intestinal-specific Islet1 mouse mutants causes impaired glucose homeostasis. Am J Physiol Gastrointest Liver Physiol [Internet]. 2014 [cited 2023 Mar 21];307(10):G979–91. Available from: https://pubmed.ncbi.nlm.nih.gov/25214396/

114. Wang Q, Elghazi L, Martin S, Martins I, Srinivasan RS, Geng X, et al. Ghrelin is a novel target of Pax4 in endocrine progenitors of the pancreas and duodenum. Dev Dyn. 2008;237(1):51–61. https://doi.org/10.1002/dvdy.21379 PMID: 18058910

115. Jiménez S, Blot F, Meunier A, Schreiber V, Giethlen C, Ghimire S, et al. Unraveling enteroendocrine cell lineage dynamics and associated gene regulatory networks during intestinal development. bioRxiv [Internet]. 2025 [cited 2025 Oct 12];2025.04.11.648332. Available from: https://www.biorxiv.org/content/10.1101/2025.04.11.648332v1

116. Yamada W, Kaji T, Onishi S, Nakame K, Yamada K, Kawano T, et al. Ghrelin improves intestinal mucosal atrophy during parenteral nutrition: an experimental study. J Pediatr Surg. 2016;51(12):2039–43. https://doi.org/10.1016/j.jpedsurg.2016.09.035 PMID: 27832865

117. Yu H, Xu G, Fan X. The effect of ghrelin on cell proliferation in small intestinal IEC-6 cells. Biomed Pharmacother. 2013;67(3):235–9.

118. Parichy DM, Elizondo MR, Mills MG, Gordon TN, Engeszer RE. Normal table of post-embryonic Zebrafish development: staging by externally visible anatomy of the living fish. Dev Dyn [Internet]. 2009 [cited 2025 Jul 2];238(12):2975. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC3030279/

119. Alsudayri A, Perelman S, Brewer M, Chura A, McDevitt M, Drerup C, et al. Gut microbiota regulate maturation and mitochondrial function of the nutrient-sensing enteroendocrine cell. Development [Internet]. 2024 [cited 2025 Oct 14];151(8). Available from: https://pubmed.ncbi.nlm.nih.gov/38577841/

120. Fawkner-Corbett D, Antanaviciute A, Parikh K, Jagielowicz M, Gerós AS, Gupta T. Spatiotemporal analysis of human intestinal development at single-cell resolution. Cell. 2021;184(3):810–26.e23.

121. Hickey JW, Becker WR, Nevins SA, Horning A, Perez AE, Zhu C, et al. Organization of the human intestine at single-cell resolution. Nature. 2023;619(7970):572–84. https://doi.org/10.1038/s41586-023-05915-x PMID: 37468586

122. Gerlai R. From genes to behavior: the question of evolutionary conservation and the role of ethology in the analysis of the Zebrafish. Front Neuroanat [Internet]. 2021 [cited 2024 Nov 3];15:809967. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC8675880/

123. Shainer I, Michel M, Marquart GD, Bhandiwad AA, Zmora N, Ben-Moshe Livne Z, et al. Agouti-related protein 2 is a new player in the teleost stress response system. Curr Biol. 2019;29(12):2009–19.e7. https://doi.org/10.1016/j.cub.2019.05.021 PMID: 31178320

124. Jeong I, Kim E, Kim S, Kim HK, Lee DW, Seong JY, et al. mRNA expression and metabolic regulation of npy and agrp1/2 in the zebrafish brain. Neurosci Lett. 2018;668:73–9.

125. Bayliss WM, Starling EH. On the uniformity of the pancreatic mechanism in vertebrata. J Physiol [Internet]. 1903 [cited 2025 Mar 30];29(2):174–80. Available from: https://pubmed.ncbi.nlm.nih.gov/16992659/

126. Cardoso JCR, Félix RC, Trindade M, Power DM. Fish genomes provide novel insights into the evolution of vertebrate secretin receptors and their ligand. Gen Comp Endocrinol [Internet]. 2014 [cited 2025 Mar 31];209:82–92. Available from: https://pubmed.ncbi.nlm.nih.gov/24906176/

127. Tam JKV, Lee LTO, Jin J, Chow BKC. Molecular evolution of GPCRS: secretin/secretin receptors. J Mol Endocrinol [Internet]. 2014 [cited 2025 Mar 31];52(3):T1–14. Available from: https://jme.bioscientifica.com/view/journals/jme/52/3/T1.xml

128. Hwang J-I, Moon MJ, Park S, Kim D-K, Cho EB, Ha N, et al. Expansion of secretin-like G protein-coupled receptors and their peptide ligands via local duplications before and after two rounds of whole-genome duplication. Mol Biol Evol. 2013;30(5):1119–30. https://doi.org/10.1093/molbev/mst031 PMID: 23427277

129. Cardoso JC, Vieira FA, Gomes AS, Power DM. The serendipitous origin of chordate secretin peptide family members. BMC Evol Biol [Internet]. 2010 [cited 2025 Mar 9];10(1). Available from: https://pubmed.ncbi.nlm.nih.gov/20459630/

130. Cardoso JCR, Pinto VC, Vieira FA, Clark MS, Power DM. Evolution of secretin family GPCR members in the metazoa. BMC Evol Biol [Internet]. 2006 [cited 2025 Mar 9];6. Available from: https://pubmed.ncbi.nlm.nih.gov/17166275/

131. Cardoso JCR, Garcia MG, Power DM. Tracing the origins of the pituitary adenylate-cyclase activating polypeptide (PACAP). Front Neurosci [Internet]. 2020 [cited 2025 Mar 9];14. Available from: https://pubmed.ncbi.nlm.nih.gov/32508559/

132. Ng SYL, Chow BKC, Kasamatsu J, Kasahara M, Lee LTO. Agnathan VIP, PACAP and their receptors: ancestral origins of today's highly diversified forms. PLoS One [Internet]. 2012 [cited 2025 Mar 9];7(9). Available from: https://pubmed.ncbi.nlm.nih.gov/22957100/

133. Erwin DH, Laflamme M, Tweedt SM, Sperling EA, Pisani D, Peterson KJ. The Cambrian conundrum: early divergence and later ecological success in the early history of animals. Science. 2011;334(6059):1091–7. https://doi.org/10.1126/science.1206375 PMID: 22116879

134. Kumar S, Hedges SB. A molecular timescale for vertebrate evolution. Nature. 1998;392(6679):917–20. https://doi.org/10.1038/31927 PMID: 9582070

135. Sherwood NM, Krueckl SL, McRory JE. The origin and function of the pituitary adenylate cyclase-activating polypeptide (PACAP)/glucagon superfamily. Endocr Rev. 2000;21(6):619–70. https://doi.org/10.1210/edrv.21.6.0414 PMID: 11133067

136. Wang Y, Huang G, Li J, Meng F, He X, Leung FC. Characterization of chicken secretin (SCT) and secretin receptor (SCTR) genes: a novel secretin-like peptide (SCT-LP) and secretin encoded in a single gene. Mol Cell Endocrinol. 2012;348(1):270–80.

137. Foreman RE, Miedzybrodzka EL, Eiríksson FF, Thorsteinsdóttir M, Bannon C, Wheller R, et al. Optimized LC-MS/MS method for the detection of ppCCK(21-44): a surrogate to monitor human cholecystokinin secretion. J Proteome Res. 2023;22(9):2950–8. https://doi.org/10.1021/acs.jproteome.3c00272 PMID: 37591880

138. Kay RG, Challis BG, Casey RT, Roberts GP, Meek CL, Reimann F, et al. Peptidomic analysis of endogenous plasma peptides from patients with pancreatic neuroendocrine tumours. Rapid Commun Mass Spectrom. 2018;32(16):1414–24. https://doi.org/10.1002/rcm.8183 PMID: 29857350

139. Foreman RE, George AL, Reimann F, Gribble FM, Kay RG. Peptidomics: a review of clinical applications and methodologies. J Proteome Res. 2021;20(8):3782–97. https://doi.org/10.1021/acs.jproteome.1c00295 PMID: 34270237

140. Lund PK, Goodman RH, Habener JF. Pancreatic pre-proglucagons are encoded by two separate mRNAs. J Biol Chem. 1981;256(13):6515–8. https://doi.org/10.1016/s0021-9258(19)69015-0 PMID: 6165720

141. Raufman JP, Singh L, Singh G, Eng J. Truncated glucagon-like peptide-1 interacts with exendin receptors on dispersed acini from guinea pig pancreas. Identification of a mammalian analogue of the reptilian peptide exendin-4. J Biol Chem. 1992;267(30):21432–7. PMID: 1328231

142. Egan JM, Clocquet AR, Elahi D. The insulinotropic effect of acute exendin-4 administered to humans: comparison of nondiabetic state to type 2 diabetes. J Clin Endocrinol Metab [Internet]. 2002 [cited 2024 Dec 5];87(3):1282–90. Available from: https://pubmed.ncbi.nlm.nih.gov/11889200/

143. Eng J, Kleinman WA, Singh L, Singh G, Raufman JP. Isolation and characterization of exendin-4, an exendin-3 analogue, from Heloderma suspectum venom. Further evidence for an exendin receptor on dispersed acini from guinea pig pancreas. J Biol Chem. 1992;267(11):7402–5. https://doi.org/10.1016/s0021-9258(18)42531-8 PMID: 1313797

144. Galvin SG, Larraufie P, Kay RG, Pitt H, Bernard E, McGavigan AK, et al. Peptidomics of enteroendocrine cells and characterisation of potential effects of a novel preprogastrin derived-peptide on glucose tolerance in lean mice. Peptides (NY) [Internet]. 2021 [cited 2021 Dec 6];140. Available from: https://pubmed.ncbi.nlm.nih.gov/33744371/

145. Orskov C, Buhl T, Rabenhøj L, Kofod H, Holst JJ. Carboxypeptidase-B-like processing of the C-terminus of glucagon-like peptide-2 in pig and human small intestine. FEBS Lett. 1989;247(2):193–6. https://doi.org/10.1016/0014-5793(89)81332-8 PMID: 2714431

146. Cardoso JCR, Félix RC, Costa C, Palma PFS, Canário AVM, Power DM. Evolution of the glucagon-like system across fish. Gen Comp Endocrinol. 2018;264:113–30.

147. Zhou L, Irwin DM. Fish proglucagon genes have differing coding potential. Comp Biochem Physiol B Biochem Mol Biol. 2004;137(2):255–64. https://doi.org/10.1016/j.cbpc.2003.11.009 PMID: 14990222

148. Mbikay M, Seidah NG, editors. Proprotein convertases, vol. 768. 1st ed. Totowa, NJ: Humana Press; 2011.

149. Mirabeau O, Perlas E, Severini C, Audero E, Gascuel O, Possenti R, et al. Identification of novel peptide hormones in the human proteome by hidden Markov model screening. Genome Res [Internet]. 2007 [cited 2024 Dec 8];17(3):320. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC1800923/

150. Reyes-Alcaraz A, Lee Y-N, Son GH, Kim NH, Kim D-K, Yun S, et al. Development of spexin-based human galanin receptor type II-specific agonists with increased stability in serum and anxiolytic effect in mice. Sci Rep. 2016;6:21453. https://doi.org/10.1038/srep21453 PMID: 26907960

151. Sur A, Wang Y, Capar P, Margolin G, Farrell JA. Single-cell analysis of shared signatures and transcriptional diversity during zebrafish development. bioRxiv [Internet]. 2023 [cited 2024 Mar 14];2023.03.20.533545. Available from: https://www.biorxiv.org/content/10.1101/2023.03.20.533545v3

152. Argenton F, Zecchin E, Bortolussi M. Early appearance of pancreatic hormone-expressing cells in the zebrafish embryo. Mech Dev. 1999;87(1–2):217–21.

153. Ahi EP, Brunel M, Tsakoumis E, Chen J, Schmitz M. Appetite regulating genes in zebrafish gut; a gene expression study. PLoS One. 2022;17(7):e0255201. https://doi.org/10.1371/journal.pone.0255201 PMID: 35853004

154. Roberts GP, Larraufie P, Richards P, Kay RG, Galvin SG, Miedzybrodzka EL, et al. Comparison of human and murine enteroendocrine cells by transcriptomic and peptidomic profiling. Diabetes. 2019;68(5):1062–72. https://doi.org/10.2337/db18-0883 PMID: 30733330

155. Wallace KN, Pack M. Unique and conserved aspects of gut development in zebrafish. Dev Biol. 2003;255(1):12–29.

156. Wallace KN, Akhter S, Smith EM, Lorent K, Pack M. Intestinal growth and differentiation in zebrafish. Mech Dev. 2005;122(2):157–73. https://doi.org/10.1016/j.mod.2004.10.009 PMID: 15652704

157. Lickwar CR, Camp JG, Weiser M, Cocchiaro JL, Kingsley DM, Furey TS, et al. Genomic dissection of conserved transcriptional regulation in intestinal epithelial cells. PLoS Biol [Internet]. 2017 [cited 2021 Nov 6];15(8). Available from: https://pubmed.ncbi.nlm.nih.gov/28850571/

158. Li HJ, Ray SK, Kucukural A, Gradwohl G, Leiter AB. Reduced neurog3 gene dosage shifts enteroendocrine progenitor towards goblet cell lineage in the mouse intestine. Cell Mol Gastroenterol Hepatol. 2021;11(2):433–48. https://doi.org/10.1016/j.jcmgh.2020.08.006 PMID: 32822913

159. Gradwohl G, Dierich A, LeMeur M, Guillemot F. neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas. Proc Natl Acad Sci U S A [Internet]. 2000 [cited 2024 Nov 3];97(4):1607–11. Available from: https://pubmed.ncbi.nlm.nih.gov/10677506/

160. Gu G, Dubauskaite J, Melton DA. Direct evidence for the pancreatic lineage: NGN3+ cells are islet progenitors and are distinct from duct progenitors. Development [Internet]. 2002 [cited 2024 Nov 3];129(10):2447–57. Available from: https://pubmed.ncbi.nlm.nih.gov/11973276/

161. Du A, McCracken KW, Walp ER, Terry NA, Klein TJ, Han A. Arx is required for normal enteroendocrine cell development in mice and humans. Dev Biol. 2012;365(1):175–88.

162. Cox HM. Peptide YY: a neuroendocrine neighbor of note. Peptides (NY) [Internet]. 2007 [cited 2022 Mar 16];28(2):345–51. Available from: https://pubmed.ncbi.nlm.nih.gov/17194503/

163. Psichas A, Reimann F, Gribble FM. Gut chemosensing mechanisms. J Clin Invest. 2015;125(3):908–17. https://doi.org/10.1172/JCI76309 PMID: 25664852

164. Kubrak O, Koyama T, Ahrentløv N, Jensen L, Malita A, Naseem MT, et al. The gut hormone Allatostatin C/Somatostatin regulates food intake and metabolic homeostasis under nutrient stress. Nat Commun [Internet]. 2022 [cited 2022 Mar 16];13(1). Available from: https://pubmed.ncbi.nlm.nih.gov/35121731/

165. Modvig IM, Andersen DB, Grunddal K V., Kuhre RE, Martinussen C, Christiansen CB, et al. Secretin release after Roux-en-Y gastric bypass reveals a population of glucose-sensitive S cells in distal small intestine. Int J Obes (Lond) [Internet]. 2020 [cited 2022 Mar 17];44(9):1859–71. Available from: https://pubmed.ncbi.nlm.nih.gov/32015474/

166. Adriaenssens A, Lam BYH, Billing L, Skeffington K, Sewing S, Reimann F, et al. A transcriptome-led exploration of molecular mechanisms regulating somatostatin-producing D-cells in the gastric epithelium. Endocrinology. 2015;156(11):3924–36. https://doi.org/10.1210/en.2015-1301 PMID: 26241122

167. Alvers AL, Ryan S, Scherz PJ, Huisken J, Bagnat M. Single continuous lumen formation in the zebrafish gut is mediated by smoothened-dependent tissue remodeling. Development [Internet]. 2014 [cited 2024 Oct 20];141(5):1110–9. Available from: https://pubmed.ncbi.nlm.nih.gov/24504339/

168. McGraw HF, Snelson CD, Prendergast A, Suli A, Raible DW. Postembryonic neuronal addition in zebrafish dorsal root ganglia is regulated by Notch signaling. Neural Dev [Internet]. 2012 [cited 2024 Oct 20];7(1). Available from: https://pubmed.ncbi.nlm.nih.gov/22738203/

169. Trapani JG, Obholzer N, Mo W, Brockerhoff SE, Nicolson T. Synaptojanin1 is required for temporal fidelity of synaptic transmission in hair cells. PLoS Genet [Internet]. 2009 [cited 2024 Nov 16];5(5). Available from: https://pubmed.ncbi.nlm.nih.gov/19424431/

170. Ye L, Robertson MA, Hesselson D, Stainier DYR, Anderson RM. Glucagon is essential for alpha cell transdifferentiation and beta cell neogenesis. Development. 2015;142(8):1407–17. https://doi.org/10.1242/dev.117911 PMID: 25852199

171. Kikuchi K, Holdway JE, Werdich AA, Anderson RM, Fang Y, Egnaczyk GF, et al. Primary contribution to zebrafish heart regeneration by gata4(+) cardiomyocytes. Nature [Internet]. 2010 [cited 2024 Dec 28];464(7288):601–5. Available from: https://pubmed.ncbi.nlm.nih.gov/20336144/

172. Wopat S, Bagwell J, Sumigray KD, Dickson AL, Huitema LFA, Poss KD, et al. Spine patterning is guided by segmentation of the notochord sheath. Cell Rep [Internet]. 2018 [cited 2024 Dec 1];22(8):2026–38. Available from: https://pubmed.ncbi.nlm.nih.gov/29466731/

173. Subedi A, Macurak M, Gee ST, Monge E, Goll MG, Potter CJ, et al. Adoption of the Q transcriptional regulatory system for zebrafish transgenesis. Methods [Internet]. 2014 [cited 2024 Nov 17];66(3):433–40. Available from: https://pubmed.ncbi.nlm.nih.gov/23792917/

174. Berger J, Currie PD. 503unc, a small and muscle-specific zebrafish promoter. Genesis [Internet]. 2013 [cited 2024 Nov 17];51(6):443–7. Available from: https://pubmed.ncbi.nlm.nih.gov/23444339/

175. Cronan MR, Beerman RW, Rosenberg AF, Saelens JW, Johnson MG, Oehlers SH, et al. Macrophage epithelial reprogramming underlies mycobacterial granuloma formation and promotes infection. Immunity [Internet]. 2016 [cited 2024 Nov 17];45(4):861–76. Available from: https://pubmed.ncbi.nlm.nih.gov/27760340/

176. Moreno-Mateos MA, Vejnar CE, Beaudoin J-D, Fernandez JP, Mis EK, Khokha MK, et al. CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nat Methods. 2015;12(10):982–8. https://doi.org/10.1038/nmeth.3543 PMID: 26322839

177. Wilson MH, Rajan S, Danoff A, White RJ, Hensley MR, Quinlivan VH, et al. A point mutation decouples the lipid transfer activities of microsomal triglyceride transfer protein. PLoS Genet. 2020;16(8):e1008941. https://doi.org/10.1371/journal.pgen.1008941 PMID: 32760060

178. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res [Internet]. 2004 [cited 2024 Nov 12];32(Web Server issue). Available from: https://pubmed.ncbi.nlm.nih.gov/15215394/

179. Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, et al. VISTA : visualizing global DNA sequence alignments of arbitrary length. Bioinformatics. 2000;16(11):1046–7. https://doi.org/10.1093/bioinformatics/16.11.1046 PMID: 11159318

180. Alliance of genome resources version: 8.1.0. neurog3sa211 [Internet]. 2025. Available from: https://www.alliancegenome.org/allele/ZFIN:ZDB-ALT-110318-1

181. Buschlab. neurog3 [Internet]. 2025. Available from: https://zmp.buschlab.org/gene/ENSDARG00000016951#sa211

182. Munjal A, Hannezo E, Tsai TYC, Mitchison TJ, Megason SG. Extracellular hyaluronate pressure shaped by cellular tethers drives tissue morphogenesis. Cell [Internet]. 2021 [cited 2025 Jul 20];184(26):6313–25.e18. Available from: https://www.sciencedirect.com/science/article/pii/S0092867421013738?via%3Dihub

183. Choi HMT, Schwarzkopf M, Fornace ME, Acharya A, Artavanis G, Stegmaier J, et al. Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. Development. 2018;145(12):dev165753. https://doi.org/10.1242/dev.165753 PMID: 29945988

184. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9(7):671–5. https://doi.org/10.1038/nmeth.2089 PMID: 22930834

185. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol. 2024;42(2):293–304. https://doi.org/10.1038/s41587-023-01767-y PMID: 37231261

186. Kurokawa T, Suzuki T, Hashimoto H. Identification of gastrin and multiple cholecystokinin genes in teleost. Peptides. 2003;24(2):227–35. https://doi.org/10.1016/s0196-9781(03)00034-2 PMID: 12668207

187. Dann J, Grützner F. Evolution of the jawed vertebrate (Gnathostomata) stomach through gene repertoire loss: findings from *Agastric* species. bioRxiv [Internet]. 2024 [cited 2024 Dec 22];2024.12.02.626498. Available from: https://www.biorxiv.org/content/10.1101/2024.12.02.626498v2

188. Dupré D, Tostivint H. Evolution of the gastrin–cholecystokinin gene family revealed by synteny analysis. Gen Comp Endocrinol. 2014;195:164–73.

189. Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, et al. The medaka draft genome and insights into vertebrate genome evolution. Nature. 2007;447(7145):714–9. https://doi.org/10.1038/nature05846 PMID: 17554307

190. Jaillon O, Aury J-M, Brunet F, Petit J-L, Stange-Thomann N, Mauceli E, et al. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature. 2004;431(7011):946–57. https://doi.org/10.1038/nature03025 PMID: 15496914

191. Zhou W, Li S, Liu Y, Qi X, Chen H, Cheng CHK, et al. The evolution of tachykinin/tachykinin receptor (TAC/TACR) in vertebrates and molecular identification of the TAC3/TACR3 system in zebrafish (*Danio rerio*). Mol Cell Endocrinol. 2012;361(1–2):202–12. https://doi.org/10.1016/j.mce.2012.04.007 PMID: 22580006

192. Westerfield M. The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio). 4th ed. Eugene: University of Oregon Press; 2000.

193. UniProt Consortium. UniProt: the universal protein knowledgebase in 2023. Nucleic Acids Res. 2023;51(D1):D523–31. https://doi.org/10.1093/nar/gkac1052 PMID: 36408920

194. Wang X, Zhang B. customProDB: an R package to generate customized protein databases from RNA-Seq data for proteomics search. Bioinformatics. 2013;29(24):3235–7. https://doi.org/10.1093/bioinformatics/btt543 PMID: 24058055

195. Heppert JK, Lickwar CR, Tillman MC, Davis BR, Davison JM, Lu HY, et al. Conserved roles for Hnf4 family transcription factors in zebrafish development and intestinal function. Genetics [Internet]. 2022 [cited 2024 Dec 1];222(4). Available from: https://pubmed.ncbi.nlm.nih.gov/36218393/

196. Pino LK, Searle BC, Bollinger JG, Nunn B, MacLean B, MacCoss MJ. The Skyline ecosystem: informatics for quantitative mass spectrometry proteomics. Mass Spectrom Rev [Internet]. 2020 [cited 2024 Nov 18];39(3):229–44. Available from: https://pubmed.ncbi.nlm.nih.gov/28691345/

197. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S, et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res [Internet]. 2022 [cited 2024 Dec 8];50(D1):D543–52. Available from: https://pubmed.ncbi.nlm.nih.gov/34723319/

198. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5:113. https://doi.org/10.1186/1471-2105-5-113 PMID: 15318951

199. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25(9):1189–91. https://doi.org/10.1093/bioinformatics/btp033 PMID: 19151095

200. Kumar S, Stecher G, Suleski M, Sanderford M, Sharma S, Tamura K. MEGA12: molecular evolutionary genetic analysis version 12 for adaptive and green computing. Mol Biol Evol. 2024;41(12).

201. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4(4):406–25. https://doi.org/10.1093/oxfordjournals.molbev.a040454 PMID: 3447015

202. Chodkowski M, Zielezinski A, Anbalagan S. A ligand-receptor interactome atlas of the zebrafish. iScience [Internet]. 2023 [cited 2024 Oct 20];26(8). Available from: http://www.cell.com/article/S258900422301386X/fulltext

203. Tostivint H, Gaillard AL, Mazan S, Pézeron G. Revisiting the evolution of the somatostatin family: already five genes in the gnathostome ancestor. Gen Comp Endocrinol. 2019;279:139–47.

204. Smith JJ, Kuraku S, Holt C, Sauka-Spengler T, Jiang N, Campbell MS, et al. Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. Nat Genet. 2013;45(4):415–21, 421e1-2. https://doi.org/10.1038/ng.2568 PMID: 23435085

205. Putnam NH, Butts T, Ferrier DEK, Furlong RF, Hellsten U, Kawashima T, et al. The amphioxus genome and the evolution of the chordate karyotype. Nature. 2008;453(7198):1064–71. https://doi.org/10.1038/nature06967 PMID: 18563158

206. Nakatani Y, Takeda H, Kohara Y, Morishita S. Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. Genome Res [Internet]. 2007 [cited 2025 Mar 29];17(9):1254–65. Available from: http://genome.cshlp.org/content/17/9/1254.full

207. Dehal P, Boore JL. Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol. 2005;3(10):e314. https://doi.org/10.1371/journal.pbio.0030314 PMID: 16128622

208. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 1985;39(4):783–91. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x PMID: 28561359

209. Zuckerkandl E, Pauling L. Evolutionary divergence and convergence in proteins. In: Evolving genes and proteins. 1965.

210. Stecher G, Tamura K, Kumar S. Molecular evolutionary genetics analysis (MEGA) for macOS. Mol Biol Evol. 2020;37(4):1237–9. https://doi.org/10.1093/molbev/msz312 PMID: 31904846