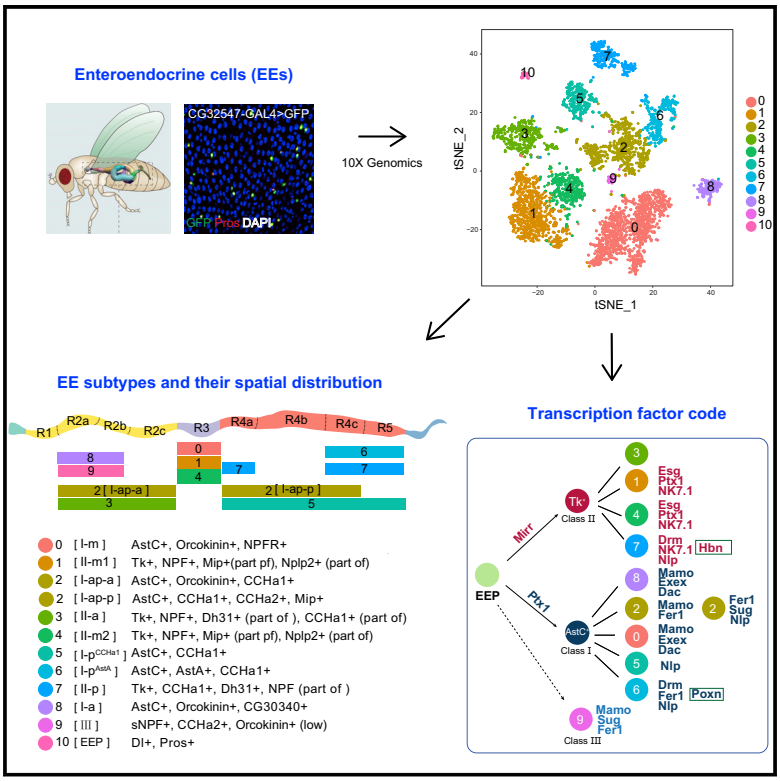
Cell Reports

Resource

The Cellular Diversity and Transcription Factor Code of *Drosophila* Enteroendocrine Cells

Graphical Abstract



Highlights

- Single-cell analysis reveals 10 EE subtypes in *Drosophila* midgut
- Most EEs produce approximately 2–5 peptide hormone classes
- The spatial distribution of EE subtypes along the length of midgut is determined
- TF code analysis and functional screens identify EE subtype regulators

Authors

Xingting Guo, Chang Yin, Fu Yang, ..., Tao Cai, Yi Rao, Rongwen Xi

Correspondence

xirongwen@nibs.ac.cn

In Brief

Through single-cell RNA-sequencing analysis and functional screens, Guo et al. provide a comprehensive characterization of the cellular diversity, spatial distribution, and transcription factor code of enteroendocrine cells in adult *Drosophila* midgut.

Guo et al., 2019, Cell Reports *29*, 4172–4185
December 17, 2019
https://doi.org/10.1016/j.celrep.2019.11.048

CellPress

OPEN ACCESS
CellPress

Cell Reports
Resource

The Cellular Diversity and Transcription Factor Code of *Drosophila* Enteroendocrine Cells

Xingting Guo,[1,4] Chang Yin,[1,4] Fu Yang,[1,4] Yongchao Zhang,[1,4] Huanwei Huang,[1] Jiawen Wang,[1] Bowen Deng,[2] Tao Cai,[1] Yi Rao,[2] and Rongwen Xi[1,3,5,*]

[1]National Institute of Biological Sciences, No. 7 Science Park Road, Zhongguancun Life Science Park, Beijing 102206, China
[2]Peking University School of Life Sciences, Beijing 100091, China
[3]Tsinghua Institute of Multidisciplinary Biomedical Research, Tsinghua University, Beijing 102206, China
[4]These authors contibuted equally to this work
[5]Lead Contact
*Correspondence: xirongwen@nibs.ac.cn
https://doi.org/10.1016/j.celrep.2019.11.048

SUMMARY

Enteroendocrine cells (EEs) in the intestinal epithelium have important endocrine functions, yet this cell lineage exhibits great local and regional variations that have hampered detailed characterization of EE subtypes. Through single-cell RNA-sequencing analysis, combined with a collection of peptide hormone and receptor knockin strains, here we provide a comprehensive analysis of cellular diversity, spatial distribution, and transcription factor (TF) code of EEs in adult *Drosophila* midgut. We identify 10 major EE subtypes that totally produced approximately 14 different classes of hormone peptides. Each EE on average co-produces approximately 2–5 different classes of hormone peptides. Functional screen with subtype-enriched TFs suggests a combinatorial TF code that controls EE cell diversity; class-specific TFs Mirr and Ptx1 respectively define two major classes of EEs, and regional TFs such as Esg, Drm, Exex, and Fer1 further define regional EE identity. Our single-cell data should greatly facilitate *Drosophila* modeling of EE differentiation and function.

INTRODUCTION

Apart from the function in food digestion and absorption, the gastrointestinal tract is also considered as the largest endocrine organ due to the resident enteroendocrine cells (EEs). In mice and humans, EEs are scattered throughout the intestinal epithelium and take up only 1% of total intestinal cells, yet they produce more than 20 types of hormones that regulate a diverse of physiological processes, such as appetite, metabolism, and gut motility. There are at least 12 major subtypes of EEs based on hormones that they produce, and due to their great regional and local cellular diversity, the complete characterization of EE specification and diversification still remains as a challenge (Furness et al., 2013; Gribble and Reimann, 2016).

The adult *Drosophila* midgut has become an attractive model system for the understanding of EE cell diversity and their regulatory mechanisms (Buchon et al., 2013; Marianes and Spradling, 2013; Micchelli and Perrimon, 2006; Ohlstein and Spradling, 2006). The EEs are scattered along the epithelium of the entire midgut, including anterior midgut (regions R1 and R2), middle midgut (the gastric region, R3) and posterior midgut (regions R4 and R5) (Buchon et al., 2013; Marianes and Spradling, 2013). They have important roles in regulating local stem cell division and lipid metabolism, as well as feeding and mating behaviors (Amcheslavsky et al., 2014; Ameku et al., 2018; Min et al., 2016; Song et al., 2014). Approximately 10 peptide hormone genes are found to be expressed in EEs, yielding more than 20 different peptide hormones (Reiher et al., 2011; Veenstra and Ida, 2014; Veenstra et al., 2008). Studies using RNA *in situ* hybridization, antibody staining, and gene reporter tools have provided a glimpse of regional EE diversity in terms of peptide hormones that they produce (Beehler-Evans and Micchelli, 2015; Chen et al., 2016a; Veenstra et al., 2008). However, due to limited availability of antibodies against all these hormones and a limit in the number of hormones that can be simultaneously analyzed, the detailed characterization of EE subtypes and peptide profiles is still lacking.

As in mammals, EEs in the fly midgut are derived from multipotent intestinal stem cells (ISCs). The initial fate determination between absorptive enterocyte versus secretory EEs is controlled by Notch signaling and appears to be regulated by the antagonistic activities of E(spl)-C genes and achaete-scute complex genes (Chen et al., 2018; Ohlstein and Spradling, 2007). This is also analogous to the antagonistic activities between Hes1 (orthologous to E(spl)) and Math1 (paralogous to AS-C) in mammalian ISCs that control the initial cell fate decision (Jensen et al., 2000; Yang et al., 2001). The committed EE progenitor cell usually divides one more time to yield a pair of EEs. Interestingly, the two EEs within each pair produce distinct hormone peptides as a result of differentially acquired Notch activity, suggesting that, at least in the posterior midgut, differential Notch signaling defines two major subtypes of EEs (Beehler-Evans and Micchelli, 2015; Chen et al., 2018; Guo and Ohlstein, 2015). The specification and commitment of EE fate requires the homeodomain transcription factor (TF) Prospero (Pros) (Biteau and Jasper, 2014; Wang et al., 2015; Yin and Xi, 2018; Zeng and Hou, 2015), and the maturation of peptide hormones in EEs requires a Neuro D family bHLH TF Dimmed (Dimm) (Beebe et al., 2015). Besides these general TFs that promote EE specification and function, little is known regarding the TFs that participate in EE subtype specification and regional EE identity.

Single-cell RNA-sequencing (scRNA-seq) has emerged as an efficient tool for revealing cell heterozygosity in different tissues and organisms (Gehart et al., 2019; Haber et al., 2017; Kumar et al., 2017; Li et al., 2017). By using scRNA-seq and a collection of recently generated peptide and receptor knockin lines, here we provide a comprehensive analysis of EE cell diversity, peptide profiling, and regional distribution along the entire length of the fly midgut at single-cell resolution. In addition, TF enrichment analysis followed by genetic screen allowed us to identify class and region EE regulators. Our results suggest a TF code composed of class-specific and region-specific TFs generates EE cell diversity. Our single-cell data and an associated online database (https://xilab.shinyapps.io/database/) should serve as an important resource and foundation upon which the differentiation, regulation, and function of EEs can be further modeled and investigated.

RESULTS

scRNA-Seq Analysis of EEs Identifies Distinct EE Populations

We performed 10X Genomics scRNA-seq analysis of EEs from adult *Drosophila* midgut (Figure 1A). We identified a GAL4 insertion line (CG32547-KI-GAL4) (Deng et al., 2019), which drives upstream activating sequence (UAS)-GFP expression specifically in EEs in the midgut, as the GFP expression matched perfectly with the expression of the EE marker Pros (Figure S1A). Using this CG32547-KI-GAL4; UAS-GFP line, EEs from 5- to 7-day-old female flies were sorted out by fluorescence-activated cell sorting (FACS), and the transcriptional profile of each single cell was analyzed by using Seurat algorithm. A total of 4,661 single cells—about four times the number of total EEs found in an entire midgut (Chen et al., 2016a)—were profiled after filtering out low-quality cells. Cells were sequenced to an average depth of 36,000 reads per cell and had an average of 1,350 genes per cell, with an average of 4,400 unique molecular identifiers (UMIs) and about 85% transcriptome mapping ratio (Figure S1B).

Seurat partitioned these EEs into 11 distinct clusters (C0–C10) by using the top 20 principle components when under the resolution of 0.4. Then we visualized these EEs by generating a t-distributed stochastic neighborhood embedding (t-SNE) plot (Figure 1B). *pros* expression was detectable in virtually all these cells, further confirming the specificity of the CG32547-GAL4 reporter line in our analysis (Figure 1C). Interestingly, EE-C10, a unique cluster, was distinguished by a relatively low level of *pros* expression and high levels of *Dl* and *Notch* expression (Figures 1D and 1E). These cells also retained the expression of ISC-enriched genes, including *Sox21a* and *Sox100b*, as well as several ISC-specific *E(spl)-C* genes such as m3 and mα (Chen et al., 2016b; Doupé et al., 2018; Guo et al., 2019; Meng and Biteau, 2015; Zhai et al., 2015) (Figure S2A, highlighted in blue), indicating that these cells are EE progenitor cells (EEPs), and each of them usually gives rise to a pair of EEs by dividing once before terminal differentiation (Chen et al., 2018).

Peptide Hormone Signatures for Each EE Cluster

To further characterize the remaining 10 mature EE populations (C0–C9), we analyzed their transcriptional signatures by choosing top 10 highly expressed genes specific for each cluster. Interestingly, many genes that showed up on the list were peptide hormone genes and peptide hormone receptors (Figure S2, highlighted in yellow). This is consistent with the idea that different EEs produce different peptide hormones. We thus analyzed the expression pattern of all peptide hormone genes among these 10 clusters.

We identified 14 peptide hormone genes out of the annotated 48 genes encoding secreted peptides (Flybase ID: FBgg0000179) that were expressed in one or more cell clusters. Interestingly, the expression pattern of these 14 peptide hormone genes showed that each of them was expressed in one or more clusters, and none of them were expressed in all clusters (Figure 2A). It is worth to note that there was an absence of one-to-one simple correlations between a specific kind of peptide hormone and a specific cluster, showing a previously unprecedented complexity of peptide hormone expression in EEs and EE subtypes and indicating a limit or pitfall in classifying EE types solely by peptide hormones that they produce.

In combination with a collection of antibodies and gene reporter lines, including the recently generated Gal4 knockin lines specific for peptide hormones and G-protein coupled receptors (GPCRs) (Deng et al., 2019), we validated the peptide hormone profiles for most EE clusters, as well as co-localization relationships among different peptide hormones. Previous studies reveal two major classes of EEs in the posterior midgut that respectively express Allatostatin C (AstC) and Tachykinin (Tk) peptides in a mutually exclusive manner (Beehler-Evans and Micchelli, 2015). We found that these two major EE classes could be applied to other midgut regions as well, except a small region at R4b, where Tk-expressing EEs could not be found. Therefore, the majority of EEs along the entire length of midgut could be divided into two major categories: class I EEs expressing AstC and class II EEs expressing Tk (Figures 2B, 2C, and 2D). To identify the co-expression pattern with other peptide hormones, Neuropeptide F (NPF, in clusters 1, 3, and 4 and a part of cluster 7) and Diuretic hormone 31 (DH31, in cluster 7) were co-localized specifically with Tk. Allatostatin A (AstA, in cluster 6) was specifically co-expressed with AstC (Figure 2A). Consistent with the peptide hormone profiles deduced from the scRNA-seq data, immunostaining with the gut also showed that NPF and DH31 were only expressed in Tk^+ $AstC^-$ cells and AstA was co-stained with AstC but not Tk (data not shown) (Beehler-Evans and Micchelli, 2015; Chen et al., 2016a). We also observed several peptide hormones that were co-localized with either Tk or AstC in different clusters. For example, CCHamide-1 (CCHa1) was co-localized with AstC in clusters 2, 5, and 6, while it was co-localized with Tk in cluster 7 (Figure 2A). Myoinhibiting peptide precursor (Mip) was expressed in $AstC^+$ cells in cluster 2 as well as Tk^+ cells in clusters 1 and 4 (Figures 2A, 2L, and 2M).

The expression pattern of 4 additional gut peptide hormone genes identified in this study was also determined. Neuropeptide-like precursor 2 (Nplp2) and Glycoprotein hormone beta 5 (Gbp5) were enriched in TK^+ cells and mainly expressed in EEs of clusters 1 and 4 (Figure 2A). Ion transport peptide (ITP) was expressed in all types of epithelial cells in the midgut, including EEs, but exhibited the highest expression level in EE cluster 7 in R4 and R5 (Figures 2A, 2N, and 2O). Consistent with this spatial expression pattern at the posterior end, ITP has been reported

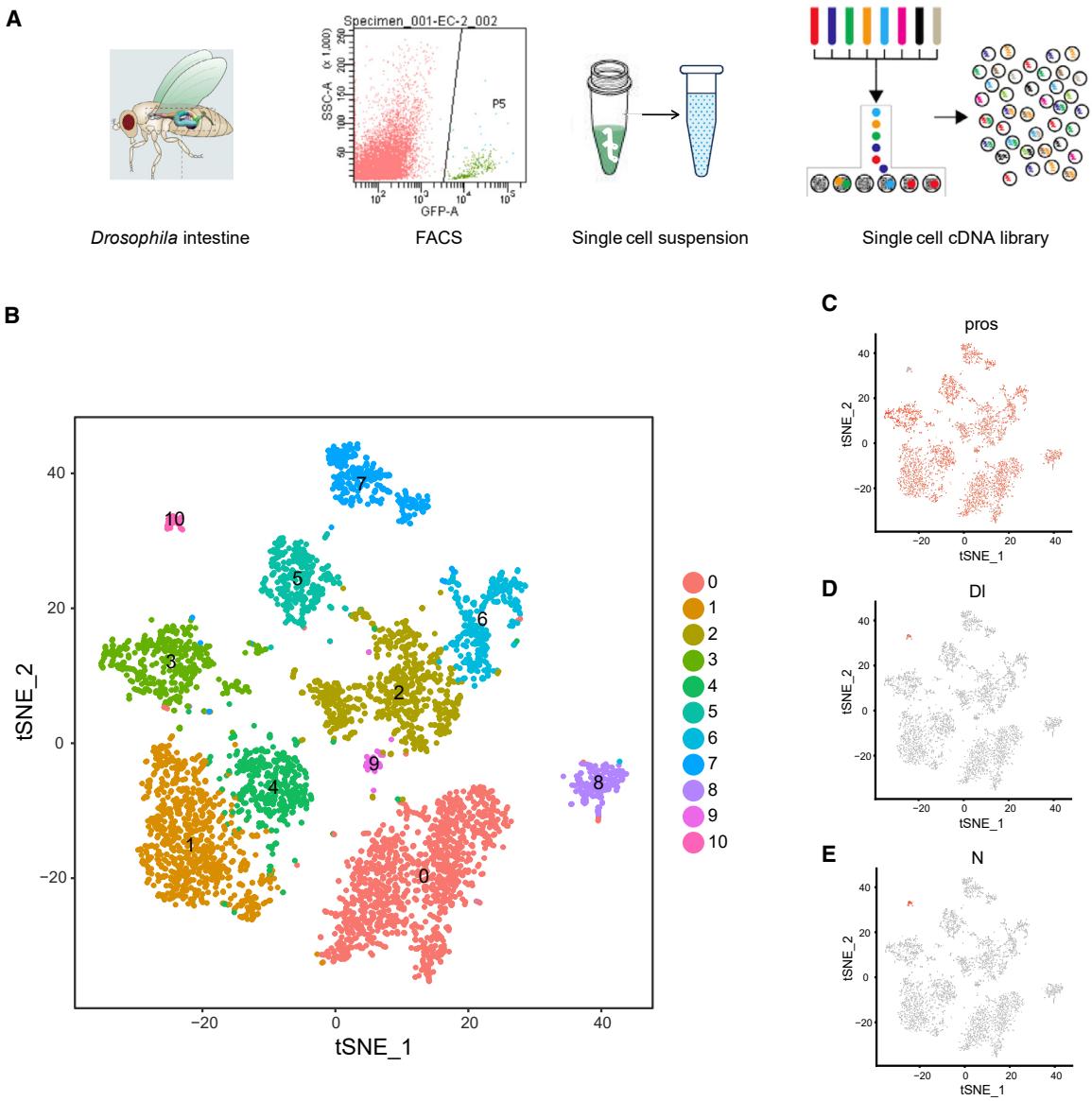


Figure 1. Single-Cell RNA-Seq Analysis of EEs in the *Drosophila* Midguts

(A) Schematic diagram of experimental design. Midguts were dissected from 5- to 7-day-old female flies, sorting EEs by FACS according to the specific expression of CG32547-GAL4 > GFP, and 10X Genomic scRNA-seq performed to establish single-cell cDNA library.

(B) EE subtypes visualized by t-SNE plot of 4,661 cells based on Seurat analysis.

(C–E) Visualization of expression patterns of key genes in t-SNE plot. Expression of *pros* could be detected in all clusters (C), and high levels of Dl (D) and Notch (E) are specifically detected in EE-C10. Color shade is according to gene expression levels.

See also Figure S1.

to regulate ion transport in hindgut (Nässel and Winther, 2010). Short neuropeptide F precursor (sNPF) was expressed in a small subset of EEs mainly belonging to cluster 9 (Figure 2A).

Interestingly, cluster 9 (blue square in Figures 2B, 2C, and 2E) may represent a specific EE population as it was not marked by either Tk or AstC (yellow arrowhead in Figure 2D). As these cells do not belong to either class I or class II, we termed this subtype as class III EEs. Peptide hormones including CCHamide-2 (CCHa2), Orcokinin, and sNPF were found to be expressed in these class III EEs (Figures 2A and 2E). Among them, CCHa-2 and Orcokinin were also found in subsets of AstC^{+} cells (Figure 2A) (in cluster 2 for CCHa2 and clusters 0, 2, and 8 for Orcokinin). By immunostaining, we found that all CCHa2^{+} EEs in the R2 region were TK^{-} AstC^{-} (Figures 2H and 2I), confirming the existence of the class III EEs. Orcokinin was generally expressed in AstC^{+} cells in R3–R5 regions (lower panel of Figure 2J). Orcokinin^{high} EE cells in the R2 region (white arrow in upper panel of Figure 2J) were also AstC^{+}, but Orcokinin^{low} cells (yellow arrow in

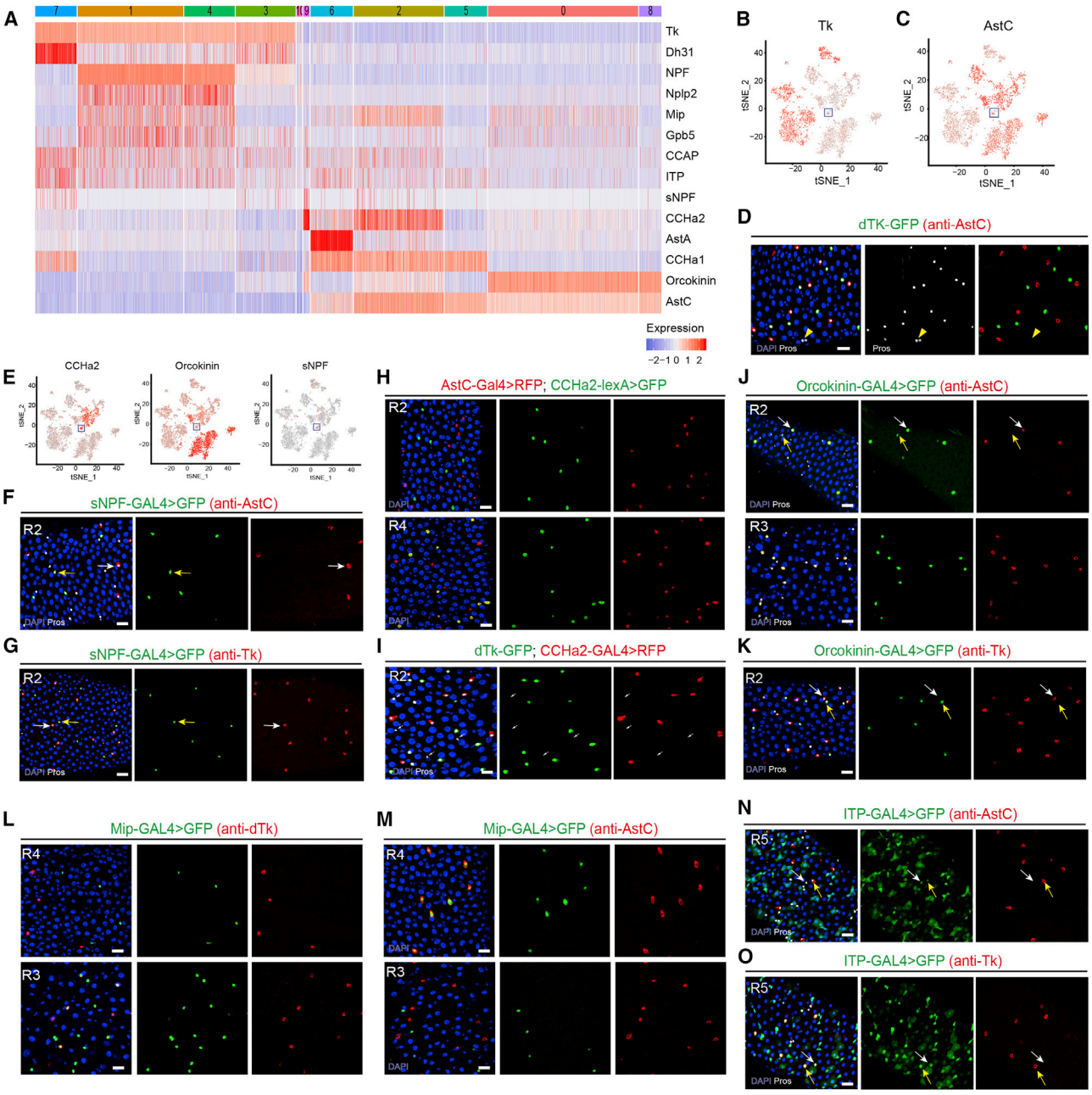


Figure 2. Peptide Hormone Signatures for Each EE Cluster

(A) Heatmap of the 14 peptide hormone genes detected in *Drosophila* EE scRNA-seq.

(B and C) Expression patterns of Tk (B) and AstC (C) in t-SNE plot. Cluster 9 is highlighted with blue square in (B) and (C).

(D) Expression of Tk and AstC are exclusive from each other, and most EEs could be separated into Tk^{+} or $AstC^{+}$, while a small set of $Pros^{+}$ EEs could not be labeled by either Tk or AstC (yellow arrowhead).

(E) Expression patterns of CCHa2 (left), Orcokinin (middle), and sNPF (right) in t-SNE plot. Cluster 9 is highlighted with blue square.

(F and G) sNPF is expressed in R2 region, and its expression is mutually exclusive with either AstC (F) or Tk (G).

(H and I) CCHa2 is co-localized with AstC in R4 (I, lower panel) but is mutually exclusive with either Tk (H) or AstC (I, upper panel) in R2 region.

(J) Orcokinin is co-localized with AstC in R3 (lower panel); in R2 region (upper panel), Orcokinin ^{high} cells (white arrow) are also co-localized with AstC, while Orcokinin ^{low} cells (yellow arrow) are $AstC^{-}$.

(K) Orcokinin expression (yellow arrow) is mutually exclusive with Tk^{+} cells (white arrow).

(L) Mip expression is partly co-localized with Tk in R3 but mutually exclusive in R4 region.

(legend continued on next page)

upper panel of Figure 2J) were AstC^{-} (Figure 2J). As Orcokinin^{+} cells were negative for Tk expression (Figure 2K), these Orcokinin^{low} EEs in R2 should belong to class III EEs. sNPF was expressed exclusively in EEs in the anterior midgut (Figure S4H), and all sNPF^{+} cells were Tk^{-} AstC^{-}, making it a possible signature hormone peptide for class III EEs (Figures 2F and 2G).

Because we can only detect on average 1,350 genes per cell, it remains possible that hormones with very low expression levels or whose expression is conditional could be missed from our analysis. For example, Bursicon is not detected in our analysis despite it reported to be expressed in some EEs in the posterior midgut (Chen et al., 2016a; Scopelliti et al., 2019). Interestingly, its expression can only be detected by antibody staining but not by mRNA using *in situ* hybridization (Chen et al., 2016a), suggesting that the transcriptional activity of *Bursicon* is extremely low but its protein product might be relatively stable.

The annotation of peptide hormone signatures for each EE subtype was shown in Figure 4C.

Peptide Hormone Co-expression Pattern in Individual EEs

The expression pattern analysis also revealed that each given EE can produce more than one type of peptide hormone. We therefore further analyzed the peptide hormone co-expression pattern at single-cell resolution. By systematically profiling peptide hormone gene expression in every single EE, we observed that approximately 80% of individual EEs co-expressed 2–5 classes of peptide hormones. Approximately 23% of individual EEs co-expressed 2 or 3 classes of peptide hormones, which is the most common scenario in the number of peptide hormones co-expressed in each EE. In rare cases, up to 7–8 different classes of peptide hormones could be detected in a single EE cell (Figure 3A). By listing all of peptide hormone combinations in individual EEs, it was clear that there were extensive variations in the frequency of peptide hormone co-expression patterns (Figure 3B). For example, the most frequent co-expressed peptide hormones were AstC and Orockinin, but their co-expression was found only in 8.3% of EEs. 13 out of the top 15 most frequent peptide hormone co-expression patterns occurred in less than 2% of EEs (Figure 3B).

Great variation in the peptide hormone expression profile was also found within a common EE cluster. For example, in the two closely related clusters 1 and 4, apart from the pan peptide hormones Tk and NPF that expressed in almost all EEs in these two clusters, their combinational expression with Nplp2, Mip, Gpb5, Crustacean cardioactive peptide (CCAP), and ITP exhibited great variations (Figures 3C and 3D). In fact, their differences in the frequency of peptide hormone co-expression patterns helped to separate the identity of these two clusters (Figures 3C and 3D). The summary of peptide hormone co-expression profiles for each cluster were listed in Figure S3.

It is well known that each peptide hormone gene usually encodes a large propeptide precursor, which is subjected to proteolytic processing to generate multiple peptide hormones with potentially different functions. As such, a given propeptide precursor could yield different sets of peptide hormones depending on the processing enzymes expressed (Steiner, 1998). We therefore also surveyed all known proprotein convertases involved in the process in *Drosophila* (Pauls et al., 2014; Reiher et al., 2011) and found that most of them showed largely unbiased expression in all EE subtypes, although their expression levels could be generally high or low (Figures 3E and 3F). However, Amontillado (Amon) and Furin 1 (Fur1) exhibited subtype-restricted expression patterns. Amon was expressed in virtually all EE subtypes except C7, the Tk^{+} DH31^{+} EEs located in R4c and R5, and Fur1 was specifically expressed both in C3 and C7 subtypes that are Tk^{+} and in C5 and C6 subtypes that are AstC^{+} (Figure 3G). The both abundant and diverse co-expression patterns of the processing enzymes in each EE subtype may add additional diversity of peptide hormones expressed and consequently additional functional diversity.

Determination of Spatial Distribution of EE Subtypes

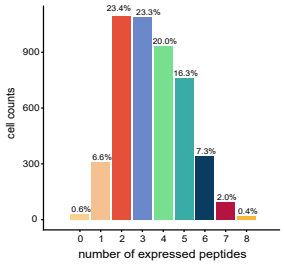
It has been observed that the gut peptide hormones show region-specific expression patterns along the length of midgut (Chen et al., 2016a; Veenstra and Ida, 2014; Veenstra et al., 2008). We also verified these regional distributions using a series of gene knockin GAL4/LexA lines (Figures S4A–S4I). For example, DH31 is mainly expressed in posterior midgut (Figure S4D), and no CCHa1 or CCHa2 positive cells could be detected in middle midgut (Figure S4F). As the spatial information for individual EEs was lost during tissue dissociation, the systemic determination of the spatial distribution pattern for all EE subtypes was challenging.

To overcome this problem, we generated a region-specific gene enrichment (RSGE) algorithm based on regional total EE RNA-seq data from previously published datasets (Dutta et al., 2015). We then determined the regional distribution for each EE cluster based on the relative expression score for the top 100 region-specific genes in R1 to R5 (Figures 4A and 4B). We also systematically investigated the expression patterns of all the cluster signature peptide hormones and GPCR receptors in the midgut using the available knockin lines (Figures S4A–S4I).

We found that the results from these two approaches matched nicely. For example, clusters 6 and 7 were mapped to R4 and R5 by the RSGE algorithm (Figure 4B), and consistent with this notation, their respective signature peptide hormones AstA and DH31 were also expressed in R4c and R5 (Figure S4C–S4D). As for cluster 9, the class III EEs that are double-negative for Tk and AstC, they were annotated to R2 only (Figure 4B), and this was in line with our immunostaining results showing that the Tk and AstC double-negative but CCHa2 positive EEs were found only in R2 (Figures 2H and 2I). We also noticed that the R3-specific genes were strongly enriched in C0, C1, and C4 clusters; these clusters also specifically expressed the TF Escargot (Esg) (Figures 4B and 6A), whose expression has been

(M) Mip expression is co-localized with AstC in R4 but mutually exclusive in R3 region.
(N and O) ITP is expressed in multiple intestinal epithelium cells, including Pros^{+} EE cells. ITP producing cells could separately co-localize with AstC (N) or Tk (O). Nuclei are labeled by DAPI staining. Pros, EE specific cell marker. Scale bars, 20 μm.
See also Figure S2.

A



B Peptides combination in the whole midgut EEs

peptides combination	peptides number	major cluster	sum_frequency
AstC+Orockinin	2	0, 5, 8	8.30%
AstC+CCHa1	2	2, 5	4.31%
Orockinin	1	0	3.37%
Tk	1	2, 7	2.19%
Mip+Orockinin	2	0	1.72%
AstC+Mip+Orockinin	3	0	1.65%
AstC+CCHa1+CCHa2+Mip	4	2	1.59%
AstA+AstC+CCHa1+CCHa2+Mip	5	2, 6	1.39%
NPF+Nplp2+Tk	3	1, 4	1.29%
NPF+Tk	2	1, 4	1.24%
Dh31+Tk	2	3, 7	1.24%
Mip+NPF+Nplp2+Tk	4	1, 4	1.20%
Mip+NPF+Tk	3	1, 4	1.16%
AstC+Gpb5+Orockinin	3	0, 8	1.12%
AstC+CCHa1+ITP	3	2, 5	1.07%
Gpb5+NPF+Nplp2+Tk	4	0, 2	1.07%
AstC+CCHa1+Orockinin	3	0, 2	1.07%

C peptides combination in cluster 1

peptides combination	number	frequency
NPF+Tk	2	5.38%
Mip+NPF+Tk	3	4.63%
Gpb5+Nplp2+NPF+Tk	4	4.51%
Nplp2+NPF+Tk	3	4.38%
Gpb5+NPF+Tk	3	4.13%
Gpb5+Mip+Nplp2+NPF+Tk	5	4.01%
Mip+Nplp2+NPF+Tk	4	4.01%
Gpb5+Mip+NPF+Tk	4	3.25%
CCAP+Gpb5+Mip+NPF+Tk	5	3.00%

D peptides combination in cluster 4

peptides combination	number	frequency
Mip+Nplp2+NPF+Tk	4	6.28%
Nplp2+NPF+Tk	3	6.02%
CCAP+Gpb5+Nplp2+NPF+Tk	5	4.19%
Mip+NPF+Tk	3	4.19%
Gpb5+Mip+Nplp2+NPF+Tk	5	3.40%
Gpb5+Nplp2+NPF+Tk	4	3.40%
Dh31+Nplp2+NPF+Tk	4	3.40%

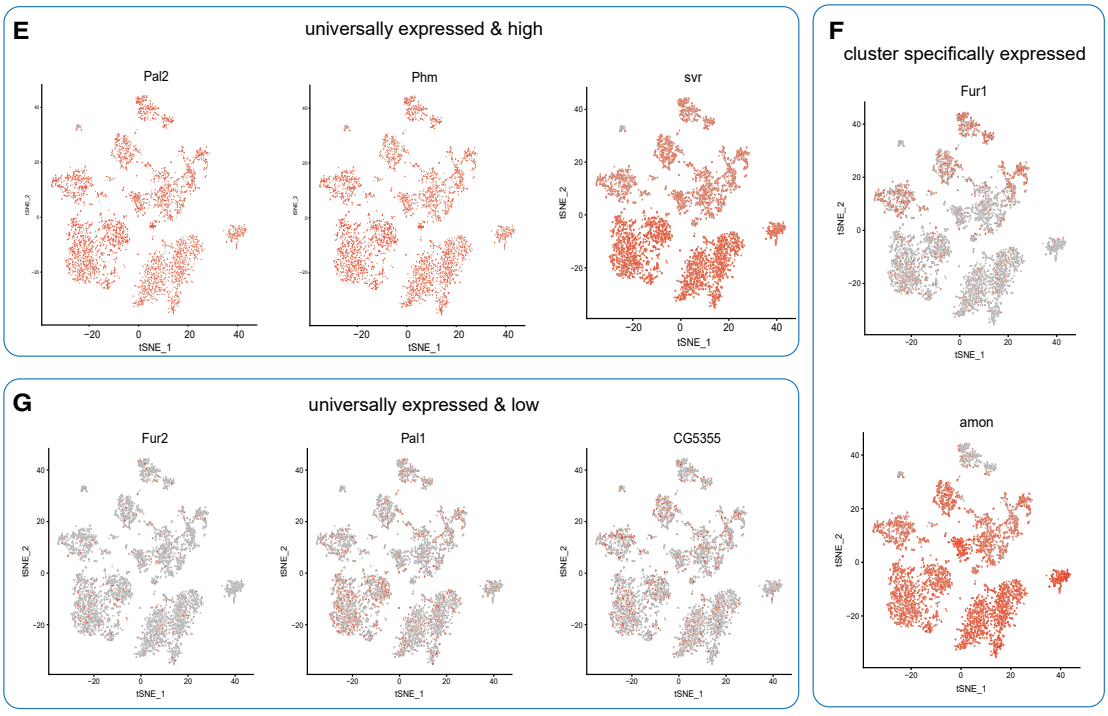


Figure 3. Peptide Hormone Co-expression Pattern in Individual EEs

(A) Distribution of the number of peptide hormones co-expressed in single EEs.

(B) Peptide hormone combinations in total midgut EEs with a percentage more than 1%.

(C and D) Peptide hormone combinations in clusters 1 (C) and 4 (D) with a percentage more than 3%. Pan-hormone peptides in each combination are highlighted in orange.

(legend continued on next page)

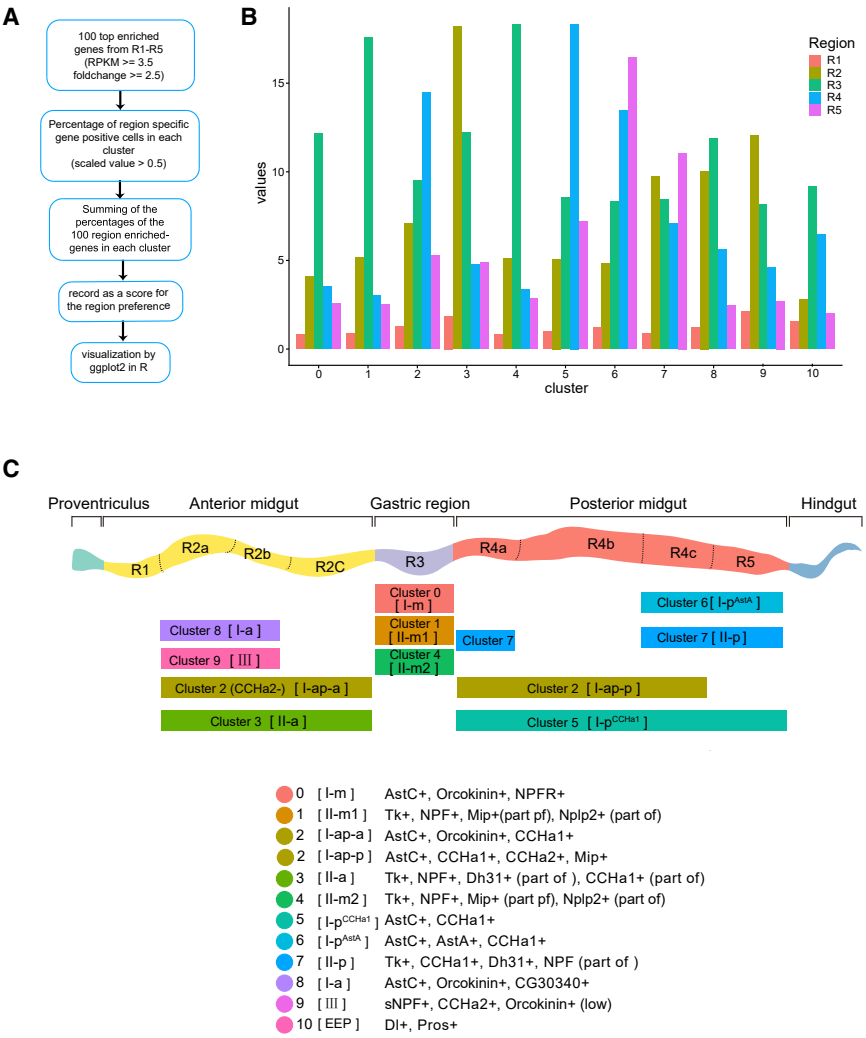


Figure 4. Determination of Spatial Distribution of EE Subtypes

(A) Workflow of the region-specific gene enrichment (RSGE) algorithm designed to calculate regional preference in each EE subtype.

(B) Visualization of regional preference of each EE subtype.

(C) A graphic summary of signature peptide hormones and spatial distributions of EE subtypes along the length of midgut.

See also Figure S4.

believed to be specific for intestinal progenitor cells. By immunostaining, we confirmed that all *esg*$^+$ EEs (either Tk$^+$ or AstC$^+$) were found in R3 only (Figures S4J and S4K). Therefore, The C0, C1, and C4 EE clusters reside in R3 only.

This combined approach allowed us to draw the spatial distribution of all EE clusters along the length of the midgut: both C8 and C9 are in the R2ab region; C3 and one part of C2 without CCHa2 expression are in the R2 region, while another part of C2 is in the R4 region; C0, C1, and C4 are in the R3 region; and C6 and C7 are at the most posterior region, from R4c to R5 of the midgut (Figure 4C). To reflect the positional information of these EE clusters, we renamed them under class I and II categories with position labels ("a" for anterior midgut, "p" for posterior, and "m" for middle) and also used peptide hormone names to distinguish subtypes when necessary (Figure 4C): the class I EEs have 5 subtypes, including I-a (corresponding to cluster 8), I-ap (cluster 2), I-m (cluster 0), I-p^{CCHa1} (cluster 5), and I-p^{AstA} (cluster 6). The I-ap subtype is distributed to both anterior and posterior midgut, and the ones in the posterior, but not anterior, express CCHa2; the class II EEs have 4 subtypes, including II-a (cluster 3), II-m1 (cluster 1), II-m2 (cluster 4), and II-p (cluster 7); the class III EEs only have one subtype, localized only in the R2 region. As restricted regional distribution appears to be a common feature for EE subtypes, the region-specific environment is likely an important factor that contributes to cellular diversity of EEs.

Expression of Peptide Hormones and Their Receptors in EE Subtypes Is Indicative of Local and Long-Range Communications

In addition to peptide hormone secretion, EE cells also function as sensors of many environmental stimuli through GPCR receptors. To investigate potential communications between peptide hormone production and perception, we systematically profiled the expression of peptide hormone receptors in EEs (Flybase ID: FBgg0000195, FBgg0000174). Out of 19 GPCRs that were found to be expressed in EEs, 5 were known receptors of the peptide hormones produced by EEs (Figure 5A, highlighted in yellow). We then individually examined the expression pattern of these peptide hormone receptors using the scRNA-seq data and available knockin lines and compared with the expression pattern of the corresponding peptide hormones.

Interestingly, the expression of peptide hormones and their corresponding receptors in general showed non-overlapping patterns. For example, NPF was expressed in Tk$^+$ cells, while its receptor NPFR was expressed in AstC$^+$ cells, and both of them exhibited the strongest expression level in R3 (Figures 5B and 5C). These observations indicate a paracrine signaling

(E–G) Visualization of peptide hormone procession enzymes in t-SNE plot. (E) Universally expressed enzymes with no subtype specificity, including Pal2, PHM, and SVR. (F) Low expressed enzymes with no subtype specificity, including Fur2, Pal1, and CG5355. (G) Amon and Fru1 exhibit subtype-restricted expression patterns.

See also Figure S3.

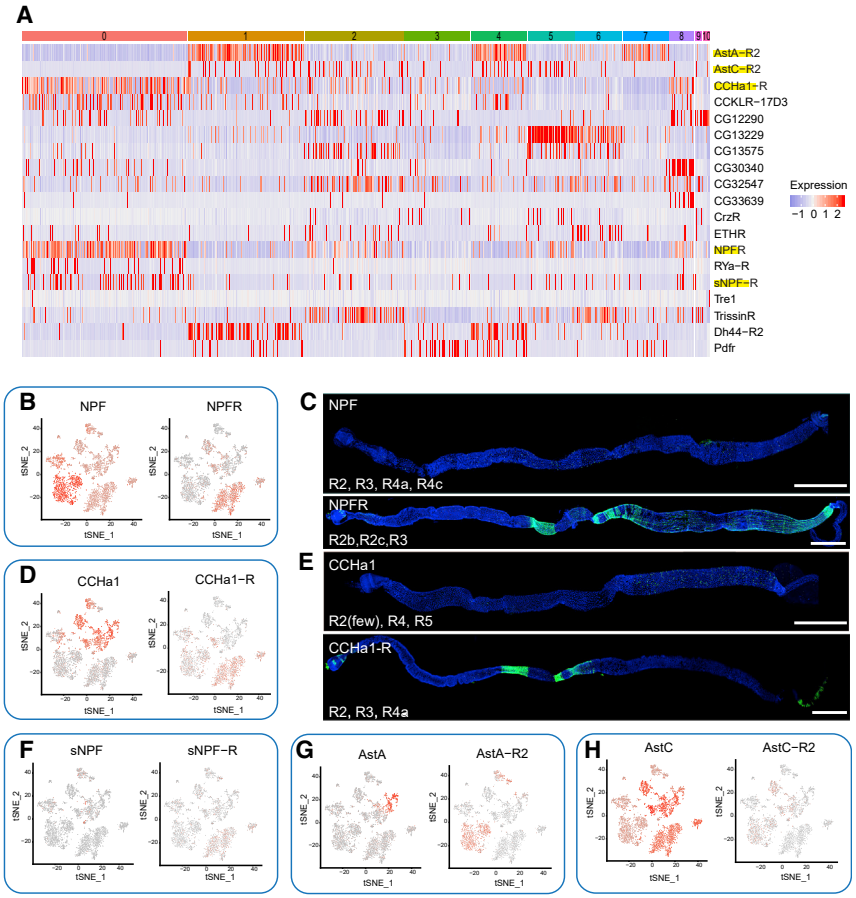


Figure 5. Expression of Peptide Hormones and Their Receptors in EE Subtypes Indicates Local and Long-range Communications

(A) Heatmap of *Drosophila* peptide hormone receptor genes detected in our EE scRNA-seq. Receptors with peptide hormone ligands expressed in EEs are highlighted in yellow.

(B and C) Visualization of NPF and its receptor NPFR expression in t-SNE plot (B) and whole midgut staining (C). Both of NPF and NPFR are expressed in R3 EEs but in different subtypes (B). NPFR is also expressed in some visceral muscle cells.

(D and E) Visualization of CCHa1 and its receptor CCHa1-R expression in t-SNE plot (D) and whole midgut staining (E). CCHa1 is mainly expressed in the posterior midgut, while CCHa1-R is expressed in both anterior and middle midgut (D). CCHa1-R also shows expression in visceral muscle cells (E).

(F–H) Visualization of sNPF (F), AstA (G), AstC (H), and their corresponding receptors in t-SNE plot. Scale bars, 500 μm.

mechanism between class I and class II EEs in R3 mediated by NPF and NPFR. The expression patterns of another peptide hormone CCHa1 and its receptor CCHa1-R indicated a potential long-range communication, as CCHa1 was mainly expressed in the posterior midgut, while its receptor CCHa1-R was mainly expressed in I-a (C8) EEs and I-m (C0) EEs (Figures 5D and 5E). Similarly, the expression patterns of sNPF, AstA, and AstC and their corresponding receptors were also indicative of their potential involvement in paracrine or long-range signaling communications among different EE clusters (Figures 5F–5H).

Although we did not intensively investigate the communication between EEs and other types of cells, it is well known that EEs send signals to many different cells and organs to exert a variety of physiological functions. We also found that NPFR and CCHa1-R were expressed in some visceral muscle cells as well, indicating signaling communications between EEs and surrounding muscles (Figures 5C and 5E). Communication between EEs and enterocytes via Tk signaling has also been implicated in controlling lipid metabolism in enterocytes (Song et al., 2014). It has been recently reported that the gut-derived NPF could also impact on the proliferation of germline stem cells in the ovary (Ameku et al., 2018). It may worth mentioning that EEs could also be signal-receiving cells from signals produced by other organs. Consistent with previous reports, Diuretic hormone 44 receptor 2 (DH44-R2) was found to be expressed in EEs (Figure 5A), and its ligand Diuretic hormone (DH44) is known to be expressed neurons of the pars intercerebralis, functioning to regulate intestinal ingestion and digestion (Cabrero et al., 2002; Dus et al., 2015).

Notch Activity Controls EE Subtype Specification, Most Likely at the EEP Stage

Notch is essential for the specification of class II Tk^+ EEs but not class I $AstC^+$ EEs. Depleting Notch activity in mature Tk^+ EEs fails to alter the expression of Tk, indicating that Notch regulates the specification, but not the maintenance, of class II EEs (Beehler-Evans and Micchelli, 2015). During the EE generation process, an EEP divides once before terminal differentiation to yield two EEs that respectively express Tk and AstC (Chen et al., 2018). In our scRNA-seq data, a certain level of Notch activity, reflected by the expression of several *E(spl)-C* genes (Figure S2), has been detected in EEP cells. Collectively, these observations indicate that Notch signaling might control the binary cell fate choice between the two daughter EEs during the process of EEP division. To test whether the low activity of Notch in EEPs participates in EE subtype specification, we depleted Notch activity in EEPs and EEs by using *pros-v1-GAL4, UAS-GFP; Tub-GAL80^{ts}* to drive *Notch-RNAi* for 5 days. By conducting immunostaining experiments followed by quantitative analysis, we found that although the percentage of $AstC^+$ cells was increased and the percentage of Tk^+ cells was proportionally decreased in Notch-depleted gut (Figure S5), this does not suggest a direct conversion of Tk^+ EEs into $AstC^+$ cells. Instead, we observed some EE pairs—two EEs juxtaposed to each other, presumably the newly formed EEs—with both cells positive for $AstC^+$ (Figure S5B, yellow arrowhead), which was likely responsible for the tilted Tk^+ EE versus $AstC^+$ EE ratios. Thus, loss of Notch is not sufficient to convert the already differentiated Tk^+ EEs into

AstC$^+$ EEs, but it may be sufficient to cause symmetric cell division of EEPs to yield EE daughters that are all AstC$^+$.

Using the same binary expression system, we also overexpressed an active form of Notch (Nicd) in EEPs and EEs for 5 days and examined the consequences. Interestingly, ectopic Notch activation did not cause a significant change of the percentage of Tk$^+$ EEs but eliminated AstC expression in the majority of EEs (Figure S5), suggesting that Notch activation is sufficient to inhibit AstC expression but not sufficient to induce Tk expression in the differentiated EEs. We occasionally spot some presumably newly formed EE pairs, and in these cases, both EEs in the pair were positive for Tk$^+$ (Figure S5D). Collectively, these data are consistent with the notion that Notch activity functions at the EEP stage to regulate asymmetric cell division of EEPs and consequently EE subtype specification.

TF Code Analysis for EE Subtype Clusters

Cell fate specification and maintenance are usually regulated by TFs. Although Notch activity specifies the two classes of EEs, the TFs that mediate the specification of these two classes are unknown. In addition, as EEs can be further classified into 10 different subtypes, the mechanism underlying this cellular diversity of EEs is also unclear. To identify TFs that participate in EE specification, we surveyed the expression of 1,052 DNA-binding motif-containing genes (The *Drosophila* TF Database) from our single-cell RNA-seq data. Using previously described methods, we minimized a set of 14 TFs, whose binary expression states (ON and OFF) alone was sufficient to classify subtype identity for all EE subtype clusters (Figure 6A) (Li et al., 2017). One exception is clusters 1 and 4, as they share an identical TF code. As mentioned earlier, these two clusters are very similar to each other, expressing similar sets of peptide hormones, although they slightly differ in peptide hormone co-expression frequency patterns (Figures 3C and 3D).

Among the 14 identified TFs, Ptx1 is a paired-type homeobox gene that has been reported to be enriched in R3 (Dutta et al., 2015; Vorbrüggen et al., 1997), and here it was also identified as an EE subtype marker for I-m (C0), II-m1 (C1), and II-m2 (C4) (Figure 6A). Drumstick (drm), a member of the Odd-skipped (Odd) family, has been reported to regulate pattern formation of *Drosophila* hindgut and expressed in the posterior-most region of midgut (Green et al., 2002; Uddin et al., 2011), which is in line with its expression in I-p^{AstA} (C6) and II-p (C7) that are located in R4c and R5, neighboring the hindgut (Figure 6A).

Mirr and Ptx1 Respectively Specify Tk$^+$ Class II and AstC$^+$ Class I EEs

Since the expression pattern of peptide hormones varied among different EE subtypes, we hypothesized that these TFs may participate in the transcriptional regulation of cluster-specific peptide hormone genes. To test this hypothesis, we individually knocked down each of these 14 TFs in EEs by pros-V1-Gal4 and performed qPCR analysis for the expression of the 14 peptide hormone genes.

As shown in Figures 6 and 7, for the majority of the TFs tested, their knockdown caused decreased expression of one or more peptide hormone genes. Interestingly, we identified two TFs that regulate the expression of Tk or AstC. Knocking down *Ptx1* specifically decreased AstC expression, whereas knocking down *mirr* specifically decreased Tk expression (Figures 6B and 6C). Immunostaining with antibodies against TK or AstC also confirmed that depletion of *mirr* globally decreased the number of Tk$^+$ cells along the whole midgut (Figures 6D–6F) and depletion of *Ptx1* led to significant loss of AstC$^+$ cells (Figures 6G–6I).

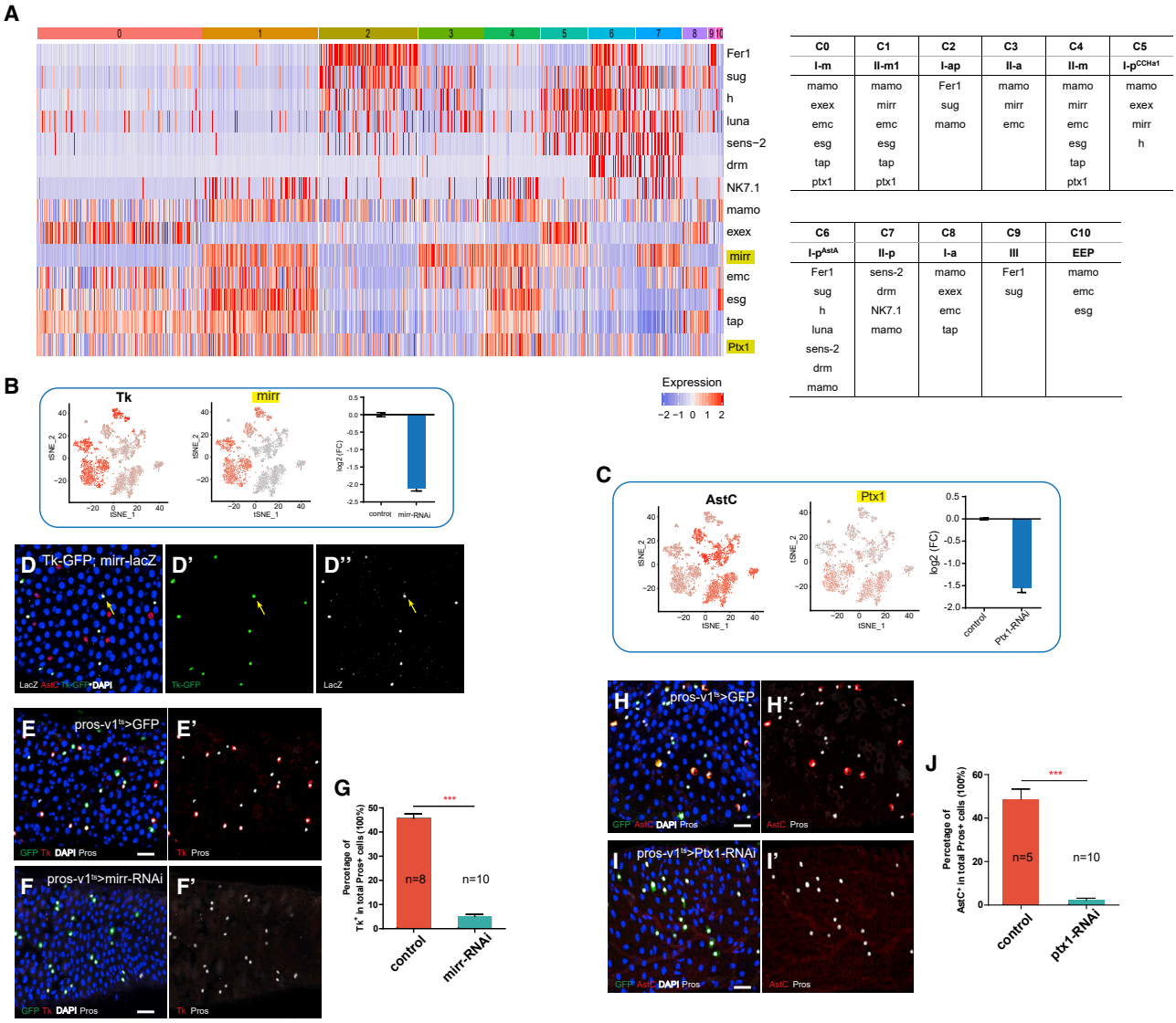
Consistent with a role for *mirr* in specifying Tk$^+$ EEs, *mirr* was found to be specifically expressed in all Tk$^+$ EE subtypes that include II-m1(C1), II-a (C3), II-m2 (C4), and II-p (C7) (Figure 6B). Using a LacZ reporter line of *mirr*, we further confirmed that *mirr* was specifically expressed in Tk$^+$ EEs (yellow arrow) but not AstC EEs (Figure 6D). Interestingly, the expression of *Ptx1* was not restricted to AstC$^+$ EE clusters. It was weakly and generally expressed in all EE clusters, with a relatively higher level in I-m (C0), II-m1(C1), and II-m2 (C4), the EEs in R3 (Figure 6C). Consistent with this expression pattern for *Ptx1*, its knockdown also caused decreased expression of several peptide hormones that are specifically expressed in class II Tk$^+$ EEs at R3, such as NPF and Nplp2 (Figures 7A and 7G).

Collectively, these results suggest that Mirr and Ptx1 participate in the selection and specification of two major EE subtypes, Tk$^+$ class II EEs and AstC$^+$ class I EEs.

TFs that Define Regional EE Identities

The above RNAi screen also identified several TFs that affect the expression of region-specific peptide hormones. As different EE clusters generally show specificity in regional distribution, we argue that the region-specific expression of peptide hormones could be regulated by regionally expressed TFs. To test this hypothesis, we systematically analyzed the expression co-relation between all TFs detected in our scRNA-seq and the 14 peptide hormones and identified a set of TFs whose expression is closely related to the expression of individual peptide hormones (Figure S6). Not surprisingly, 11 identified peptide-hormone-related TFs were also EE subtype signature TFs described earlier. By combining qPCR results with the co-relation map, we found that expression of most peptide hormones could be significantly affected by knocking down their closely related TFs. For example, Fer1 was mainly expressed in CCHa2 expressing cells, and its knocking down significantly downregulated CCHa2 expression (Figure 7C). We also noticed that certain peptide hormones could be modulated by more than one TF, and one TF could affect the expression of several different peptide hormones. For example, expression of Tk, NPF, and DH31 were all downregulated upon depletion of *mirr* (Figures 6B, 7F, and 7G). Orcokinin expression could be regulated by several TFs including Ptx1, Mamo, Exex, and Mirr. Of note, Mirr might negatively regulate Orcokinin expression, as Mirr expression was detected mainly in Orcokinin^{low} EE clusters and its depletion caused significant upregulation of Orcokinin (Figure 7E).

The expression of TFs and their requirements on peptide hormones also showed regional specificity, such as *esg*, which was expressed specifically in R3 EEs and specifically affected the expression of R3-specific peptide hormones NPF and Nplp2 (Figures 7A and 7F). Another example is *drm*. It is required for hindgut specification and specifically expressed in the posterior-most part of midgut (Green et al., 2002; Uddin et al., 2011).



C0	C1	C2	C3	C4	C5
I-m	**II-m1**	**I-ap**	**II-a**	**II-m**	**I-p^{CCHa1}**
mamo	mamo	Fer1	mamo	mamo	mamo
exex	mirr	sug	mirr	mirr	exex
emc	emc	mamo	emc	emc	mirr
esg	esg			esg	h
tap	tap			tap	
ptx1	ptx1			ptx1	

C6	C7	C8	C9	C10
I-p^{AstA}	**II-p**	**I-a**	**III**	**EEP**
Fer1	sens-2	mamo	Fer1	mamo
sug	drm	exex	sug	emc
h	NK7.1	emc		esg
luna	mamo	tap		
sens-2				
drm				
mamo				

Figure 6. Analysis of EE Subtype TF Codes Reveals Mirr and Ptx1 that Respectively Specify Tk$^+$ Class II and AstC$^+$ Class I EEs

(A) Heatmap of TF genes that construct minimal combinatorial codes for EE subtype specification. Signature TFs for each EE subtype are also listed on the right-hand side of the heatmap.

(B) Expression patterns of Tk and mirr. qRT-PCR results of Tk mRNA level after depletion of *mirr* using *pros-V1-GAL4^{ts}*. Three independent replicates were carried out for each sample.

(C) Expression patterns of AstC and Ptx1. qRT-PCR results of AstC mRNA level after depletion of Ptx1 using *pros-v1-GAL4^{ts}*. Three independent replicates were carried out for each sample.

(D) *In vivo* immunostaining of mirr-lacZ. Mirr is specifically co-localized with Tk producing EE cells (yellow arrow) and is mutually exclusive with AstC.

(E and F) Compared to control (E), conditional knocking down *mirr* in *Pros-V1$^+$* EE cells for 5 days leads to significant decrease of Tk$^+$ cells (F).

(G) Quantification of Tk$^+$ EE percentages in total EEs. Mean ± SEM; n = 8 for *pros-V1^{ts} > GFP* midguts, and n = 10 for *pros-V1>mirr-RNAi* midguts. ***p < 0. 001 (Student's t test).

(H and I) Compared to control (H), conditional knocking down *Ptx1* in *Pros-V1$^+$* cells for 5 days resulted in significant decrease of AstC$^+$ cells (I).

(J) Quantification of AstC$^+$ EE percentages in total EEs. Mean ± SEM; n = 5 for *pros-V1^{ts} > GFP* midguts, and n = 10 for *pros-V1>mirr-RNAi* midguts. ***p < 0. 001 (Student's t test).

Scale bars, 20 μm.

See also Figure S6 and Tables S1 and S2.

Its depletion significantly downregulated ITP expression (Figure 7D), a hormone mainly expressed in II-p (C7), and participates in ion transport in hindgut (Nässel and Winther, 2010). These observations strongly indicate a role for II-p EEs in regulating hindgut function via paracrine ITP signaling. In some cases, the subtype specificity for some TFs occurred at a

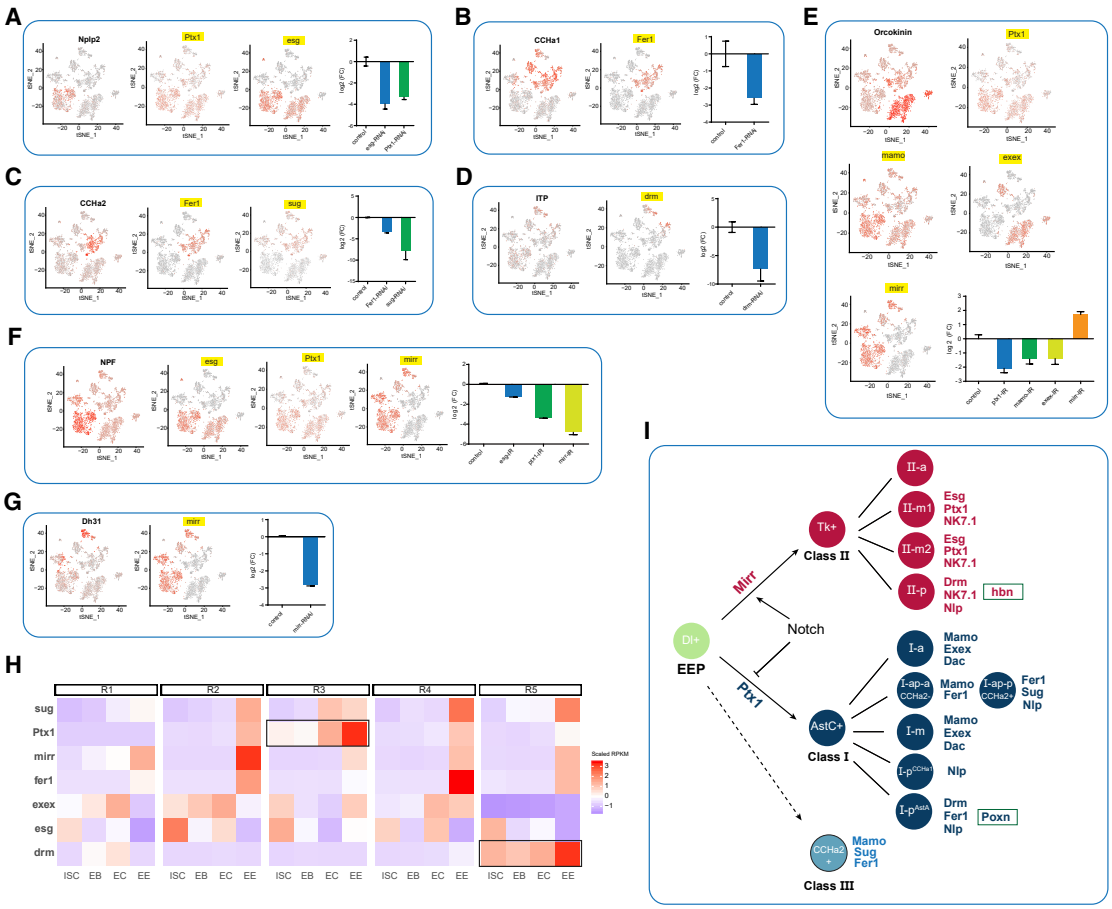


Figure 7. Region-Specific TFs Control EE Regional Identity

(A) Knocking down Ptx1 and esg, two R3 enriched transcription factors, leads to significant decrease of Nplp2 expression.

(B) Fer1 is specifically expressed in class I and III EEs, and knocking down Fer1 significantly downregulates CCHa1 expression.

(C) Expression of Fer1 and sug is highly similar to CCHa2, and their depletion downregulates CCHa2 expression.

(D) Both of ITP and drm are mainly expressed in posterior-most region of midgut, and knocking down drm downregulates ITP expression.

(E) Ptx1, mamo, and exex are required for expression of orcokinin, while knocking down mirr upregulates orcokinin level.

(F) Knocking down mirr leads to significant downregulation of DH31 expression.

(G) Knocking down esg, Ptx1, or mirr leads to downregulation of NPF.

Three independent replicates were carried out for each sample.

(H) Expression of peptide hormone-regulating TFs in all gut epithelium cell types (ISC, enteroblast [EB], enterocyte [EC], EE) of different midgut regions (R1–R5). Expression of Ptx1 is enriched in R3 region of most cell types, and Drm is enriched in R5 region of all cell types. These two TFs are highlighted in red squares.

(I) A model for EE subtype specification, classification, and associated transcription factor code. TFs in green square are potential repressors of the related EE subtypes.

See also Figures S5 and S6 and Tables S1 and S2.

class-specific level. Despite CCHa1 being expressed in both class I (Iap, Ip^{CCHa1}, Ip^{AstA}) and class II (IIa, IIp) EE cells, depletion of Fer1 only altered CCHa1 expression in class I cells, as Fer1 is only expressed in $AstC^{+}$ cells (Figure 7B).

In addition to the 14 cluster signature TFs, we also functionally tested 7 additional TFs that showed high co-relation scores with hormone peptides. Among them, several highly co-related TF-peptide pairs, including *NK7.1*-Tk and *dac*-Orcokinin, showed significant regulatory relationships (Figures S6C and S6D). Depletion of *nlp*, a generally expressed TF in different EE subtypes, leads to downregulation of both CCHa1 and ITP (Figures S6G and S6H). We also observed several possible transcriptional repressors for peptide hormones. As is shown in Figures S6E and S6F, knocking down *poxn* or *hbn* caused increased expression of AstA and DH31, respectively. All the qPCR results from the RNAi screen were summarized in Table S2.

By combining the co-expression results between TFs and peptide hormones and between TFs and EE subtypes, along with the results from the RNAi screen, we were able to allocate multiple TFs that contribute to the regional diversity of EEs. For example, Mamo and Exex help to define I-a subtype identity, Esg and Ptx1 cooperatively define II-m1 and II-m2 subtypes, and Drm contributes to both $I\text{-}p^{AstA}$ and II-p subtype identity.

Interestingly, the I-m subtype is not regulated by Esg; instead, it is regulated by Mamo and Exex. By analyzing the gut regional RNA-seq data published previously (Dutta et al., 2015), we found that among the identified regional TFs, Ptx1 and Drm showed relative expression specificity in most types of epithelial cells found in their local regions (R3 and R5, respectively), whereas Sug, Mirr, and Fer1 showed relative expression specificity in EEs only (Figure 7H). Our results collectively suggest that EE cellular diversity is collectively regulated by both class-specific TFs and region-specific TFs, and the specificity of later ones sometimes can be applied to other cell types found in the local region. The identified TFs that may contribute to EE subtype identity were summarized in Figure 7I. As our single-cell method detects only about 1,350 genes per cell, the less abundantly expressed TFs could be missed in our study. As such, the TFs responsible for some EE subtypes are still missing. Nevertheless, our results provide an important foundation and framework for complete characterization of EE cell diversity and TF code in the future.

DISCUSSION

Using single-cell transcriptomics in combination with a collection of reporter lines, here we have provided a comprehensive characterization of the EE population in the entire midgut of adult *Drosophila*. In addition to the two major classes of EEs that respectively produce TK and AstC peptide hormones, we identify a third class of EEs that reside only in the anterior midgut (R2) and produce sNPF and CCHa2. We identify 10 EE subtypes that generally show region-specific distributions. In addition, functional screens with subtype-specific TFs have revealed class- and region-specific TFs that regulate subtype specification. Our single-cell data should serve as an important resource for further understanding the differentiation, regulation, and function of EEs using *Drosophila* midgut as a genetic model system.

Our single-cell data reveal 14 classes of peptide hormone genes that are expressed in EEs, compared to the previously known 10 classes (Chen et al., 2016a; Reiher et al., 2011; Veenstra and Ida, 2014; Veenstra et al., 2008). The midgut expression patterns of all these peptide hormones, including several peptide hormones whose gut expression patterns have not been clearly defined, such as Gbp5, ITP, and Nplp2 as well as sNPF, are also determined. As EEs perform their endocrine function by secreting various peptide hormones, the types of peptide hormones that they produced are usually used to classify EE subtypes in mammals. Indeed, the exclusive expression pattern of Tk and AstC is sufficient to distinguish between class I EEs and class II EEs. However, although different EE subtypes show distinct peptide hormone expression profiles, the types of peptide hormone expressed and the EE subtypes are not strictly correlated. In fact, the peptide hormone co-expression patterns are highly variable among individual EEs, even for EEs that belong to the same cluster or subtype. For example, for the II-m (C4) subtype, although they commonly produce Tk and NPF, their expression for Mip, Nplp2, and CCAP is highly variable. The external stimuli, such as stress and microbiota, may have an impact on the expression status of these variable peptide hormone genes. Alternatively, EEs could be plastic and change their peptide hormone expression profiles with age. Recent studies demonstrate that the mammalian EEs are plastic and can switch their hormone profiles as they differentiate and migrate upward along the crypt-villus axis (Beumer et al., 2018; Park et al., 2016).

One major limit associated with the scRNA-seq technology is that the spatial information of the cells is lost during tissue dissociation. In a way to overcome this limit, we have developed a RSGE algorithm based on the region- and cell-type-specific transcriptome database from flygut-seq (Dutta et al., 2015). As confirmed, for the various peptide hormone markers, including GAL4 knockin lines and antibodies, this algorithm has allowed us to generate a reliable distribution map a for all the EE subtype clusters long the length of the midgut. The determination of the spatial distribution of EE subtypes should greatly facilitate the understanding of their regulation and function. For instance, DH31 and ITP expressing EEs are found be located in the posterior-most region of the midgut, and their location is clearly consistent with their known function: DH31 is known to regulate fluid secretion in Malpighian tubules, and ITP is known to regulate ion transport in hindgut (Nässel and Winther, 2010). As regional difference for a common cell type is likely a general phenomenon in diverse tissues of many organisms, our algorithm here could provide an example of possible approaches for acquiring the lost spatial information of cells when conducting this type of single-cell analysis.

By analyzing the TF code for the EE subtypes followed by functional screen, our study has identified a number of TFs that participate in the specification of EE subtypes, including the class-I- and class-II-specific TFs Mirr and Ptx1 for the two major classes of EEs and region-specific TFs such as Esg, Drm, Fer1, and Sug that define regional EE identity. Previous studies in the posterior midgut have revealed that class I and II EEs are specified by differential Notch signaling (Beehler-Evans and Micchelli, 2015; Guo and Ohlstein, 2015; Ohlstein and Spradling, 2007). Here, cell-type specific manipulating of Notch activity allows us to conclude that Notch must function transiently at the progenitor stage, between the two immediate daughters of an EEP, to define the two classes of EEs. As Mirr and Ptx1 are expressed only in differentiated EEs, the sequential activity of Notch and Mirr/Ptx1 indicates that these two TFs act downstream of Notch to specify class I versus class II EE type. The regional diversity of EEs is then further specified by region-specific TFs and possibly impacted by other environmental factors (Figure 7I). We propose that EE cellular diversity is generated by a combination of class-specific and region-specific TFs, with class-specific TFs regulated by Notch signaling and region-specific TFs determined by anterior-posterior body planning during early development. The local EE diversity could also be regulated by environmental changes and age-related cell plasticity, possibilities that remain to be explored in the future.

Collectively, our single-cell data have provided a comprehensive characterization of EE cell diversity and their peptide hormone expression profiles. The TF code analysis also provides insights into EE diversity mechanisms. Our data should greatly facilitate functional annotations of EE subtypes and gut peptide

hormones under diverse physiological and pathological conditions, such as mating, starvation, bacterial infection, and so on. The Perrimon lab recently conducted single-cell transcriptomes for all types of midgut cells using the inDrop method. As EEs only represent a small fraction of total cells analyzed, their analysis primarily focused on progenitor cells and enterocytes (Hung et al., 2018). Therefore, our data and their scRNA-seq data should serve as complementary resources for understanding *Drosophila* gut cells. An online searchable database has been established to facilitate the use of these single-cell data (https://xilab.shinyapps.io/database/).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Fly Stocks and Cultivation
- METHOD DETAILS
  - Generation of Transgenic Fly
  - Immunostaining
  - FACS and Single-Cell RNA-Sequencing (scRNA-seq)
  - Quantitative PCR (qPCR) Analysis of Peptide Hormones Expression
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - ScRNA-Seq Data Analysis
  - Region-Specific Gene Enrichment (RSGE) Algorithm
  - Statistical Analysis
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2019.11.048.

ACKNOWLEDGMENTS

We thank Dr. Dick Nassel (Stockholm University) for providing anti-Tk antibody; Dr. Jan-Adrianus Veenstra (Université de Bordeaux) for anti-AstC antibody; Dr. Fernando Jiménez Díaz-Benjumea (CBMSO) for anti-CCAP antibody; the Bloomington Drosophila Stock Center (BDSC), the Tsinghua Fly Center, and Development Studies Hybridoma Bank (DSHB) for fly strains and antibodies; and Dr. Yanzhuang Wang (University of Michigan) and members of the Xi laboratory for helpful discussions. This work was supported by the National Key Research and Development Program of China (2017YFA0103602) and the National Basic Research Program of China (2014CB850002) from the Chinese Ministry of Science and Technology.

AUTHOR CONTRIBUTIONS

Conceptualization, X.G., F.Y., C.Y., and R.X.; Methodology, X.G., C.Y., F.Y., Y.Z., H.H., J.W., and T.C.; Software, Y.Z.; Formal Analysis, Y.Z., H.H., and J.W.; Investigation, X.G., C.Y., Y.Z., and F.Y.; Writing – Original Draft, X.G., C.Y., F.Y., Y.Z., and R.X.; Writing – Review & Editing, X.G. and R.X.; Funding Acquisition, R.X.; Resources, B.D. and Y.R.; Supervision, R.X.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 5, 2019
Revised: October 13, 2019
Accepted: November 12, 2019
Published: December 17, 2019

REFERENCES

Adryan, B., and Teichmann, S.A. (2006). FlyTF: a systematic review of site-specific transcription factors in the fruit fly Drosophila melanogaster. Bioinformatics *22*, 1532–1533.

Amcheslavsky, A., Song, W., Li, Q., Nie, Y., Bragatto, I., Ferrandon, D., Perrimon, N., and Ip, Y.T. (2014). Enteroendocrine cells support intestinal stem-cell-mediated homeostasis in Drosophila. Cell Rep. *9*, 32–39.

Ameku, T., Yoshinari, Y., Texada, M.J., Kondo, S., Amezawa, K., Yoshizaki, G., Shimada-Niwa, Y., and Niwa, R. (2018). Midgut-derived neuropeptide F controls germline stem cell proliferation in a mating-dependent manner. PLoS Biol. *16*, e2005004.

Beebe, K., Park, D., Taghert, P.H., and Micchelli, C.A. (2015). The Drosophila Prosecretory Transcription Factor dimmed Is Dynamically Regulated in Adult Enteroendocrine Cells and Protects Against Gram-Negative Infection. G3 (Bethesda) *5*, 1517–1524.

Beehler-Evans, R., and Micchelli, C.A. (2015). Generation of enteroendocrine cell diversity in midgut stem cell lineages. Development *142*, 654–664.

Beumer, J., Artegiani, B., Post, Y., Reimann, F., Gribble, F., Nguyen, T.N., Zeng, H., Van den Born, M., Van Es, J.H., and Clevers, H. (2018). Enteroendocrine cells switch hormone expression along the crypt-to-villus BMP signalling gradient. Nat. Cell Biol. *20*, 909–916.

Biteau, B., and Jasper, H. (2014). Slit/Robo signaling regulates cell fate decisions in the intestinal stem cell lineage of Drosophila. Cell Rep. *7*, 1867–1875.

Buchon, N., Osman, D., David, F.P., Fang, H.Y., Boquete, J.P., Deplancke, B., and Lemaitre, B. (2013). Morphological and molecular characterization of adult midgut compartmentalization in Drosophila. Cell Rep. *3*, 1725–1738.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

Cabrero, P., Radford, J.C., Broderick, K.E., Costes, L., Veenstra, J.A., Spana, E.P., Davies, S.A., and Dow, J.A.T. (2002). The Dh gene of Drosophila melanogaster encodes a diuretic peptide that acts through cyclic AMP. J. Exp. Biol. *205*, 3799–3807.

Chen, J., Kim, S.M., and Kwon, J.Y. (2016a). A Systematic Analysis of Drosophila Regulatory Peptide Expression in Enteroendocrine Cells. Mol. Cells *39*, 358–366.

Chen, J., Xu, N., Huang, H., Cai, T., and Xi, R. (2016b). A feedback amplification loop between stem cells and their progeny promotes tissue regeneration and tumorigenesis. eLife *5*, e14330.

Chen, J., Xu, N., Wang, C., Huang, P., Huang, H., Jin, Z., Yu, Z., Cai, T., Jiao, R., and Xi, R. (2018). Transient Scute activation via a self-stimulatory loop directs enteroendocrine cell pair specification from self-renewing intestinal stem cells. Nat. Cell Biol. *20*, 152–161.

Deng, B., Li, Q., Liu, X., Cao, Y., Li, B., Qian, Y., Xu, R., Mao, R., Zhou, E., Zhang, W., et al. (2019). Chemoconnectomics: Mapping Chemical Transmission in Drosophila. Neuron *101*, 876–893.e4.

Doupé, D.P., Marshall, O.J., Dayton, H., Brand, A.H., and Perrimon, N. (2018). *Drosophila* intestinal stem and progenitor cells are major sources and regulators of homeostatic niche signals. Proc. Natl. Acad. Sci. USA *115*, 12218–12223.

Dus, M., Lai, J.S., Gunapala, K.M., Min, S., Tayler, T.D., Hergarden, A.C., Geraud, E., Joseph, C.M., and Suh, G.S. (2015). Nutrient Sensor in the Brain Directs the Action of the Brain-Gut Axis in Drosophila. Neuron *87*, 139–151.

Dutta, D., Xiang, J., and Edgar, B.A. (2013). RNA Expression Profiling from FACS-Isolated Cells of the Drosophila Intestine. Curr. Protoc. Stem Cell Biol. *27*, 2F.2.1–2F.2.12.

Dutta, D., Dobson, A.J., Houtz, P.L., Gläßer, C., Revah, J., Korzelius, J., Patel, P.H., Edgar, B.A., and Buchon, N. (2015). Regional Cell-Specific Transcriptome Mapping Reveals Regulatory Complexity in the Adult Drosophila Midgut. Cell Rep. *12*, 346–358.

Furness, J.B., Rivera, L.R., Cho, H.J., Bravo, D.M., and Callaghan, B. (2013). The gut as a sensory organ. Nat. Rev. Gastroenterol. Hepatol. *10*, 729–740.

Gehart, H., van Es, J.H., Hamer, K., Beumer, J., Kretzschmar, K., Dekkers, J.F., Rios, A., and Clevers, H. (2019). Identification of Enteroendocrine Regulators by Real-Time Single-Cell Differentiation Mapping. Cell *176*, 1158–1173.e16.

Green, R.B., Hatini, V., Johansen, K.A., Liu, X.-J., and Lengyel, J.A. (2002). Drumstick is a zinc finger protein that antagonizes Lines to control patterning and morphogenesis of the Drosophila hindgut. Development *129*, 3645–3656.

Gribble, F.M., and Reimann, F. (2016). Enteroendocrine Cells: Chemosensors in the Intestinal Epithelium. Annu. Rev. Physiol. *78*, 277–299.

Guo, Z., and Ohlstein, B. (2015). Stem cell regulation. Bidirectional Notch signaling regulates Drosophila intestinal stem cell multipotency. Science *350*, aab0988.

Guo, X., Huang, H., Yang, Z., Cai, T., and Xi, R. (2019). Division of Labor: Roles of Groucho and CtBP in Notch-Mediated Lateral Inhibition that Controls Intestinal Stem Cell Differentiation in Drosophila. Stem Cell Reports *12*, 1007–1023.

Haber, A.L., Biton, M., Rogel, N., Herbst, R.H., Shekhar, K., Smillie, C., Burgin, G., Delorey, T.M., Howitt, M.R., Katz, Y., et al. (2017). A single-cell survey of the small intestinal epithelium. Nature *551*, 333–339.

Hung, R., Kirchner, Y.H.R., Li, F., Xu, C., Comjean, A., Tattikota, S., Song, W., Sui, S., and Perrimon, N. (2018). A cell atlas of the adult Drosophila midgut. bioRxiv. https://doi.org/10.1101/410423.

Jensen, J., Pedersen, E.E., Galante, P., Hald, J., Heller, R.S., Ishibashi, M., Kageyama, R., Guillemot, F., Serup, P., and Madsen, O.D. (2000). Control of endodermal endocrine development by Hes-1. Nat. Genet. *24*, 36–44.

Kumar, P., Tan, Y., and Cahan, P. (2017). Understanding development and stem cells using single cell-based analyses of gene expression. Development *144*, 17–32.

Li, H., Horns, F., Wu, B., Xie, Q., Li, J., Li, T., Luginbuhl, D.J., Quake, S.R., and Luo, L. (2017). Classifying Drosophila Olfactory Projection Neuron Subtypes by Single-Cell RNA Sequencing. Cell *171*, 1206–1220.e22.

Lin, G., Xu, N., and Xi, R. (2008). Paracrine Wingless signalling controls self-renewal of Drosophila intestinal stem cells. Nature *455*, 1119–1123.

Marianes, A., and Spradling, A.C. (2013). Physiological and stem cell compartmentalization within the Drosophila midgut. eLife *2*, e00886.

Meng, F.W., and Biteau, B. (2015). A Sox Transcription Factor Is a Critical Regulator of Adult Stem Cell Proliferation in the Drosophila Intestine. Cell Rep. *13*, 906–914.

Micchelli, C.A., and Perrimon, N. (2006). Evidence that stem cells reside in the adult Drosophila midgut epithelium. Nature *439*, 475–479.

Min, S., Chae, H.-S., Jang, Y.-H., Choi, S., Lee, S., Jeong, Y.T., Jones, W.D., Moon, S.J., Kim, Y.J., and Chung, J. (2016). Identification of a Peptidergic Pathway Critical to Satiety Responses in Drosophila. Curr. Biol. *26*, 814–820.

Nässel, D.R., and Winther, A.M. (2010). Drosophila neuropeptides in regulation of physiology and behavior. Prog. Neurobiol. *92*, 42–104.

Ohlstein, B., and Spradling, A. (2006). The adult Drosophila posterior midgut is maintained by pluripotent stem cells. Nature *439*, 470–474.

Ohlstein, B., and Spradling, A. (2007). Multipotent Drosophila intestinal stem cells specify daughter cell fates by differential notch signaling. Science *315*, 988–992.

Park, J.-H., Chen, J., Jang, S., Ahn, T.J., Kang, K., Choi, M.S., and Kwon, J.Y. (2016). A subset of enteroendocrine cells is activated by amino acids in the Drosophila midgut. FEBS Lett. *590*, 493–500.

Pauls, D., Chen, J., Reiher, W., Vanselow, J.T., Schlosser, A., Kahnt, J., and Wegener, C. (2014). Peptidomics and processing of regulatory peptides in the fruit fly Drosophila. EuPA Open Proteom. *3*, 114–127.

Reiher, W., Shirras, C., Kahnt, J., Baumeister, S., Isaac, R.E., and Wegener, C. (2011). Peptidomics and peptide hormone processing in the Drosophila midgut. J. Proteome Res. *10*, 1881–1892.

Scopelliti, A., Bauer, C., Yu, Y., Zhang, T., Kruspig, B., Murphy, D.J., Vidal, M., Maddocks, O.D.K., and Cordero, J.B. (2019). A Neuronal Relay Mediates a Nutrient Responsive Gut/Fat Body Axis Regulating Energy Homeostasis in Adult Drosophila. Cell Metab. *29*, 269–284.e10.

Siviter, R.J., Coast, G.M., Winther, Å.M.E., Nachman, R.J., Taylor, C.A.M., Shirras, A.D., Coates, D., Isaac, R.E., and Nässel, D.R. (2000). Expression and functional characterization of a Drosophila neuropeptide precursor with homology to mammalian preprotachykinin A. J. Biol. Chem. *275*, 23273–23280.

Song, W., Veenstra, J.A., and Perrimon, N. (2014). Control of lipid metabolism by tachykinin in Drosophila. Cell Rep. *9*, 40–47.

Steiner, D.F. (1998). The proprotein convertases. Curr. Opin. Chem. Biol. *2*, 31–39.

Uddin, S.N., Yano, M., and Murakami, R. (2011). The drumstick gene acts cell-non-autonomously and triggers specification of the small intestine in the Drosophila hindgut. Int. J. Dev. Biol. *55*, 945–952.

Veenstra, J.A., and Ida, T. (2014). More Drosophila enteroendocrine peptides: Orcokinin B and the CCHamides 1 and 2. Cell Tissue Res. *357*, 607–621.

Veenstra, J.A., Agricola, H.J., and Sellami, A. (2008). Regulatory peptides in fruit fly midgut. Cell Tissue Res. *334*, 499–516.

Vorbrüggen, G., Constien, R., Zilian, O., Wimmer, E.A., Dowe, G., Taubert, H., Noll, M., and Jäckle, H. (1997). Embryonic expression and characterization of a Ptx1 homolog in Drosophila. Mech. Dev. *68*, 139–147.

Wang, C., Guo, X., Dou, K., Chen, H., and Xi, R. (2015). Ttk69 acts as a master repressor of enteroendocrine cell specification in Drosophila intestinal stem cell lineages. Development *142*, 3321–3331.

Yang, Q., Bermingham, N.A., Finegold, M.J., and Zoghbi, H.Y. (2001). Requirement of Math1 for secretory cell lineage commitment in the mouse intestine. Science *294*, 2155–2158.

Yin, C., and Xi, R. (2018). A Phyllopod-Mediated Feedback Loop Promotes Intestinal Stem Cell Enteroendocrine Commitment in Drosophila. Stem Cell Reports *10*, 43–57.

Zeng, X., and Hou, S.X. (2015). Enteroendocrine cells are generated from stem cells through a distinct progenitor in the adult Drosophila posterior midgut. Development *142*, 644–653.

Zhai, Z., Kondo, S., Ha, N., Boquete, J.P., Brunner, M., Ueda, R., and Lemaitre, B. (2015). Accumulation of differentiating intestinal stem cell progenies drives tumorigenesis. Nat. Commun. *6*, 10219.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-Pros	Developmental Studies Hybridoma Bank (DSHB)	Cat#MR1A; RRID: AB_528440
Rabbit polyclonal anti-Tachykinin	Siviter et al., 2000	RRID: AB_2569591
Rabbit polyclonal anti-AstC	Veenstra et al., 2008	RRID: AB_2753141
Rabbit polyclonal anti-LacZ	Cappel	Cat#0855976; RRID: AB_2334934
Chemicals, Peptides, and Recombinant Proteins		
Elastase	Sigma-Aldrich	Cat#E0258
RNAiso Plus	Takara	Cat#9109
SYBR Green PCR Master Mix	Applied Biosystems	Cat#4309155
Critical Commercial Assays		
Direct-zol RNA MiniPrepkit	Zymo Research	Cat#R2050
5X All-In-One RT Master Mix	abm	Cat#G485
Deposited Data		
Raw and processed EE scRNA-seq data	This study	GSE132274
Experimental Models: Organisms/Strains		
ProsV1-Gal4, UAS-GFP	from Bruce Edgar	
UAS-Notch-RNAi	BDSC	7078
UAS-Nicd	from Ting Xie	
UAS-RFP, lexAop-GFP	BDSC	32229
esg-GFP	from Lynn Cooley	esg-P01986
Dtk-GFP	This study	NA
mirrr-lacZ	BDSC	10880
Fer1-IR	Tsinghua Stock Center	TH01959.N
		HMC03073
Sug-IR	Tsinghua Stock Center	TH02012.N
		HMC03866
Hairy-IR	Tsinghua Stock Center	THU1541
		HMS01313
Luna-IR	Tsinghua Stock Center	THU2473
		HMJ23655
Sens-IR	Tsinghua Stock Center	TH02005.N
		HMC03997
Drm-IR	Tsinghua Stock Center	THU4947
		HMJ02120
Mamo-IR	Tsinghua Stock Center	TH02114.N
		HMS02823
Exex-IR	Tsinghua Stock Center	TH04061.N
		HMC04898
Mirr-IR	Tsinghua Stock Center	THU2284
		HMC06139
Emc-IR	Tsinghua Stock Center	THU2364
		GL00724

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Esg-IR	Tsinghua Stock Center	TH03106.N2
		HMS02538
Tap-IR	Tsinghua Stock Center	THU1614
		HMS01395
Ptx1-IR	Tsinghua Stock Center	TH01869.N
		HMC04809
Hbn-IR	Tsinghua Stock Center	THU2283
		JF02195
Hth-IR	Tsinghua Stock Center	THU2741
		JF02733
Odd-IR	Vienna Drosophila Resource Center	v107257
		KK105970
Dac-IR	Tsinghua Stock Center	THU1651
		HMS01435
Fru-IR	Tsinghua Stock Center	THU4021
		JF01182
NK7.1-IR	Tsinghua Stock Center	THU2285
		JF02197
Poxn-IR	Tsinghua Stock Center	THU2237
		JF02136
Tup-IR	Tsinghua Stock Center	TH02017.N
		HMC03317
CCHa2-lexA	Deng et al., 2019	NA
AstC-Gal4	Deng et al., 2019	NA
dTK-Gal4	Deng et al., 2019	NA
AstA-Gal4	Deng et al., 2019	NA
DH31-Gal4	Deng et al., 2019	NA
CCHa1-Gal4	Deng et al., 2019	NA
CCHa2-Gal4	Deng et al., 2019	NA
Mip-Gal4	Deng et al., 2019	NA
NPFR-Gal4	Deng et al., 2019	NA
CG30340-Gal4	Deng et al., 2019	NA
CG32547-Gal4	Deng et al., 2019	NA
CCHa1-R-Gal4	Deng et al., 2019	NA
ITP-Gal4	Deng et al., 2019	NA
sNPF-Gal4	Deng et al., 2019	NA
Orcokinin-Gal4	Deng et al., 2019	NA
Oligonucleotides		
Oligos used in this study are listed in Table S1	This study	NA
Software and Algorithms		
Region specific gene enrichment (RSGE) algorithm code	This study	https://github.com/fentouxungui/Fly-Midgut-EEs-scRNAseq
Cell Ranger	10x Genomics	https://github.com/10XGenomics/cellranger
R v3.6.1	R Development Core Team	https://www.r-project.org/
Rstudio v1.1.456	RStudio, Inc	https://rstudio.com/
Seurat v2.3.4	Butler et al., 2018	https://satijalab.org/seurat/
ggplot2 v3.2.1	Hadley Wickham	https://cran.r-project.org

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
corrgram	Kevin Wright	https://cran.r-project.org
FACSdiva software	BD Biosciences	https://www.bdbiosciences.com/en-us
ABI PRISM 7500 Fast Real-Time PCR System	Applied Biosystems	https://7500-fast-real-time-pcr-system.html
FlyPN	felixhorns	https://github.com/felixhorns/FlyPN
Other		
Single-cell data associated online database	This study	https://xilab.shinyapps.io/database/
RNA-seq data of cell types in Drosophila midgut	Dutta et al., 2015	http://flygutseq.buchonlab.com/resources
Drosophila TF Database	Adryan and Teichmann, 2006	http://flytf.gen.cam.ac.uk/
Drosophila Transcriptome	BDGP6.87	iGenomes

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Rongwen Xi (xirongwen@nibs.ac.cn). All unique reagents generated in this study are available from the Lead Contact without restriction.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Fly Stocks and Cultivation

Flies were cultured on standard media with yeast paste added to the food surface. The culture temperature was 25°C unless otherwise noted. Strains used in this study were listed in Key Resources Table.

Crosses for RNAi knockdown and overexpression experiments were performed at 18°C before adulthood, and F1 generation flies with the correct genotype were transferred to 29°C for 7 days before dissection.

METHOD DETAILS

Generation of Transgenic Fly

To generate the Tk-GFP reporter fly, approximately a 2.5 kb of DNA fragment ranging from −47 upstream to +2381 downstream of Tk transcription start site was amplified from the genomic DNA of *Drosophila melanogaster* with the following primers:

TkG-F: 5′-GAAGATCTgggacgactgtcaacctttgcaca-3′;
TkG-R: 5′-GGGGTACCgcgtaccgaacgcgccgagaagtat-3′.

The fragment was then inserted into the MCS site of pStinger vector (DGRC, #1017) using BglII(5′) and KpnI (3′) restriction enzymes. The embryos of w^{1118} flies were used for subsequent microinjection following the standard protocol for P-element-mediated germline transformation. The eclosed flies were then crossed with w^{1118} flies and their progenies with w^{+} (red eyes) were selected for further validation by examining GFP expression in the gut and by co-immunostaining with anti-Tk (Siviter et al., 2000) and anti-Pros antibodies.

Immunostaining

Immunostaining of *Drosophila* midgut was performed as previously described (Lin et al., 2008). Briefly, 10-15 adult female midguts were dissected in ice-cold PBS and fixed in 4% paraformaldehyde for 30 min at room temperature. After dehydration in methanol and rehydration in PBT containing 0.1% Triton X-100, samples were incubated with primary antibodies in 5% NGS-PBT solution overnight at 4°C. Secondary antibodies were incubated at room temperature for 2 hours, followed by 5 min DAPI staining. The intestines were mounted using 70% glycerol and images were captured using Nikon A1-R, Leica SP8 and Zeiss LSM800 confocal microscope. Images were adjusted in Adobe Photoshop and assembled in Adobe Illustrator. To visualize the entire length of the midgut, images shown in Figures 5C, 5E, S4E, S4G, and S4H are results of multiple (7 to 12) images manually stitched together in Adobe Illustrator, and images shown in Figures S1A, S4A–D, S4F, and S4I are automatically stitched images captured by the Nikon A1R microscope. All scale bars indicate 20 μm unless otherwise noted. Antibodies used in this study were listed in Key Resources Table.

FACS and Single-Cell RNA-Sequencing (scRNA-seq)

FACS purification of specific cell types were performed according to the method described previously (Dutta et al., 2013). The line of CG32547-GAL4 > GFP was used to harvest enteroendocrine cells in this study. Briefly, about 200 midguts of female flies aged 5-7 days old were dissected in cold PBS within 1 hour and digested in 1 mg /ml Elastase (Sigma, #E0258) solution at 25°C for

1 hr. Dissociated cells were pelleted at 400 g for 20 min, resuspended in PBS with 0.2% BSA, filtered with 70 μm filters (BD Falcon) and sorted using a FACS Aria III sorter (BD Biosciences). GFP was used to sort EE cell population and PI to distinguish live cells, while cells from w^{1118} midgut was used to set negative fluorescence gate of the GFP panel. GFP^+ EE cells were sorted into PBS solution with 0.2% BSA, and after pelleting (400 g for 20 min) and resuspension, TC20 automated cell counter (BIO-RAD) was used to count for cell number and live/dead cell ratio.

About 8000 live cells (ratio of live cells in the suspension >90%) were added for 10X genomics library preparation following the manufacture's manual. Briefly, single cells and GemCode Gel Beads were encapsulated into droplets by using the GemCode Single Cell Platform, and in these droplets, cell lysis and barcoded reverse transcription were carried out. Next, cDNA was amplified for 12 cycles, generating about 100ng cDNA. Standard process of library preparation for next-generation deep sequencing was carried out afterward, and the library was sequenced on the Illumina X10 system (Novogene).

Quantitative PCR (qPCR) Analysis of Peptide Hormones Expression

$Gal80^{ts}$; ProsV1-Gal4, UAS-GFP strain was used to specifically knock down TFs in EE cells. Crosses were performed at 18°C to avoid premature expression. F1 females with correct genotypes were collected and cultivated at 29°C for 5 days. Afterward, whole midguts were dissected in cold DEPC-PBS and homogenize in 300 μL RNAiso Plus (Takara, # 9109) buffer using glass pestle. 15-20 midguts were used for each sample, and 3 independent replicates were carried out for all the RNAi lines. Total RNA was extracted using Direct-zol RNA MiniPrepkit(Zymo Research, #R2050) according to the manufacturer's instructions, and on-column DNase I treatment was also carried out to avoid genomic DNA contamination. For each sample, 1 ug RNA was used to synthesize cDNA using 5X All-In-One RT Master Mix (abm, #G485) according to the manufacturer's instructions, and qPCR was then carried out using SYBR Green PCR Master Mix (Applied Biosystems, #4309155) on an ABI PRISM 7500 Fast Real-Time PCR System (Applied Biosystems). Expression levels of peptide hormones were normalized to rp49, and qPCR primers used in this study were listed in Table S1.

QUANTIFICATION AND STATISTICAL ANALYSIS

ScRNA-Seq Data Analysis

Raw reads were performed with sample de-multiplexing, barcode processing, and single cell 3′ gene counting by Cell Ranger Software (v2.0). Seurat V2.3 was used for the downstream analysis (Butler et al., 2018), keeping only cells with detected genes ranging from 200 to 3000 and percentage of mitochondrial genes less than 0.15%. Besides, genes expressed in less than 3 cells were also omitted. We used default parameters to normalize expression data and obtained highly variable genes, and then scaled the data and regressed out variations caused by "nUMI" and "percent.mito." The variable genes were used for performing linear dimensional reduction. The top 20 principal components were used for the t-SNE projection and clustering analysis (resolution = 0.4). Top 10 marker genes of each cluster were picked out by default parameters, and a large portion of these signature genes were peptide hormones and receptors. Peptide hormones and receptors with total UMI counted more than 200 in all EE cells were used for further analysis.

To investigate peptide hormones co-expression patterns at single cell level, we set the scaled value 0.5 as the threshold to distinguish whether a peptide is expressed or not. Totally 14 peptide hormones were composed in this step. The barplot of peptides combination was made by R package ggplot2.

To determine TF codes for each EE subtype, we systematically profiled all TFs (The *Drosophila* TF Database) in our scRNA-seq data (Adryan and Teichmann, 2006). To identify minimal sets of TFs that can specify different EE subtypes, we performed an information theory-based analysis by using the Combinatorial Code (https://github.com/felixhorns/FlyPN) (Li et al., 2017). The total UMI counts of each cell were normalized to 1e4, and the cutoff of genes expression ON/OFF was set to log2 (normalized UMI + 1) = 0.3, the minimal TF codes had >0.90 cumulative information. Via this approach, we mapped a list of 14 TFs, and their ON/OFF states could distinguish each EE subtypes except clusters 1 and 4.

To calculate the correlation scores of TFs and peptide hormones, the normalized data from Seurat object was further processed by R package corrgram, and the top 10 most co-related TFs for each peptide hormone were displayed.

Region-Specific Gene Enrichment (RSGE) Algorithm

To evaluate the regional preference of each cluster, we used the bulk EE RNA-seq data of different midgut regions from the Flygut-seq database (http://flygutseq.buchonlab.com/resources) (Dutta et al., 2015). For each region, genes with RPKM > = 3.5 and fold enrichment over other four regions more than 2.5 were profiled, and the top 100 genes according to fold enrichment were selected as the region-specific gene sets to perform subsequent analysis. To evaluate whether cells from scRNA-seq data express certain genes, we set a cutoff with the scaled value of 0.5, and then calculated the percentages of the cells expressing these region-specific genes in each cluster. The summed percentages of the 100 genes of each region were represented as the regional enrichment score for each cluster.

Statistical Analysis

All quantification and qPCR data were presented in the form of mean ± SEM. GraphPad Prism 5 software (GraphPad Software Inc.) was used to calculate p values by unpaired Student's t test, ***indicates $p < 0.001$. The number of intestines for calculation and qPCR replicates were labeled on figures or related figure legends.

DATA AND CODE AVAILABILITY

The raw and processed datasets reported in this study have been deposited in the Gene Expression Omnibus (GEO). The accession number is GEO: GSE132274. The RSGE code generated in this study can be accessed at GitHub (https://github.com/fentouxungui/Fly-Midgut-EEs-scRNAseq).