Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Insect Science

Defining cell types and lineage in the *Drosophila* midgut using single cell transcriptomics

Ruei-Jiun Hung[1,3], Joshua Shing Shun Li[1,3], Yifang Liu[1] and Norbert Perrimon[1,2]

The *Drosophila* midgut has emerged in recent years as a model system to study stem cell renewal and differentiation and tissue homeostasis. Histological, genetic and gene expression studies have provided a wealth of information on gut cell types, regionalization, genes and pathways involved in cell proliferation and differentiation, stem cell renewal, and responses to changes in environmental factors such as the microbiota and nutrients. Here, we review the contribution of single cell transcriptomic methods to our understanding of gut cell type diversity, lineage and behavior.

Addresses
[1] Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA 02115, United States

[2] Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, United States

Corresponding author:
Perrimon, Norbert (perrimon@genetics.med.harvard.edu)
[3] Equal contribution.

Current Opinion in Insect Science 2021, **47**:12–17

This review comes from a themed issue on **Molecular physiology**

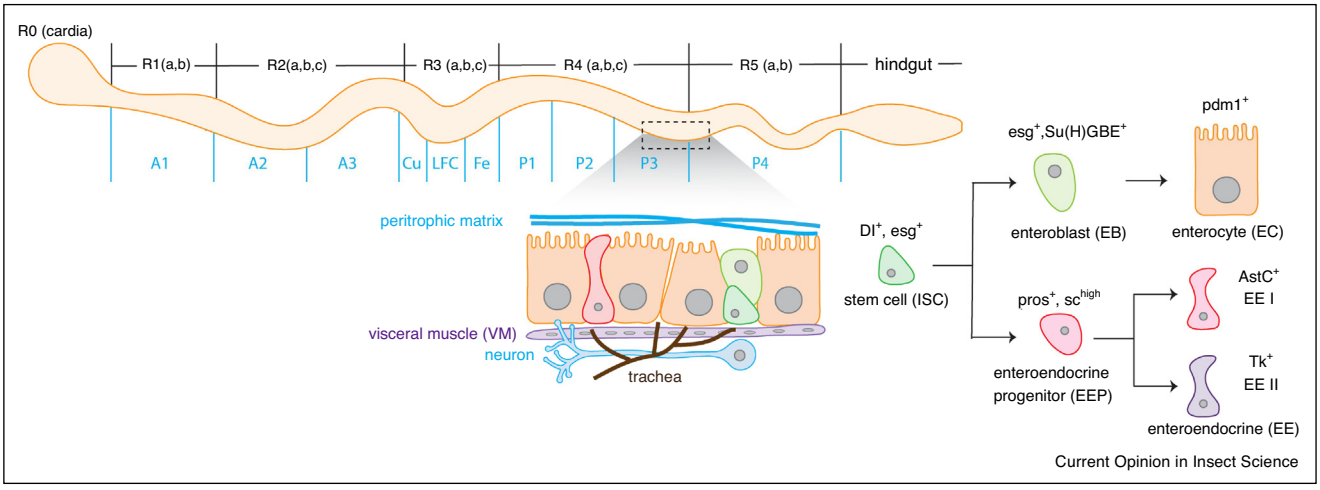
Edited by **Julian Dow** and **Aylin Rodan**

https://doi.org/10.1016/j.cois.2021.02.008

The *Drosophila* adult gut is a complex tissue that is subdivided into three major compartments: foregut, midgut and hindgut. The midgut is the major site for digestion while the foregut and hindgut are involved in food storage and water reabsorption, respectively. The midgut contains self-renewable intestinal stem cells (ISCs) that replenish cells in the midgut over the lifetime of a fly. ISCs differentiate into two diverging transient states, enteroblasts (EBs) or enteroendocrine progenitor cells (EEPs), depending on the level of Notch signaling. High Notch signaling directs ISCs to develop into absorptive enterocytes (ISC → EB → EC), while low Notch signaling toward hormone-secreting enteroendocrine cells (ISC → EEP → EE) (Figure 1) [1–3]. Studies using antibody staining, RNA *in situ* hybridization and gene reporter lines have revealed that EEs are scattered along the gut and express more than 10 gut hormones in adult midgut (Allatostatins (AstA, AstB/Mip, AstC), Tachykinin (Tk), neuropeptide F (NPF), diuretic hormone 31(DH31), CCHa1, CCHa2, Orcokinin B, and Bursicon (Burs) [4–7]. Strikingly, the midgut displays significant regional differences along the antero-posterior axis. Early studies based on the unique orange fluorescence of epithelial cells in copper-fed flies divided the midgut into the anterior, middle and posterior midgut [8]. More recent studies based on electron microscopy, cell lineage and transcriptome analyses of dissected gut sections have led to a further subdivision of the midgut into 10–14 regions [9,10] (Figure 1). Each midgut region is characterized by specific histological and cellular features (microvilli length, lumen width, EC architectures), cell types, stem cell proliferation rate, physical properties such as luminal pH, and specific gene expression profiles that reflect their function [9–11].

Two studies monitored the transcriptome of specific gut subregions, one via Affymetrix microarray and the other via bulk-RNAseq analysis [9,10]. Each subregion was defined by the enriched expression of 50–150 genes. Transcripts encoding enzymes involved in the processing of complex macromolecules are enriched in the anterior part of the gut. In contrast, those involved in the processing of simpler nutrients are more abundant in the posterior midgut, possibly reflecting the sequential breakdown and absorption of food along the gut. The region-specific gene profiles imply distinct roles for the gut subregions. Interestingly, many digestive genes are organized in large genomic clusters and are sequentially expressed in different gut regions. For instance, the trypsin genes *alpha-Try*, *beta-Try*, and *gamma-Try* are highly expressed in the anterior midgut, *theta-Try* is highly expressed in the middle midgut, and *kappa-Try*, *lambda-Try*, and *iota-Try* are enriched in the posterior midgut. ISCs also differ between subregions in their division rate, their cell shape, their numbers relative to the number of EEs and ECs, and their abilities to produce tumors [10]. Interestingly, clonal analyses revealed that the subregions of the midguts resemble cellular compartments [10], reminiscent to compartmentalization observed in imaginal tissues [12]. Most ISCs generate progeny that do not cross the subregion borders, suggesting that differences in ISCs may contribute to midgut regionalization.

Figure 1



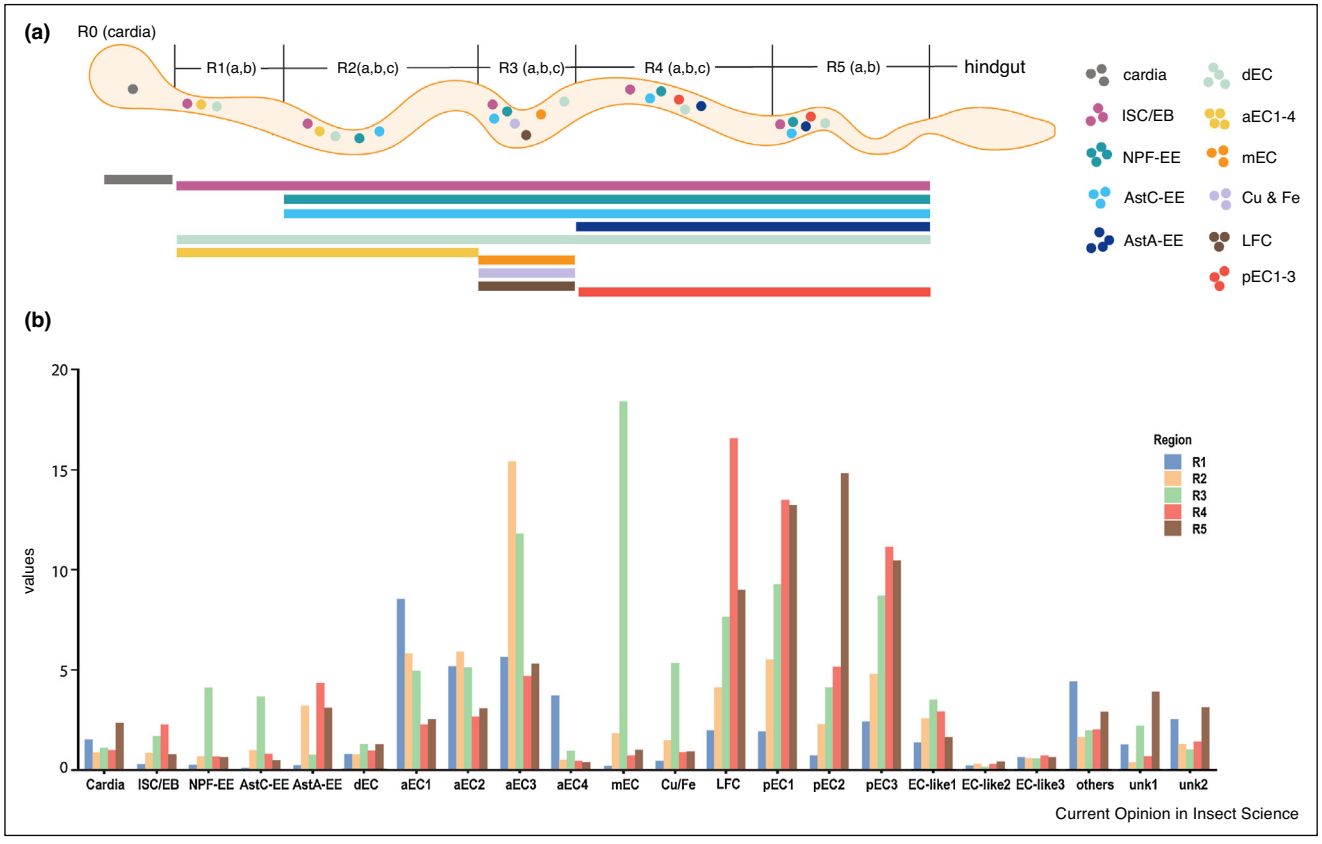
The *Drosophila* adult midgut and its cell types. Schematic of the adult midgut compartments, as defined by Buchon *et al.* (2013) (above, black) and by Marianes & Spradling (2013) (below, blue). The general cellular composition of the digestive tract consists of a monolayer of epithelial cells in close contact to the visceral muscle (VM), neurons and trachea. ISCs proliferate and differentiate into enteroblasts (EBs) or enteroendocrine progenitors (EEPs) that subsequently terminally differentiate into enterocytes (ECs) or enteroendocrine cells (EEs), respectively. EEs develop into specific subtypes depending on the combinatorial expression of transcription factors (TFs) (see text).

To further understand the contribution of specific cell types to gut function and regional differences, Dutta *et al.* generated cell type specific transcriptomes using fluorescence-activated cell sorting (FACS) of each cell type within five gut regions (R1–R5) (Figure 1). Specific cell types were isolated using cell-type specific Gal4-based GFP expression: ISCs (*Dl-gal4*), EBs (*Su(H)-GBE-gal4*), ECs (*Myo1A-gal4*), EEs (*pros-gal4*), and visceral muscle (VM) (*How-gal4*). Interestingly, cell types displayed transcriptional variation along the length of the gut. In particular, R1 and R3 ISCs are quite distinct from R2, R4, and R5 ISCs. For example, ISCs in the acidic R3 region express vacuolar H+ ATPases, indicating an adaptation of ISCs to their local environment. EBs and ECs also showed clear regional specificities, including altered expression of genes involved in metabolism and digestion along the gut. Among these cell types, EEs showed the most differences between regions due to the diversity of the gut hormones they produce. Interestingly, several morphogens (*Wnt1*, *Wnt4*, *Wnt6* and *WntD*) are expressed in VM cells in a gradient dependent manner, which might provide regional specific niches for ISCs. In addition, an unexpected role of EEs and ECs in producing different combination of AMP secretion upon bacterial infection was observed. Finally, genes that were preferentially expressed in a cell type specific manner were identified and their functions were validated by RNAi. For example, *sna* was shown to regulate stem cell differentiation in the R4–R5 region.

Altogether, the transcriptomes captured by these previous studies have uncovered fundamental differences between the cell-types within and between regions of the fly midgut. However, both RNAseq of specific regions [9,10] and RNAseq of FACS cell types have limitations. RNAseq of specific regions requires visual discrimination of regional boundaries within the midgut, which can be prone to inconsistencies between individual guts. Additionally, bulk sequencing does not distinguish between cell types within a dissected region, and differences between quiescent and transient states of ISCs undergoing differentiation are likely missed. With regards to RNAseq of FACS cell types, different expression in GFP levels associated with Gal4-based GFP expression can affect sorting and cell recovery. Further, with FACS, the highly heterogeneous EE subtypes are reduced dimensionally into a single population.

To address some of the limitations of the previous studies, single-cell RNA sequencing (scRNA-seq) was applied on whole guts, as it provides an unbiased approach to survey cell type diversity and function, and to define relationships among cell types [13••]. A total of 22 clusters were identified, including one cluster of ISC/EBs, three clusters of EEs, 14 clusters of ECs, one cluster of cardia, and three unknown clusters. Among the 14 ECs clusters, four clusters (aEC1-4) map to the anterior midgut, three clusters (pEC1-3) to the posterior midgut, one cluster of mEC to the middle midgut, one cluster to copper and iron cells, one cluster to large flat cells (LFC), one cluster to differentiating EC (dEC) and three EC-like clusters (Figure 2). Genes categorized as encoding members of signaling pathways, transcription factors, cytoskeletal proteins, and RNA-binding proteins

Figure 2

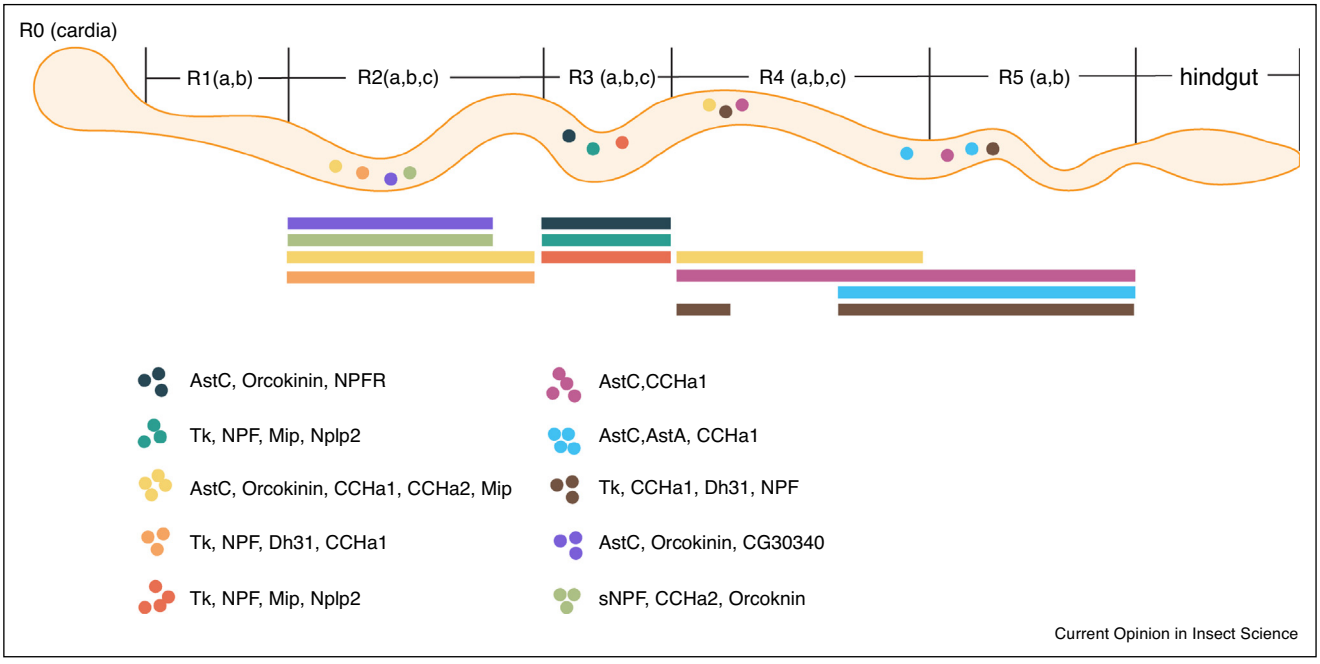


Summary of the spatial distribution of intestinal cell clusters defined by Hung *et al.* (2020). (a) The regional distribution of 22 cell clusters generated by Hung *et al.* (2020) was mapped using the region-specific gene enrichment (RSGE) algorithm developed by Guo *et al.* (2019). (b) The bar graph depicts the regional enrichment score for the top 100 region-specific genes in R1 to R5 (derived from Dutta *et al.* (2015)).

are enriched in ISC/EB progenitor cells. Genes encoding proteins involved in vesicle docking, fusion and secretion are enriched in EEs, reflecting the molecular machinery for hormone secretion. Genes that encode proteins involved in metabolic processes, serine proteases, and transporters are enriched in ECs. Different EC clusters are enriched for different types of metabolic processes. For example, aECs and cardia are enriched in genes involved in carbohydrate metabolism, galactose metabolism, and binding to chitin, a component of the peritrophic matrix secreted by cardia cells. Genes involved in mannosidase activity, galactose metabolism and lipid metabolism are enriched in pECs. Interestingly, ribosomal proteins (RpL and RpS) are enriched in the ISC/EB and dEC clusters, and transmission electron micrographs of the midgut confirmed that ISCs exhibit more ribosomes than mature ECs. These observations are consistent with several studies indicating that embryonic stem cells maintain a high level of free inactive ribosomes, which might be essential for rapidly rewiring and priming gene expression toward differentiation [14]. In addition, a study of *Drosophila* germline stem cells demonstrated that the transition from self-renewal to differentiation relies on increased ribosome biogenesis and protein synthesis [15], suggesting that high levels of ribosomes is one characteristic of stem cells.

Analysis of scRNA-seq data also identified previously known and new markers, such as *klu*, *lola* and *Eip75B* [13••]. In particular, *klu is* expressed in cells that are positive for *Su(H)GBE-LacZ*, confirming that *klu* is a novel EB marker [13••]. Knocking down *klu* in progenitors resulted in an increase in EEs, which is consistent with an independent study showing that *klu* maintains the EB commitment to EC lineage and suppresses EE fate via indirectly repressing the levels of *scute* (*sc*) [13••,16•]. Further, *lola* expression overlaps with *esg*, a ISC/EBs marker, and was later found to function in ISC/EBs to suppress ISC proliferation [17]. Finally, *Eip75B* is expressed preferentially in ISC/EBs. Two subsequently studies have shown that Eip75B acts downstream of the ecdysone receptor (EcR) to drive ISC daughter cells toward absorptive EC lineage to ensure epithelial growth induced by the steroid hormone ecdysone [18•,19•].

Figure 3



Summary of the signature peptide hormones and spatial distribution of EE subtypes identified by Guo *et al.* (2019).

In addition to scRNA-seq of the whole midgut, Guo *et al.* (2019) conducted scRNA-seq of EEs by isolating EEs via FACS [20••]. They identified 10 major EE subtypes in the midgut and comprehensively characterized the repertoires in these subtypes of the gut hormones (Tk, Dh31, NPF, Nplp2, Gpb5, CCAP, ITP, sNPF, CCHa2, CCHa1, AstA, Orcokinin and AstC). Single-cell analysis revealed that most individual EEs produce approximately 2–5 gut hormones ([20••], consistent with [13••]). To retrieve the spatial distribution of all EE subtypes, which is lost during tissue dissociation, Guo generated a region-specific gene enrichment (RSGE) algorithm based on regional bulk EE RNA-seq from previously published data [21] (Figure 3). Interestingly, the expression of gut hormones and their corresponding receptors in EE subtypes revealed both local (paracrine) and long-range communication between gut regions [13••,20••]. For example, AstC-EEs also express NPF receptors, suggesting that NPF from neighboring cells may regulate AstC-EE in a paracrine manner. In addition, the gut hormone CCHa1 is expressed in the posterior midgut while its receptor CCHa1-R is expressed in the anterior and middle midgut, suggesting potential long-range communication. Alternatively and noteworthy, as the gut folds into a stereotypical 3D structure inside the abdomen, gut hormone producing EEs and gut cells expressing their receptors may in some cases be very close to each other.

Transient expression of *sc* in ISCs specifies *pros*-expressing EEPs, which divide one more time before terminal differentiation to generate two EEs (pros+) that express AstC (class I EE) or Tk (class II EE) [22,23]. This binary cell fate is specified by Notch activity [22,24,25]. To identify the transcription factors (TFs) that govern EE subtype specification, Guo identified a set of 14 TFs whose binary expression states (ON or OFF) alone are sufficient to classify the subtypes for all EE clusters (with the exception of 2 clusters that share very similar expression signatures). Depleting individual TFs resulted in decreased expression of at least one or more gut hormones. For example, knocking down *Ptx1* in EEs decreased AstC expression and loss of AstC expressing EEs. This approach identified two TFs, Ptx1 and Mirror (Mirr), involved in the specification of class I and class II EEs, respectively. In addition, this study identified several regionally expressed TFs that define spatial identity of EEs (e.g. *esg* in the middle midgut).

Although these findings suggest that EE subtypes can be defined by a combinatorial expression of TFs, expression of some gut hormones did not strictly correlate with EE subtype clusters. One explanation is that extrinsic factors, such as infection or mechanical stress, could influence the expression of gut hormones. Alternatively, EEs could be plastic, as shown for mammalian EEs, which switch their hormone expression as they differentiate [26]. Importantly, the directed approach of isolating cells of interest for scRNA-seq led to a better resolution of EE subtypes than whole gut scRNA-seq had provided. Thus, a more directed approach of isolating cells for scRNA-seq by

using FACS (either based on reporter expression or *in situ* probes [27]), is likely to provide further characterization of cell types.

An important feature of scRNA-seq is to provide information on lineage relationships between cell types, which have classically relied on genetic recombination and fluorescent reporters to label the starting cells and their progeny. Computational methods, such as Monocle [28], slingshot [29], and STREAM [30], can be applied to scRNA-seq data to construct developmental trajectories and extract transient states from pseudo-temporal predictions. For example, Hung *et al.* (2020) used slingshot to map cell differentiation trajectories of early states (ISCs) to terminal states (EEs and ECs). Three trajectories of ISCs were identified: Firstly, ISC/EBs → mECs → dECs (differentiating ECs) → aECs; secondly, ISC/EBs → mECs → dECs → pECs; and finally, ISC/EBs → EE. Slingshot also helped to classify dEC as an intermediate state, as dECs did not resemble either ISCs or mature ECs. In the future, the use of DNA sequence barcodes (instead of fluorescent reporters) to encode clonal information should help capture temporal information in the context of lineage tracing. In particular, CRISPR-Cas9 has been used to develop several lineage tracing strategies, including GESTALT [31], LINNAEUS [32], MEMOIR [33], and ScarTrace [34]. In general, these methods use Cas9-induced stochastic mutations to create insertions or deletions (during DNA repair following double-strand breaks) at predefined target sites. Since the alterations occur in the DNA, these methods rely on inferring cell lineage from changes observed over multiple cell divisions. While this works well for constructing trajectories for cell-types that arise from distant progenitors, it falls short in capturing the transient-states over the lifetime of a cell. One possible solution is to use Cas-related technologies to create alterations in predefined RNA reporters. These edited sequences could then be used as barcodes for determining the transient state of single cells and be readily combined with other sequencing-based omics measurements to construct the trajectory of stem cell differentiation.

Surprisingly, while scRNA-seq was able to characterize extensive regional differences in ECs and EEs, it was not able to distinguish regional differences among ISC/EBs clusters. Possibly, the preferential expression of cell type specific genes in ISCs/EBs is more profound than that of regional specific genes. Another possibility is that regional difference genes in ISCs are expressed at low levels, which are not easily detected by scRNA-seq using the commonly used 10× Genomics platform. In this case, use of a more in-depth sequencing approach such as smart-seq [35] to sequence the full-length transcripts rather than only 3′ ends (10× Genomics), and a more targeted approached, such as use of GFP-sorted cells or Probe-seq to enrich for a specific cell type, could be used in future studies.

Future scRNA-seq studies should enable the discovery of the gene regulatory networks that underlie ISC proliferation and differentiation. This would involve mapping TFs that regulate each other and their downstream targets. A few studies have attempted to address this by inferring co-expression networks from scRNA-seq data [36,37]. More recently, SCENIC (Single-Cell rEgulatory Network Inference and Clustering) integrated co-expression analyses with *cis*-regulatory motif analyses to identify TF-to-target modules (regulons) enriched in single cell clusters [38]. What remains unavailable are computational methods that can map regulon expression onto lineage trajectory predictions. Putative TF-cascade predictions derived from such approaches will be crucial to our understanding of gene networks that underly ISC differentiation.

Finally, although scRNA-seq has the advantage of providing substantial transcriptomics information of individual cells, the spatial information of individual cells in the tissue is missing. Thus, region-specific markers identified by scRNA-seq need to be validated to retrieve spatial information. The wealth of existing data from large-scale *in situ* RNA hybridization studies and gene expression reporters available for *Drosophila* will considerably facilitate such studies.

Conflict of interest statement

Nothing declared.

Acknowledgments

We thank Stephanie Mohr for helpful comments on the manuscript. JSSL is supported by a Croucher fellowship for Postdoctoral Research from the Croucher Foundation and NP is an investigator of the Howard Hughes Medical Institute.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Micchelli CA, Perrimon N: **Evidence that stem cells reside in the adult Drosophila midgut epithelium**. *Nature* 2006, **439**:475-479.
2. Ohlstein B, Spradling A: **The adult Drosophila posterior midgut is maintained by pluripotent stem cells**. *Nature* 2006, **439**:470-474.
3. Ohlstein B, Spradling A: **Multipotent Drosophila intestinal stem cells specify daughter cell fates by differential notch signaling**. *Science* 2007, **315**:988-992.
4. Veenstra JA: **Peptidergic paracrine and endocrine cells in the midgut of the fruit fly maggot**. *Cell Tissue Res* 2009, **336**:309-323.
5. Veenstra JA, Ida T: **More Drosophila enteroendocrine peptides: Orcokinin B and the CCHamides 1 and 2**. *Cell Tissue Res* 2014, **357**:607-621.
6. Veenstra JA, Agricola HJ, Sellami A: **Regulatory peptides in fruit fly midgut**. *Cell Tissue Res* 2008, **334**:499-516.
7. Chen J, Kim SM, Kwon JY: **A systematic analysis of Drosophila regulatory peptide expression in enteroendocrine cells**. *Mol Cells* 2016, **39**:358-366.

8. Poulson DF, Bowen VT, Hilse RM, Rubinson AC: **The copper metabolism of Drosophila**. *Proc Natl Acad Sci U S A* 1952, **38**:912-921.

9. Buchon N, Osman D, David FP, Fang HY, Boquete JP, Deplancke B, Lemaitre B: **Morphological and molecular characterization of adult midgut compartmentalization in Drosophila**. *Cell Rep* 2013, **3**:1725-1738.

10. Marianes A, Spradling AC: **Physiological and stem cell compartmentalization within the Drosophila midgut**. *Elife* 2013, **2**:e00886.

11. Strand M, Micchelli CA: **Quiescent gastric stem cells maintain the adult Drosophila stomach**. *Proc Natl Acad Sci U S A* 2011, **108**:17696-17701.

12. Dahmann C, Basler K: **Opposing transcriptional outputs of Hedgehog signaling and engrailed control compartmental cell sorting at the Drosophila A/P boundary**. *Cell* 2000, **100**:411-422.

13. •• Hung RJ, Hu Y, Kirchner R, Liu Y, Xu C, Comjean A, Tattikota SG, Li F, Song W, Ho Sui S *et al.*: **A cell atlas of the adult Drosophila midgut**. *Proc Natl Acad Sci U S A* 2020, **117**:1514-1523
scRNA-seq was used on whole adult gut to characterize cell types and states and identified 22 distinct clusters representing ISCs, EBs, EEs and ECs. This study identified molecular signatures of each cluster, new markers for EBs, new gut hormones for EEs, their combinational expression diversity, and paracine function of EEs. This is the first study of single cell transcriptomics on whole gut.

14. You KT, Park J, Kim VN: **Role of the small subunit processome in the maintenance of pluripotent stem cells**. *Genes Dev* 2015, **29**:2004-2009.

15. Sanchez CG, Teixeira FK, Czech B, Preall JB, Zamparini AL, Seifert JR, Malone CD, Hannon GJ, Lehmann R: **Regulation of ribosome biogenesis and protein synthesis controls germline stem cell differentiation**. *Cell Stem Cell* 2016, **18**:276-290.

16. • Korzelius J, Azami S, Ronnen-Oron T, Koch P, Baldauf M, Meier E, Rodriguez-Fernandez IA, Groth M, Sousa-Victor P, Jasper H: **The WT1-like transcription factor Klumpfuss maintains lineage commitment of enterocyte progenitors in the Drosophila intestine**. *Nat Commun* 2019, **10**:4123
The authors show that klu maintains EC lineage commitment in EBs by suppressing the EE lineage. This study together with Hung *et al.* (2020) are the first to show klu as a marker for EBs.

17. Hao X, Wang S, Lu Y, Yu W, Li P, Jiang D, Guo T, Li M, Li J, Xu J *et al.*: **Lola regulates Drosophila adult midgut homeostasis via non-canonical hippo signaling**. *Elife* 2020, **9**.

18. • Zipper L, Jassmann D, Burgmer S, Gorlich B, Reiff T: **Ecdysone steroid hormone remote controls intestinal stem cell fate decisions via the PPARgamma-homolog Eip75B in Drosophila**. *Elife* 2020, **9**
The authors found that ecdysone acts on the midgut to increase ISC proliferation after mating. Together with Ahmed *et al.* (2020), these studies are the first to show communication between ovaries and the midgut.

19. • Ahmed SMH, Maldera JA, Krunic D, Paiva-Silva GO, Penalva C, Teleman AA, Edgar BA: **Fitness trade-offs incurred by ovary-to-gut steroid signalling in Drosophila**. *Nature* 2020, **584**:415-419
The authors found that the increase in ecdysone production in ovaries, after mating, acts on ISCs to promote proliferation. This paper together with Zipper *et al.* (2020) are two pioneering studies that show communication between ovaries and the midgut.

20. •• Guo X, Yin C, Yang F, Zhang Y, Huang H, Wang J, Deng B, Cai T, Rao Y, Xi R: **The cellular diversity and transcription factor code of Drosophila enteroendocrine cells**. *Cell Rep* 2019, **29**:4172-4185 e4175
Using FACS to isolate EE cells for scRNA-seq, the authors identify 10 major EE subtypes and computationally map them to specific regions in the adult Drosophila midgut. This study is the first to show a combinatorial TF code that controls EE cell diversity.

21. Dutta D, Buchon N, Xiang J, Edgar BA: **Regional cell specific RNA expression profiling of FACS isolated drosophila intestinal cell populations**. *Curr Protoc Stem Cell Biol* 2015, **34** 2F 2 1–2F 14.

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

23. Zeng X, Hou SX: **Enteroendocrine cells are generated from stem cells through a distinct progenitor in the adult Drosophila posterior midgut**. *Development* 2015, **142**:644-653.

24. Beehler-Evans R, Micchelli CA: **Generation of enteroendocrine cell diversity in midgut stem cell lineages**. *Development* 2015, **142**:654-664.

25. Guo Z, Ohlstein B: **Stem cell regulation. Bidirectional Notch signaling regulates Drosophila intestinal stem cell multipotency**. *Science* 2015, **350**.

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

27. Amamoto R, Garcia MD, West ER, Choi J, Lapan SW, Lane EA, Perrimon N, Cepko CL: **Probe-Seq enables transcriptional profiling of specific cell types from heterogeneous tissue by RNA-based isolation**. *Elife* 2019, **8**.

28. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL: **The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells**. *Nat Biotechnol* 2014, **32**:381-386.

29. Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, Purdom E, Dudoit S: **Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics**. *BMC Genomics* 2018, **19**:477.

30. Chen H, Albergante L, Hsu JY, Lareau CA, Lo Bosco G, Guan J, Zhou S, Gorban AN, Bauer DE, Aryee MJ *et al.*: **Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM**. *Nat Commun* 2019, **10**:1903.

31. McKenna A, Findlay GM, Gagnon JA, Horwitz MS, Schier AF, Shendure J: **Whole-organism lineage tracing by combinatorial and cumulative genome editing**. *Science* 2016, **353**:aaf7907.

32. Raj B, Gagnon JA, Schier AF: **Large-scale reconstruction of cell lineages using single-cell readout of transcriptomes and CRISPR-Cas9 barcodes by scGESTALT**. *Nat Protoc* 2018, **13**:2685-2713.

33. Frieda KL, Linton JM, Hormoz S, Choi J, Chow KK, Singer ZS, Budde MW, Elowitz MB, Cai L: **Synthetic recording and in situ readout of lineage information in single cells**. *Nature* 2017, **541**:107-111.

34. Alemany A, Florescu M, Baron CS, Peterson-Maduro J, van Oudenaarden A: **Whole-organism clone tracing using single-cell sequencing**. *Nature* 2018, **556**:108-112.

35. Picelli S, Faridani OR, Bjorklund AK, Winberg G, Sagasser S, Sandberg R: **Full-length RNA-seq from single cells using Smart-seq2**. *Nat Protoc* 2014, **9**:171-181.

36. Guo M, Wang H, Potter SS, Whitsett JA, Xu Y: **SINCERA: a pipeline for single-cell RNA-Seq profiling analysis**. *PLoS Comput Biol* 2015, **11**:e1004575.

37. Moignard V, Woodhouse S, Haghverdi L, Lilly AJ, Tanaka Y, Wilkinson AC, Buettner F, Macaulay IC, Jawaid W, Diamanti E *et al.*: **Decoding the regulatory network of early blood development from single-cell gene expression measurements**. *Nat Biotechnol* 2015, **33**:269-276.

38. Aibar S, Gonzalez-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F, Marine JC, Geurts P, Aerts J *et al.*: **SCENIC: single-cell regulatory network inference and clustering**. *Nat Methods* 2017, **14**:1083-1086.