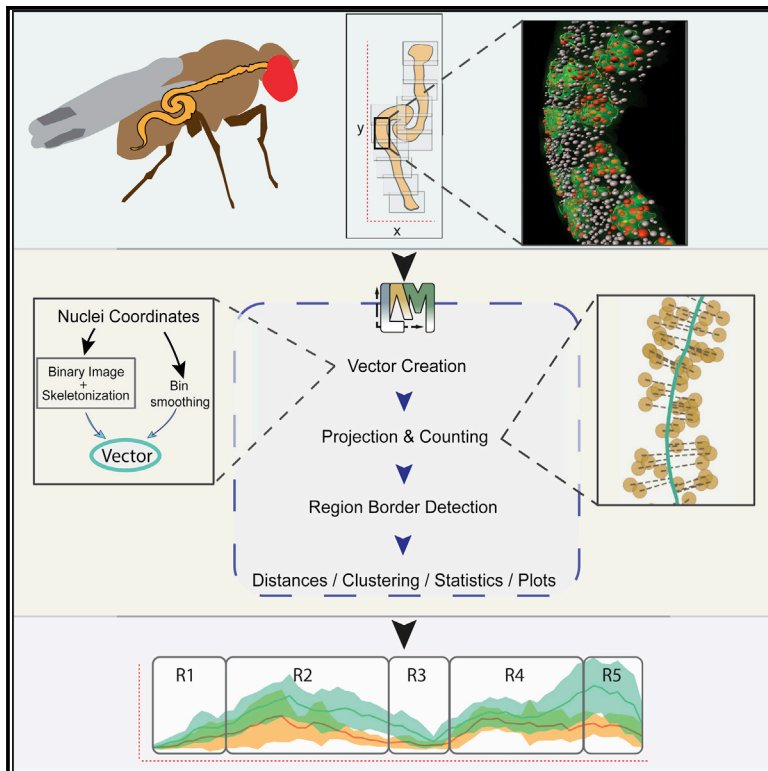


# An image analysis method for regionally defined cellular phenotyping of the *Drosophila* midgut

## Graphical abstract



## Authors

Arto Viitanen, Josef Gullmets, Jack Morikka, Pekka Katajisto, Jaakko Mattila, Ville Hietakangas

## Correspondence

jaakko.i.mattila@helsinki.fi (J.M.), ville.hietakangas@helsinki.fi (V.H.)

## In brief

The intestine is divided into functionally distinct regions along its anteroposterior axis. Here, Viitanen and co-workers develop a quantitative image analysis approach to map and compare cellular phenotypes of *Drosophila* midgut with subregional resolution.

## Highlights

- LAM is an image analysis method for quantitative analysis of the whole midgut
- LAM allows automated identification of the major regional borders
- Midguts have a high density of enteroendocrine cells at major border regions
- Midgut stem cells display regionally heterogeneous injury response



## Article

# An image analysis method for regionally defined cellular phenotyping of the *Drosophila* midgut

Arto Viitanen,<sup>1,2</sup> Josef Gullmets,<sup>1,2</sup> Jack Morikka,<sup>1,2</sup> Pekka Katajisto,<sup>1,2</sup> Jaakko Mattila,<sup>1,\*</sup> and Ville Hietakangas<sup>1,2,3,\*</sup><sup>1</sup>Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki 00790, Finland<sup>2</sup>Institute of Biotechnology, University of Helsinki, Helsinki 00790, Finland<sup>3</sup>Lead contact\*Correspondence: [jaakko.i.mattila@helsinki.fi](mailto:jaakko.i.mattila@helsinki.fi) (J.M.), [ville.hietakangas@helsinki.fi](mailto:ville.hietakangas@helsinki.fi) (V.H.)<https://doi.org/10.1016/j.crmeth.2021.100059>

**MOTIVATION** Because of the combination of small size and regionalized structure, the *Drosophila* midgut is an ideal model for organ-wide analyses of intestinal cells, including stem cells. Although imaging of the whole midgut is feasible by using fast and affordable tile scan imaging, the downstream analysis remains challenging. Because cellular phenotypes are inherently variable, it is necessary to quantitatively analyze several replicate midguts, which requires proper alignment of the spatially resolved cellular data. Each midgut has a unique morphology, which brings about the need for time-consuming and subjective manual work to identify, align, and compare the intestinal regions of replicate samples. Because of these technical challenges, the potential of *Drosophila* midgut for organ-wide analysis has not been taken to full use.

## SUMMARY

The intestine is divided into functionally distinct regions along the anteroposterior (A/P) axis. How the regional identity influences the function of intestinal stem cells (ISCs) and their offspring remain largely unresolved. We introduce an imaging-based method, “Linear Analysis of Midgut” (LAM), which allows quantitative, regionally defined cellular phenotyping of the whole *Drosophila* midgut. LAM transforms image-derived cellular data from three-dimensional midguts into a linearized representation, binning it into segments along the A/P axis. Through automated multivariate determination of regional borders, LAM allows mapping and comparison of cellular features and frequencies with subregional resolution. Through the use of LAM, we quantify the distributions of ISCs, enteroblasts, and enteroendocrine cells in a steady-state midgut, and reveal unprecedented regional heterogeneity in the ISC response to a *Drosophila* model of colitis. Altogether, LAM is a powerful tool for organ-wide quantitative analysis of the regional heterogeneity of midgut cells.

## INTRODUCTION

The intestine has a critical role in regulating organismal metabolism and immunity (Miguel-Aliaga et al., 2018). These functions are dynamically modulated by environmental factors, such as nutrition and microbes. Uncovering the mechanistic basis of the underlying regulation requires tractable *in vivo* model systems. The *Drosophila* midgut, analogous to the mammalian small intestine, has proved to be a powerful model for understanding intestinal physiology (Miguel-Aliaga et al., 2018). The midgut is composed of four cell types: the absorptive enterocytes (ECs), their differentiating progenitor cells, called enteroblasts (EBs), the hormone-secreting enteroendocrine (EE) cells, and the mitotic intestinal stem cells (ISCs) (Miguel-Aliaga et al., 2018). The midgut is an adaptive regenerative organ whose cellular turnover and composition is affected by diet, sex, inflammation, age, and reproductive status (Biteau et al., 2008; Buchon et al., 2009, 2013; Hudry et al., 2016; Reiff et al., 2015). Previous

studies have uncovered regulatory pathways involved in the control of intestinal homeostasis through inter- and intracellular signaling (Gervais and Bardin, 2017; Guo et al., 2016).

To perform the functions of digestion, absorption, metabolism, nutrient sensing, and signaling in a sequentially coordinated manner, the animal intestine is compartmentalized into regions along its anteroposterior (A/P) axis (Miguel-Aliaga et al., 2018; O’Brien, 2013). Moreover, human intestinal pathophysiologicals, such as cancer or inflammatory disorders, often manifest in a region-specific manner (Missiaglia et al., 2014; Mowat and Agace, 2014). Therefore, the mechanisms that establish, maintain, and modulate the regionalized functions of the intestine are of high biological and medical relevance. The *Drosophila* midgut regions have been distinguished on the basis of anatomical characteristics, differential staining with histological dyes, and region-specific gene expression patterns (Buchon et al., 2013; Dimitriadis, 1991; Marianes and Spradling, 2013). Buchon et al. (2013) divided the midgut into six major regions (R0 to R5), which can



be distinguished on the basis of cross-intestinal anatomy. R1–R5 were further divided into 14 subregions on the basis of morphological, histological, and gene expression differences. In a parallel study, [Marianes and Spradling \(2013\)](#) divided the midgut into ten zones, with significant overlap to the 14 subregions defined by [Buchon et al. \(2013\)](#).

Molecular analyses of the intestinal cell types have given more detailed insight into midgut regionalization. Consistent with sequentially coordinated digestion and absorption, the digestive enzyme and nutrient transporter genes display strictly region-specific expression patterns in the ECs ([Dutta et al., 2015](#)). The EE cells, mediating the signaling function of the intestine, can be divided into ten subtypes displaying region-specific distribution ([Guo et al., 2019](#)). In addition to the differentiated cell types, it has also been proposed that the function of undifferentiated ISCs depends on regional identity. The ISCs display regional autonomy, i.e., their differentiated daughter cells do not cross most region boundaries ([Marianes and Spradling, 2013](#)). The ISCs in different midgut regions also display distinct morphological features as well as differential gene expression, exemplified by the finding that more than 900 genes show regional expression variation in the ISCs ([Dutta et al., 2015](#); [Marianes and Spradling, 2013](#)). The acidic R3 region, often termed the stomach of *Drosophila*, contains stem cells that have been deemed quiescent in unchallenged conditions but activated in response to stressful stimuli, such as heat shock or pathogen ingestion ([Strand and Micchelli, 2011](#)). Despite the evidence strongly implying regional ISC heterogeneity, most studies on ISCs focus on one specific region (mostly R4), and the possible impact of regional identity and tissue environment on ISC regulation is largely overlooked.

Achieving representative data of the midgut requires unbiased quantitative analysis of all midgut regions. Rapid development of affordable and fast tile scan imaging has made it feasible to collect high-resolution imaging data from the whole midgut. High phenotypic variation between midguts limits the reproducibility of qualitative analysis and sets the requirement for robust quantitative analysis of replicate samples. However, achieving quantitative and regionally defined data from midgut cells has remained a major bottleneck, hampering the use of organ-wide analysis. Here we describe a widely applicable phenotyping method called LAM (Linear Analysis of Midgut) to achieve spatially defined quantitative data on midgut cells. LAM transforms data from three-dimensional (3D) midgut images into one dimension by an algorithm design that couples cellular identities into a specific position at a linear representation of the midgut. This enables binning of cell-specific data along the A/P axis and joining of replicate samples into spatially resolved data matrices. The use of one-dimensional (1D) data enables automatization of the regional boundary identification, allowing accurate alignment of corresponding regions. These features enable LAM to achieve robust quantitative phenotyping of midguts with subregional resolution. To facilitate the downstream data analysis, LAM includes various options for visualization, statistical analysis, and data subsetting. A graphical user interface, user manual, and tutorial videos make LAM accessible to all researchers. As a proof of concept, we use LAM to quantitatively analyze regional distributions of ISCs, EBs, and EE cells.

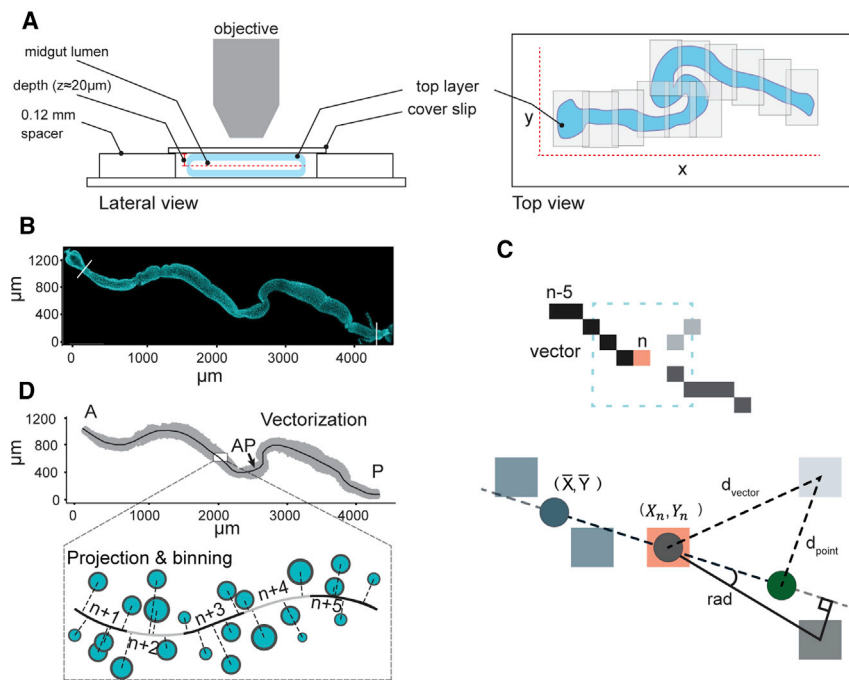
We also demonstrate the regional heterogeneity of the injury response to a well-established colitis model, dextran sulfate sodium (DSS) treatment. The organ-wide analysis by using LAM revealed several features of DSS-induced response, including a failure of regenerative stem cell activation in R3, a regionally discordant pattern of stem cell division and differentiation in R4 versus R5, and an increase in EE cell numbers in the posterior R4/R5 region. By making unbiased, quantitative, organ-wide analysis highly feasible, LAM is expected to open new avenues for the analysis of regional heterogeneity of midgut cells.

## RESULTS

### An approach for spatially defined quantitative phenotyping of the *Drosophila* midgut

To analyze the spatial heterogeneity of intestinal cell responses in an unbiased and reproducible manner, we developed an intestinal phenotyping approach that is automated, quantitative, and regionally defined. For imaging the nuclei of pseudostratified midgut epithelium, fixed 4',6-diamidino-2-phenylindole (DAPI)-stained tissues were mounted in between a coverslip and a microscope slide with 0.12-mm spacers. Flattening the intestinal tube into two epithelial layers while still separated by its lumen allowed z-stack acquisition of one layer, saving time and reducing file size ([Figure 1A](#)). As an initial step, we sought a means to reduce the tile scan stacks of nonlinear midguts into a linearized representation ([Figure 1B](#)). To this end, an algorithm that approximates the midlines of partially uncoiled midguts along their A/P axis was used. The algorithm first transformed coordinates of nuclei belonging to the midgut into a binary image onto which pixel erosion was applied to produce a pixel-wide skeleton. The pixels of the skeleton were then iteratively scored to produce a linear representation of the midgut ([Figure 1C](#)), which we colloquially call vectors, as they reduce the data into 1D arrays. Subsequently, any object in the 3D space, such as nuclei, and any associated characteristics could be projected and have their x:y:z coordinates reduced to a linear reference. Thereby, the projection point's normalized distance along the midline vector directly corresponds to the location along the A/P axis of the linearized midgut ([Figure 1D](#)). The vector and data were then divided into bins, the number of which can be adjusted to a desired spatial resolution. Because of the linear referencing and binning of the measured cellular features, data collected from different intestines could be joined as biological replicates in a spatially relevant data matrix, where each row corresponds to the same biological location. As a result, the binning and joining of data allowed spatially defined quantitative representation and statistical analysis between sample groups.

Next, we wanted to address whether the quantitative information obtained by using our algorithm allowed automatic determination of the borders of midgut regions. Midgut regions are characterized by differences in enterocyte ploidy and density ([Figure 2A](#)) ([Marianes and Spradling, 2013](#)) and are separated by constrictions of the midgut radius ([Buchon et al., 2013](#)). We first separated the polyploid enteroblast and enterocyte population from diploid cells, based on nuclear area ([Figure 2B](#)). The filtered cell population was then projected onto the A/P vector along with associated data on polyploid nuclear area and



**Figure 1. A pipeline for regionally defined quantification of *Drosophila* midguts**

(A) Schematic presentation of the whole midgut imaging.

(B) A representative tile scan image of DAPI (cyan)-stained midgut. After imaging and stitching of the tiles, the image is processed to exclude any features lying outside the area of interest. Subsequently, the image is analyzed for DAPI spots by, for example, the spot-detection algorithm of the Imaris software, or by StarDist segmentation.

(C) Pixel selection in skeleton vector creation. The vector is a piecewise line starting from leftmost pixels of the binary image skeleton. The vector is extended with pixel coordinates based on a scoring system that gives penalties depending on the pixel's directional change and distance. With  $n$  as the last pixel of the vector, a direction-giving line is formed based on coordinates of  $n$  and the average coordinate of  $n-1$  and  $n-2$ . On this line, a projection point (green circle) is created equidistant from  $n$  as the average coordinate. For each candidate pixel, distances to  $n$  ( $d_{\text{vector}}$ ) and the projection point ( $d_{\text{point}}$ ) are determined, both contributing equally to the penalty. Additionally, the absolute radian changes of each pixel in relation to  $n$  and the direction line is multiplied by 10 and added to the distance scores to give the full penalty. The pixel with the smallest penalty is added to the vector, and subsequently the

algorithm would follow the pixel in darker gray.

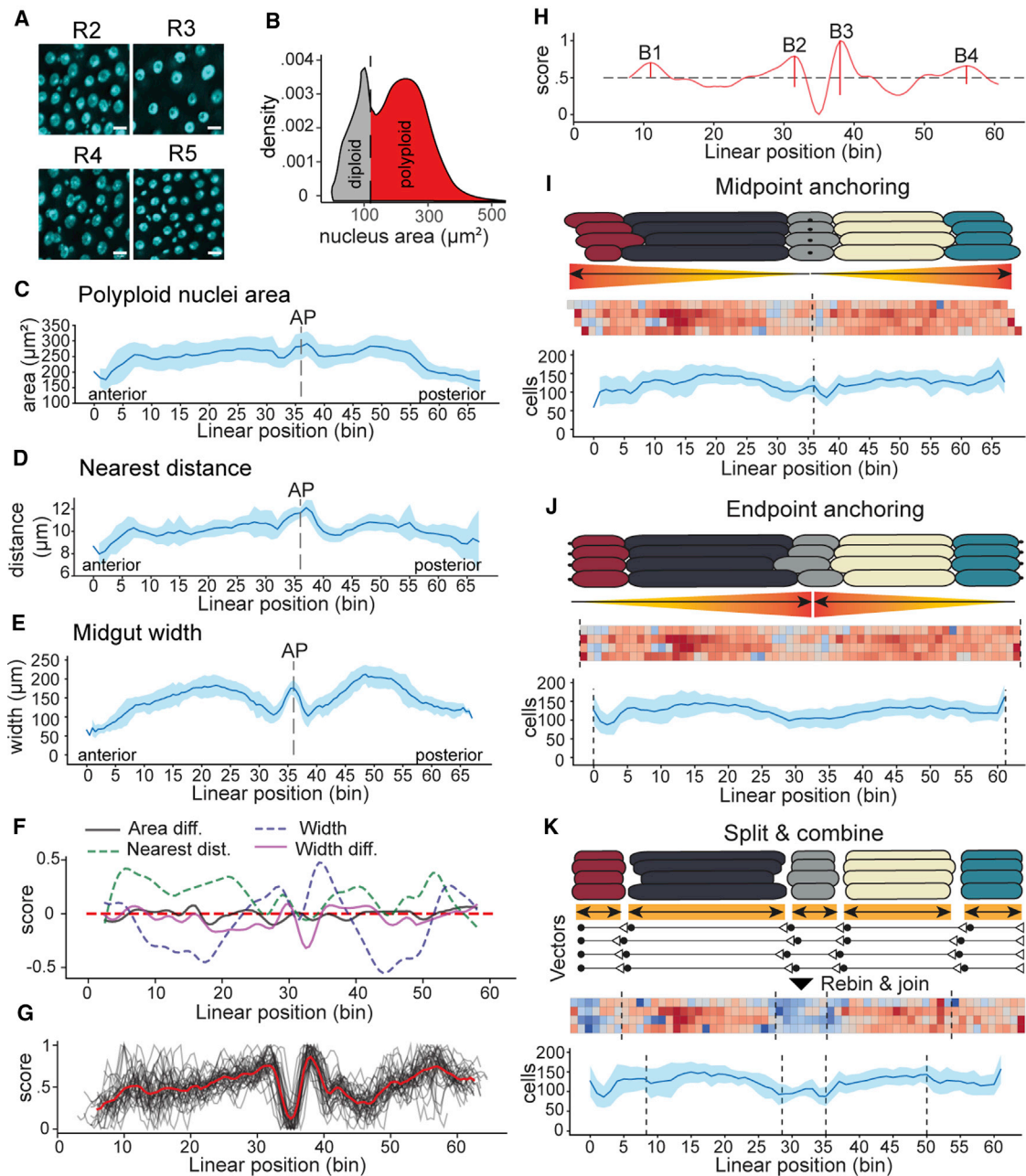
(D) Projection and linearization. The spots, and any accompanying data, are projected onto the vector. The vector is then binned, where the number of bins is chosen on the basis of the desired resolution. An “anchoring point” (AP) is introduced into a morphologically distinct place, such as the border between the copper cell region (CCR) and the large flat cells (LFC) region of the middle midgut (arrow).

nucleus-to-nucleus nearest distances. Together with midgut width computed from the projection distances of all nuclei along the vector, these data enabled multivariate mapping for the detection of borders. Because of high variability of morphology, it was not possible to reliably detect the borders from individual midguts (data not shown). This led us to explore border detection from combined measurements of several replicate samples. As even minor variation in region proportions could produce a compounding error resulting in misalignment of border signals between samples, we sought to create more accurate bin-to-bin correspondence. Consequently, we introduced an anchoring point (AP) in the middle of the midgut, located at the border of the copper cell region (CCR) and large flat cell (LFC) region, which can be easily identified on the basis of the difference in nuclear distance (Figure 1D). Alignment of the projected polypliod areas, nearest distances, and midgut widths from several replicate midguts by using the APs revealed characteristic midgut profiles, as described by Buchon et al. (2013) (Figures 2C–2E). Although the borders of all regions are not uniform, they are characterized by sudden, localized changes in values. Therefore, we fitted a Chebyshev polynomial to the normalized data to simulate background context and subtracted it from the values as an adjustment (Figure 2F). After scoring each replicate by summing the values of its weighted variables, distinct patterns could be detected in the joined scores (Figure 2G). Smoothing and peak detection with average values of each group allowed for robust identification of four peaks corresponding to region borders B1–B4 (Figure 2H).

Midgut total length, as well as the length of the individual regions, is variable (Buchon et al., 2013). This poses a challenge for aligning corresponding regions of replicate samples. Accordingly, the utilization of a single alignment point in the middle midgut, i.e., the point where the vectors of different samples are anchored together, can lead to an imprecise alignment of the regions toward the anterior and posterior ends (Figure 2I). On the other hand, anchoring the samples from the ends will reduce the accuracy of the alignment in the middle regions (Figure 2J). To minimize noise introduced by the variable length, we utilized region border analysis to apply several independent alignment points, resulting in a more optimal comparison of midgut regions. In this “split and combine” approach the vectors and projected data were cut on the basis of region border detection, aligned separately, and rejoined back together (Figure 2K). Although this pipeline could lead to slight discrepancy in bin lengths between different regions, it improved accuracy in regional comparisons between the midguts.

We have implemented the analysis tools described above into a Python package, called “Linear Analysis of Midgut” or LAM (<https://github.com/hietakangas-laboratory/LAM>). LAM provides various options for analyzing midgut image-derived feature data, such as object coordinates for measuring cell-to-cell distances and cell clustering (Figures 3A and 3B), object size, and object intensities in a regional manner. It also provides various options for plotting and statistical analysis between sample groups (Figure 3C). We also provide a separate tool for stitching tile images for large-scale datasets

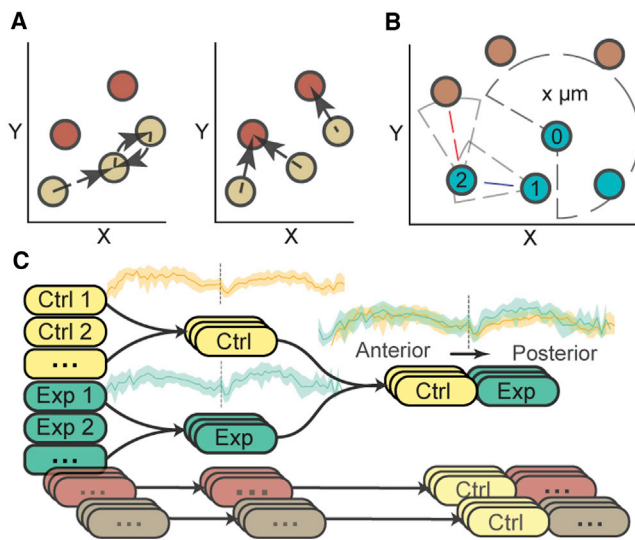




**Figure 2. Border detection and alignment of midguts for pairwise comparison by LAM**

(A) Midgut regions have distinct enterocyte size and density. Representative images of DAPI (cyan)-stained nuclei from the midgut R2–R5 regions. Scale bar, 10  $\mu\text{m}$ .  
 (B) Nuclei area profile from whole midguts determined by identification of DAPI spot areas from the Imaris spot-detection algorithm. For the subsequent analysis of midgut region borders, the diploid cells were filtered out from the dataset.  
 (C) Polyploid nuclei area profile along the A/P axis of midgut. AP, anchoring point.  $n = 32$  midguts. Light-blue shading is the standard deviation.  
 (D) Nuclei nearest distance profile along the A/P axis of midgut. The distance between nuclei is a proxy for cell density. AP, anchoring point.  $n = 32$  midguts. Light-blue shading is the standard deviation.  
 (E) Width profile along the A/P axis of midgut. Midgut width is approximated by following the vector bin-by-bin and summing the average projection distances of the most distant decile of cells on both sides of the vector. AP, anchoring point.  $n = 32$  midguts. Light-blue shading is the standard deviation.  
 (F–H) Border-detection algorithm performs a multivariate border region detection for each sample and outputs average border locations for each sample group. (F) Smoothed scores of default border-detection variables along A/P axis of a sample. The variables are scored on the basis of weighted divergence from expected values, i.e., from a fitted fifth degree Chebyshev polynomial. The variable scores are summed to provide a total score for each location of a sample, which are then rescaled to interval [0, 1] in order to give comparable peak locations despite phenotypic differences. (G) Total scores of samples belonging to one sample

(legend continued on next page)



**Figure 3. Functionalities of LAM**

(A and B) Feature-to-feature distances and clustering. Both functionalities are based on calculating distances to neighbors of each feature. (A) Feature-to-feature distance calculations determine the nearest neighbors of a channel's features either on one channel's dataset (left) or compared with a target channel's dataset (right). In practice, the functionality can be used in determining cell densities and differences in cell dynamics. In the schematics, the colored circles indicate feature locations of different channels, and the arrows show the nearest features in the channel that is under analysis. (B) The clustering algorithm is based on finding neighbors of each feature on one channel to form "cluster seeds." The seeds are then merged on the basis of shared feature IDs to form the final clusters (blue circles). In the figure, the centroid of feature number 1 falls within the cluster seed of feature 0, whereas feature 2 does not. However, as feature 2 is within the proximity of feature 1, during the merging of seeds all these numbered features are joined into one cluster. (C) Pairwise sample group comparisons in LAM. All groups are first analyzed alone and then compared against the control group. LAM analysis can include any number of sample groups, but each group is statistically tested only against the control group.

(<https://github.com/hietakangas-laboratory/Stitch>). Finally, LAM is accompanied by a step-by-step guide, tutorial videos (<https://www.youtube.com/playlist?list=PLjv-8Gzxh3AynUtI3HaahU2o ddMbDpgtx>), and a user-friendly graphical user interface.

### Region-specific cellular profiling of the steady-state *Drosophila* midgut

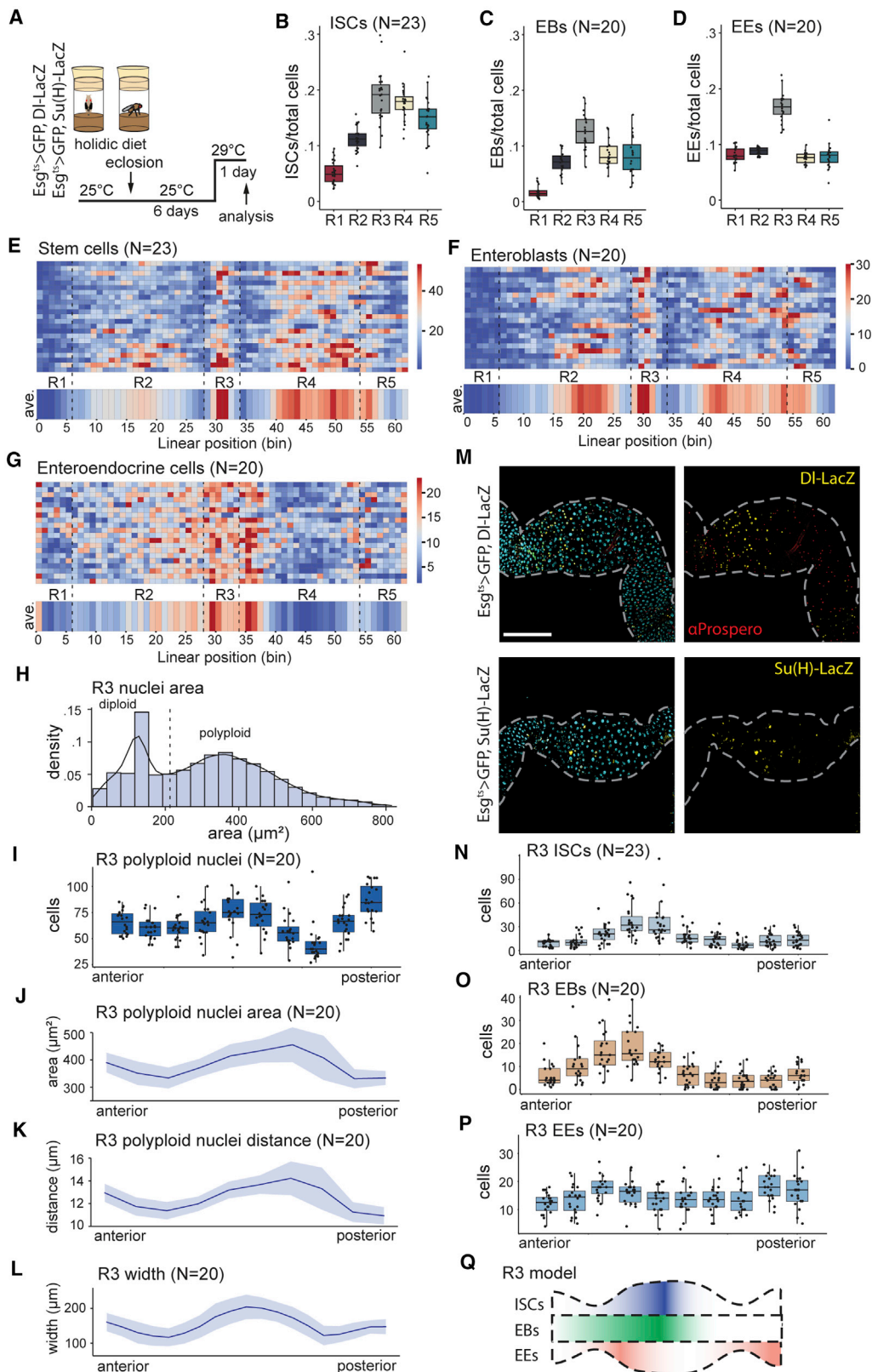
To date, no quantitative data on regional distribution of cell types in a steady-state midgut are available. We used LAM to establish

such a dataset for mated young (7 days old) females, grown on chemically defined holidic medium (Figure 4A) (Piper et al., 2014). With the chosen experimental settings, we expect the midguts to be in a gradually renewing steady state. The border-detection algorithm was used to identify regions R1–R5. To identify intestinal cell types, we used specific markers for ISCs (Delta-LacZ), EBs (Su(H)-LacZ), and EE cells (anti-Prospero) along with Esg-Gal4,UAS-GFP,tub-Gal80<sup>ts</sup> (Esg<sup>ts</sup>), which marks ISCs and EBs (Figures S1A and S1B) (Jiang et al., 2009). The relative (normalized to total cell number) and total cell numbers within regions R1–R5 were calculated (Figures 4B–4D and S1C–S1F). The analysis shows clear regional variation in the proportional numbers of distinct cell types—for example, the EE cells were most concentrated in R3 (Figure 4D). The overall regional pattern of Delta-positive ISC and Su(H)-positive EB distributions largely overlap with each other (Figures 4E and 4F). The relative number of ISCs and EBs are high in the middle, and the posterior of R4 (corresponding to R4bc) as well as in the anterior R5 (corresponding to the R5a). In R2, ISCs and EBs are most abundant in the middle of the region (corresponding to the R2b). Notably, R1 contains very low numbers of ISCs and EBs compared with the rest of the midgut (Figures 4E and 4F).

As the LAM analysis was performed at the resolution of 62 bins per midgut, we were able to identify even more fine-structured patterns of cellular distribution. For example, R3 is divided into the acid-secreting CCR and the LFC region flanked by intestinal constrictions. Plotting the polyploid EC nuclei number, area, nuclei-to-nuclei distance, and midgut width revealed typical topology of the CCR and LFC along the R3 region (Figures 4H–4L). Interestingly, the anterior side of R3, composed of the CCR, displayed high relative numbers of ISCs and EBs. However, their respective distributions within this region differed slightly: ISCs were most abundant in the middle and posterior parts of CCR, whereas EBs were primarily clustered in the anterior end of the CCR, adjacent to the R2/R3 border (Figures 4M–4O). This is in line with the findings that the CCR can be subdivided into molecularly distinct regions (Strand and Micchelli, 2011) and suggests the existence of localized signals directing the balance between stem cell renewal and differentiation in the CCR. In addition to ISCs and EBs, EE cells displayed specific patterns in the middle midgut (Figures 4M and 4P). A high density of EE cells was present in a narrow stripe at the anterior CCR, as well as directly after the R3/R4 border (Figure 4P). The latter stripe corresponded to the so-called iron cell region, which contains enterocytes highly expressing the iron storage protein Ferritin (Marianes and Spradling, 2013). Interestingly, additional

group (n = 32). The red line is the median score of the sample group and black lines are individual samples. Although individual samples have great variation in score, grouping of samples leads to emergence of trends that can be used for peak detection. (H) Peak detection performed on a sample group's median scores (red line) shows approximate locations of border regions, as defined by value changes in multiple variables. The group's score is smoothed and rescaled to [0,1] for peak detection. The vertical red lines at peak locations show their prominence. The marked borders from left to right are B1, B2, B3, and B4.

(I–K) Anchoring of midgut samples for regional alignment. (I) Midpoint anchoring. Using a single anchoring point in a distinct morphological site, such as the border between copper cells and large flat cells, results in accurate alignment close to the anchoring point but propagates error toward the distal regions. The anchoring point is a user-defined image coordinate that is projected onto the normalized [0,1] vector. The vector is then divided into a user-defined number of bins that is equal for each sample. The samples are aligned within a data matrix by assigning them to indices according to the bin of their projected anchoring point. Note the unequal alignment of the midgut ends due to varying proportions of regions, and variable lengths at either side of the anchoring point. (J) Endpoint anchoring. Aligning the samples from both ends propagates the error toward the middle of the midgut. In this method, a user-defined anchoring point is not necessary. (K) Split and combine anchoring. In this method, border peak analysis determines vector cut points. This allows splitting, realigning, and rejoining of the vectors with more accurate regional comparison of different midgut samples.



(legend on next page)



enrichments of the EE cells were observed in the distal ends of the midgut, at the border between the crop and R1, and at the border between the midgut and hindgut (Figure 4G). Taken together, profiling of the cellular distributions along the steady-state midgut A/P axis by LAM revealed unprecedented patterns of cell organization and demonstrated the performance of LAM in quantitative analysis of subregional phenotypic features (Figure 4Q).

### DSS feeding results in regional changes to midgut morphology and ISC differentiation

As a further proof of concept of the functionalities in LAM, we analyzed the injury response of ISCs in a widely used colitis model, oral administration of DSS (Figure 5A). DSS treatment has been reported to induce regenerative ISC proliferation and accumulation of Su(H)-positive enteroblasts, and there were no significant changes in numbers of Delta- or Prospero-positive cells (Amcheslavsky et al., 2009). An analysis of the morphological features of the midgut revealed that DSS feeding results in significant, region-specific changes in midgut morphology. Midgut width and length were affected in several regions, especially in R3 and R4, R3 displaying the strongest relative reduction in midgut length (Figures 5B and 5C). Furthermore, the size and patterning of nuclei were altered in a region-specific manner (Figure 5D). These changes somewhat compromised border detection, in particular preventing reliable detection of the first border (B1, Figure 5E). One of the most striking consequences of DSS feeding was the prominent reduction of R3 cell numbers (Figures 5F and 5G). This implies that the ECs of R3 are more sensitive and/or that the R3 ISCs are not equally capable of maintaining homeostatic regeneration upon DSS treatment. In line with EC loss in R3, DSS treatment resulted in significant loss of polyploid cells, whereas the number of smaller diploid nuclei was less affected (Figure 5H). Consistent with the notion of the stem cells' inability to divide and compensate for cell death, the number of ISC-derived GFP-marked cells was not significantly increased in the R3 region upon acute DSS treatment (Figure 5I). As a consequence of these changes, the typical subregional R3 morphology, including differential patterning and number of ECs in the CCR and LFC region, was lost in the DSS-treated flies (Figures 5J and 5K). Altogether, based on the anal-

ysis of the morphological and cellular parameters, our results indicate severe sensitivity of the R3 region to acute DSS treatment concomitant with impaired stem cell activation to compensate for the cell loss.

To further investigate the regional heterogeneity of ISC differentiation during DSS-induced injury, we used cell-type-specific markers for ISCs (Delta-LacZ), EBs (Su(H)-LacZ), and EE cells (anti-Prospero). Consistent with earlier findings (Amcheslavsky et al., 2009), DSS treatment led to accumulation of Su(H)-positive EBs (Figures 6A and 6B). However, the accumulation of EBs displayed region-specific differences, being most prominent in R5 and particularly low in R1, and in the anterior parts of R4 (Figure 6B). In contrast to the previous report (Amcheslavsky et al., 2009), we detected widespread accumulation of Delta-positive ISCs, especially in R2 and R4 (Figures 6C and 6D). Notably, the regional pattern of Delta- and Su(H)-positive cells did not fully correlate. This might be explained by a regional difference in the prevalence of symmetric ISC-ISC divisions (high Delta in the R4) and asymmetric ISC-EB divisions (high Su(H) in the R5) (Figures 6B and 6D). Interestingly, we also noticed that the nuclei of the Su(H)-positive cells were significantly larger in R5 compared with R4. This might reflect impaired regulation of the Notch signaling pathway with failure to switch off Su(H) expression in the differentiating enterocytes in the R5 region (Figures 6A and 6E). Consistent with the low amount of stem cells in R1 during steady state, few Delta-positive cells were detected in the anterior parts of the midgut after the DSS treatment (Figure 6D). The levels of Prospero-positive EE cells remained stable upon the DSS treatment in most of the midgut area (Figure 6F). Interestingly, however, an area ranging from posterior R4 to anterior R5 displayed significantly elevated numbers of Prospero-positive cells after the DSS treatment (Figure 6F). In conclusion, the DSS-induced injury response displays prominent regional heterogeneity in terms of stem cell activation, division, and differentiation profiles.

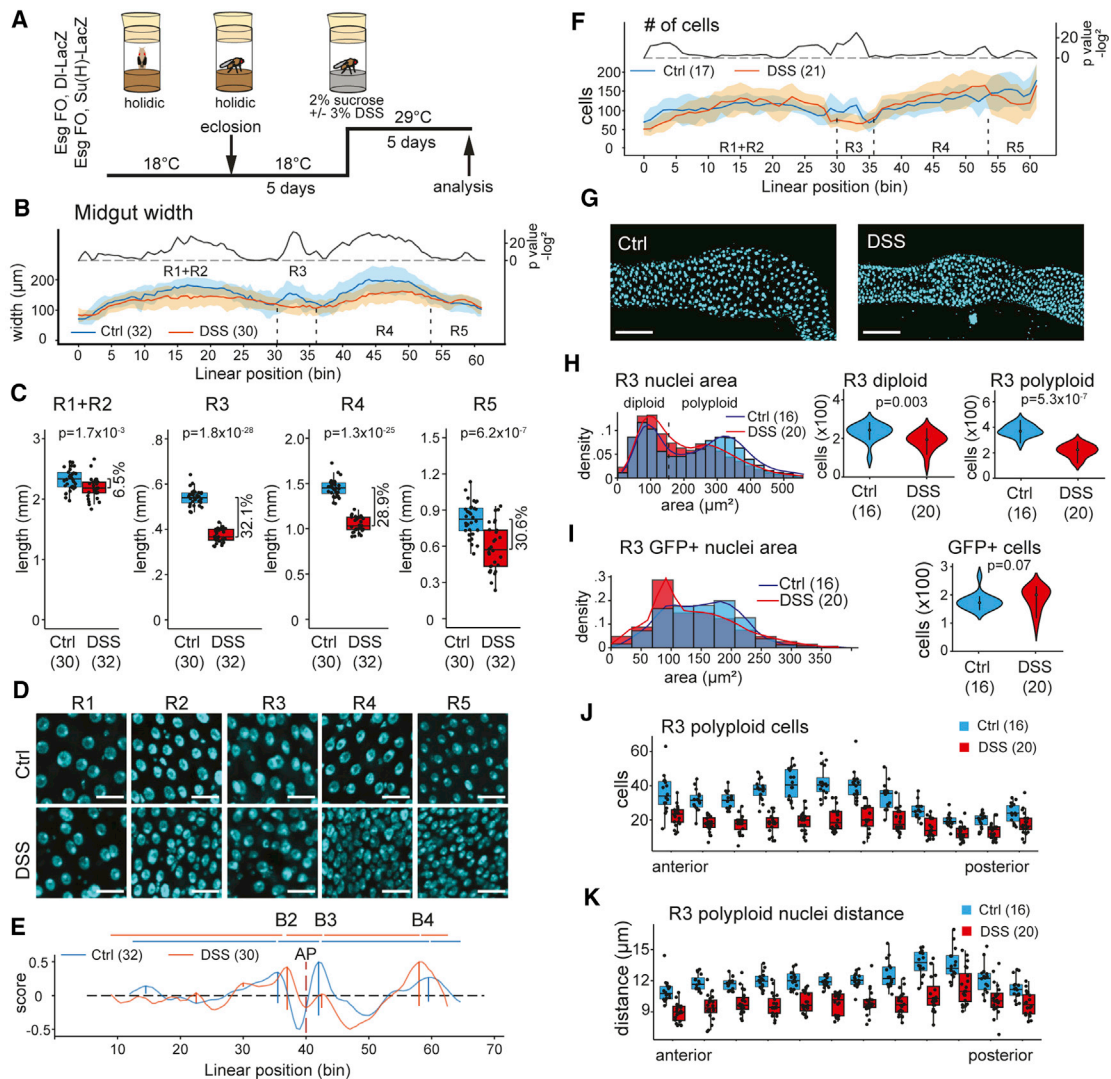
### DISCUSSION

Here, we present an approach to quantitatively study cellular phenotypes of the whole *Drosophila* midgut. In combination with fast tile scan imaging and efficient image feature detection

#### Figure 4. Cellular profiling of a steady-state midgut by LAM

- (A) Experimental design used for the regional steady-state midgut profiling. Age-matched, mated females of *Esg-Gal4<sup>ts</sup>*, *UAS-GFP*, *Delta-LacZ* or *Esg-Gal4<sup>ts</sup>*, *UAS-GFP*, and *Su(H)-LacZ* genotypes were kept at +25°C for 6 days and then shifted to +29°C for 1 day.
- (B–D) Relative numbers of ISCs (B), EBs (C), and EE cells (D) in R1–R5 regions, calculated as the number of specific cells per total number of cells.
- (E–G) Sample and average heatmaps of cellular distributions along the midgut A/P axis for ISCs (E), EBs (F), and EE cells (G).
- (H) Area distribution of R3 nuclei.
- (I) Polyploid nuclei number along the R3 A/P axis.
- (J) Polyploid nuclei area along the R3 A/P axis. Light-blue shading is the standard deviation.
- (K) Average polyploid nuclei-to-nuclei distance along the R3 A/P axis. Light-blue shading is the standard deviation.
- (L) R3 width along the A/P axis. Light-blue shading is the standard deviation.
- (M) Representative images of R3 region showing the localization of *DI-lacZ*-positive ISCs and Prospero-positive EE cells (upper panels) and *Su(H)-lacZ*-positive EBs (lower panels). DNA is stained with DAPI and is shown in cyan. Scale bar, 100 μm.
- (N) ISC number along the R3 A/P axis.
- (O) EB number along the R3 A/P axis.
- (P) EE cell number along the R3 A/P axis.
- (Q) Schematic model displaying the steady-state distribution of ISCs, EBs, and EE cells in the R3 region. Segmentation of the images was performed by Imaris software. See also Figure S1.





**Figure 5. DSS feeding results in regional changes to midgut morphology**

(A) Experimental design of the DSS feeding experiment. Age-matched, mated females of Esg FO, UAS-GFP, Delta-LacZ or Esg FO, UAS-GFP, and Su(H)-LacZ genotypes were kept at the restrictive temperature (+18°C) for 5 days and then shifted to the permissive temperature (+29°C) to induce the flip-out clones in the presence of 3% DSS. The intestines were then analyzed after 5 days.

(B) LAM width profile of the control and DSS-treated midguts. Light-blue and orange shadings are the standard deviations. p values are calculated by Mann-Whitney-Wilcoxon U test using a 3-bin window.

(C) Length of R1 + R2, R3, R4, and R5 midgut regions of control and DSS-fed flies.

(D) Representative images of R1–R5 midgut regions of control and DSS-fed flies. DNA is stained with DAPI and is shown in cyan. Scale bar, 20 μm.

(E) Border peak analysis of midguts of control and DSS-fed flies.

(F) LAM profile of number of cells between control and DSS-treated midguts. Light-blue and orange shadings are the standard deviations. p values are calculated by Mann-Whitney-Wilcoxon U test using a 3-bin window.

(G) Representative images of the R3 region of control (left panel) and DSS-fed (right panel) flies. DNA is stained with DAPI and is shown in cyan. Scale bar, 100 μm.

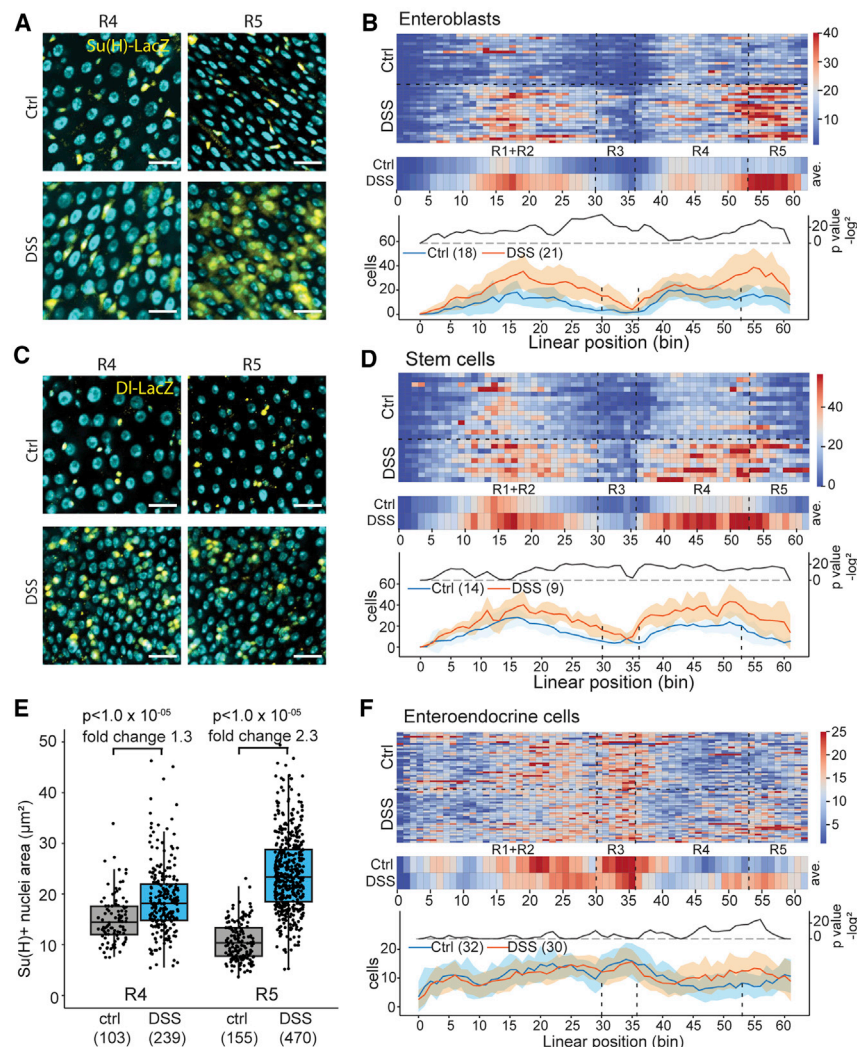
(H) Area distribution of R3 nuclei in midguts of control and DSS-fed flies. Comparison of the number of R3 diploid and polyloid nuclei between midguts of control and DSS-fed flies.

(I) Area distribution of R3 GFP-positive nuclei in midguts of control and DSS-fed flies. Comparison of the number of R3 GFP-positive nuclei between midguts of control and DSS-fed flies.

(J) Number of polyloid nuclei in R3. Anterior to the left, posterior to the right.

(K) Mean distance between polyloid nuclei in R3. Anterior to the left, posterior to the right.

p values in (C) are calculated by the two-sample t test. p values in (H) and (I) are calculated by Mann-Whitney-Wilcoxon U test. Segmentation of the images was performed by Imaris software except in (F), where StarDist was used.



**Figure 6. DSS feeding results in regional changes to ISC differentiation**

(A) Representative images of the R4 (left panels) and R5 (right panels) regions of control and DSS-fed midguts from Esg-Gal4<sup>1s</sup>, UAS-GFP, and Su(H)-LacZ flies. DNA is stained with DAPI and is shown in cyan. Scale bar, 20  $\mu\text{m}$ .

(B) Regional quantification of Su(H)-LacZ-positive EBs of midguts of control and DSS-fed flies.

(C) Representative images of the R4 (left panels) and R5 (right panels) regions of control and DSS-fed midguts from Esg-Gal4<sup>1s</sup>, UAS-GFP, and Delta-LacZ flies. DNA is stained with DAPI and is shown in cyan. Scale bar, 20  $\mu\text{m}$ .

(D) Regional quantification of Delta-LacZ-positive ISCs of midguts of control and DSS-fed flies.

(E) Nuclei area quantification of Su(H)-positive cells in R4 and R5 regions.

(F) Regional quantification of Prospero-positive EE cells of midguts of control and DSS-fed flies.

p values in (E) are calculated by two-way ANOVA followed by Tukey's test. p values in (B), (D), and (F) are calculated by Mann-Whitney-Wilcoxon U test using a 3-bin window. Segmentation of the images was performed by Imaris software.

algorithms, LAM enables, for the first time, quantitative and regionally defined automated phenotyping of all cells in the whole midgut. LAM allows (1) coupling of cellular identities to a specific position along the A/P axis, (2) automated detection of regional boundaries, and consequently (3) quantitative and statistical analysis of cellular phenotypes along the regions of the midgut, with subregional resolution. In doing so, LAM (4) opens the path for organ-wide studies of midgut cells and eliminates the bias caused by selective analysis of a specific midgut area. Through these advances, LAM will allow the exploration of regional heterogeneity of midgut cells, including the ISCs, and will significantly increase the representativeness of midgut phenotypic data. The graphical user interface makes LAM accessible even for scientists with limited experience in computational image analysis.

### Variations in regional distribution of midgut cells

We tested the performance of LAM by analyzing the distribution of ISCs, EBs, and EE cells in a steady-state midgut of mated young females. This analysis revealed several new features of

high variation of phenotypes between individual midguts, it would have not been possible to reliably detect such features by qualitative analysis of individual midguts. This demonstrates the ability of LAM to detect variable phenotypes with high subregional resolution. The resolution of LAM is influenced by the numbers of bins, which can be freely adjusted by the user. The optimal number of bins depends on the density of input data points as well as data quality, which influences the accuracy of alignment of individual midguts.

### Regional heterogeneity of the injury response in a *Drosophila colitis* model

As another proof of principle, we employed a widely used *Drosophila colitis* model induced by DSS feeding. Use of LAM allowed us to identify several new features of stem cell activation and differentiation not previously documented in the literature, providing new insight into the previously reported models of midgut injury response (Amcheslavsky et al., 2009; Jiang et al., 2016). We noticed a significant reduction of the total cell numbers in R3, which coincided with low activation of stem cells in R3 when

cellular distributions, including partially overlapping clusters of Delta- and Su(H)-positive cells within the CCR/R3ab subregion. In addition, EE cells were observed to cluster around the main regional boundaries, including the cardia-R1, R2-R3, R3-R4, and R5-hindgut boundaries, suggesting a common regional organizer for the specification of EE cell fate in these regions. One such signal could be the Wg signaling pathway, whose activity has been shown to localize to these regions (Tian et al., 2016). Notably, due to the

compared with the neighboring R2 and R4 regions. As our experiment focused on an acute 5-day DSS response, it remains possible that R3 ISCs react with slower kinetics or that R3 ISCs differentially depend on other environmental factors, such as nutrition. In fact, a previous study has demonstrated that the R3 stem cells are capable of inducing a regenerative response during a slightly longer (>1 week) DSS administration period on a diet with 20% sucrose (in contrast to 2% in our study) (Wang et al., 2014). Outside R3, DSS treatment led to a relatively uniform increase in Delta-positive cells, except in R1, which contains fewer Delta-positive cells to start with. Notably, our data differ from those of an earlier study (Amcheslavsky et al., 2009) showing no accumulation of Delta-positive cells upon acute DSS treatment. A possible technical explanation for the discrepancy is the perdurance of the  $\beta$ -Galactosidase protein, expressed by the Delta-LacZ reporter. Whereas the overall pattern of Su(H)-positive cells was similar to that of the Delta-positive cells, the posterior midgut displayed interesting quantitative differences. Most of R4 showed only a modest increase in Su(H)-positive cells, but an area from the posterior end of R4 to R5 displayed a very high increase in relative numbers of EBs in response to DSS. Comparison of the regional profiles of Delta-LacZ and Su(H)-LacZ reporters is consistent with the conclusion that the differentiation rates of ISCs display significant regional differences, with the ISCs of R5 being more prone to EB fate. In addition, the Su(H)-positive cells in the R5 region of the DSS-treated midguts showed enlarged nuclei compared with the EBs in other regions. Although the molecular details explaining the difference in ISC fate between R4 and R5 are as yet unknown, regional transcriptome mapping has revealed existing gene expression differences between the ISCs of these regions (Dutta et al., 2015). One candidate in regulating ISC fate in these regions is the transcription factor Snail, whose expression is relatively high in R5 ISCs. Forced expression of Snail prevented EB differentiation into ECs, leading to an accumulation of EBs (Dutta et al., 2015). Hence, it will be interesting to learn whether intrinsic differences in Snail expression, or possible region-specific extrinsic factors, underlie the region-specific differentiation patterns in the injured midgut.

### Concluding remarks

In addition to the physiology of midgut regionality, the unbiased organ-wide analysis with LAM can improve representativeness of midgut data in general. Considering the concern of confirmation bias throughout the scientific literature, there is a risk that studies focusing on a narrow (often undefined) area of the midgut primarily record and present data from areas that give the strongest phenotypes. Considering our DSS experiment, a focused analysis of only one (sub)region would have yielded several different, and sometimes even mutually contradictory, biological conclusions, depending on the region chosen. Therefore, one should exercise caution when making generalized conclusions based on the findings of a small subset of ISCs. We propose an approach whereby the phenotypic response for a given treatment/genotype is first quantitatively analyzed and reported at the level of the whole midgut, with more detailed follow-up experiments concentrated on the specific region(s) of interest. In conclusion, we expect that the unbiased organ-wide analysis offered by LAM will allow the pursuit of more representative

data and uncover the extent of tissue context-dependence of stem cell regulation as well as increasing the understanding of the physiological roles of intestinal regionalization.

### Limitations of study

The performance of LAM is dependent on the quality of the midgut preparations, image acquisition, and image segmentation for cellular objects. Each step is to be carefully considered for successful application of LAM. In our experiments, the intestines were mounted between a microscope slide with 0.12-mm spacers and a coverslip. Images were obtained to capture half of the midgut circumference, thus assuming the cellular heterogeneity to be equal on each side. When mounted, the midgut is not always equally flat along the A/P axis. Special care is needed to avoid disproportional recording of the midgut circumference in different regions. To circumvent any bias from disproportional imaging, it is possible to extend the z stacks to include the full circumference of the midgut if required. Although LAM allows the recording of all imaged cells, the projection of objects to the midline vector is dimensionally restricted, and LAM does not account for orientation on the z axis. Consequently, the information on cell stratification as well as the 3D geometry of the intestinal cylinder is not used during object counting. z-axis coordinates are, however, taken into account when calculating the object distances and clustering, allowing reliable data acquisition around the intestinal circumference.

The algorithm for detecting region borders is based on the morphological characteristics of the regions, such as midgut constrictions, as well as the nuclear size and distance, which were previously applied to manually map borders between physiologically distinct compartments (Buchon et al., 2013). Although the parallel use of multiple parameters increases the robustness of border detection, it remains possible that experimental conditions, such as those influencing visceral muscle function or changing the ratio between cell types (e.g., EB accumulation), might influence the border detection. As was the case with DSS treatment, a subset of borders can often still be reliably detected. Specific ad hoc solutions, such as region-specific GFP traps or elimination of EBs from the analysis (by using a marker), might be applied under such circumstances. Object segmentation is a critical step for calculating the nuclear features characteristic of different regions. The performance of the traditional object segmentation methods, such as intensity thresholding, is compromised by high cellular densities, cell-size variation, cell stratification, and intensity differences. The variation in the success of nuclear segmentation possibly hampered our attempts to reliably detect regional borders from individual midguts. We overcame this limitation by applying sample group average values to locate the borders of individual samples. Deep-learning-based nuclear segmentation algorithms, such as StarDist (Schmidt et al., 2018; Weigert et al., 2020), are likely to further improve the accuracy.

### STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY



- Lead contact
- Materials availability
- Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - *Drosophila* stocks and husbandry
- **METHOD DETAILS**
  - DSS treatment
  - Immunohistochemistry
  - Microscopy and image processing
  - Methods in LAM
  - Vector creation
  - Projection and counting
  - Feature-to-feature distances
  - Feature clustering
  - Gut width measurement
  - Automatic border detection
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Statistics in LAM
  - Other statistical analysis
- **ADDITIONAL RESOURCES**

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.crmeth.2021.100059>.

### ACKNOWLEDGMENTS

We thank Bloomington *Drosophila* Stock Center and Bruce Edgar for providing fly stocks. Heini Lassila, Jorge Fuentes Rodriguez, and Esa Pitkänen are acknowledged for technical assistance. We thank Paula Jouhten for critical reading of the manuscript. Funding was provided by the Academy of Finland through the MetaStem Center of Excellence funding to V.H. (312439) and a project grant (137530) to J. Mattila, Sigrid Jusélius Foundation (V.H. and J. Mattila), Erkkö Foundation (V.H. and P.K.), and Novo Nordisk Foundation (NNF19OC0057478 to V.H.). This study was facilitated by the University of Helsinki *Drosophila* core facility (Hi-Fly) and the Light Microscopy Unit supported by Biocenter Finland and Helsinki Institute of Life Science.

### AUTHOR CONTRIBUTIONS

Conceptualization, A.V., J. Mattila, and V.H.; methodology, A.V. and J. Mattila; software, A.V.; investigation, J. Mattila, J.G., and J. Morikka; data curation, A.V. and J. Mattila; writing – original draft, A.V., J. Mattila, and V.H.; writing – review & editing, A.V., J. Mattila, P.K., and V.H.; visualization, A.V. and J. Mattila; supervision, V.H.; funding acquisition, V.H.

### DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 16, 2021  
Revised: June 1, 2021  
Accepted: June 30, 2021  
Published: July 30, 2021

### REFERENCES

Amcheslavsky, A., Jiang, J., and Ip, Y.T. (2009). Tissue damage-induced intestinal stem cell division in *Drosophila*. *Cell Stem Cell* 4, 49–61.  
Biteau, B., Hochmuth, C.E., and Jasper, H. (2008). JNK activity in somatic stem cells causes loss of tissue homeostasis in the aging *Drosophila* gut. *Cell Stem Cell* 3, 442–455.

Buchon, N., Broderick, N.A., Chakrabarti, S., and Lemaître, B. (2009). Invasive and indigenous microbiota impact intestinal stem cell activity through multiple pathways in *Drosophila*. *Genes Dev.* 23, 2333–2344.  
Buchon, N., Osman, D., David, F.P.A., Yu Fang, H., Boquete, J.-P., Deplancke, B., and Lemaître, B. (2013). Morphological and molecular characterization of adult midgut compartmentalization in *Drosophila*. *Cell Rep.* 3, 1725–1738.  
Dimitriadis, V.K. (1991). Fine structure of the midgut of adult *Drosophila auraria* and its relationship to the sites of acidophilic secretion. *J. Insect Physiol.* 37, 167–177.  
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.  
Furriols, M., and Bray, S. (2001). A model Notch response element detects Suppressor of Hairless-dependent molecular switch. *Curr. Biol.* 11, 60–64.  
Gervais, L., and Bardin, A.J. (2017). Tissue homeostasis and aging: new insight from the fly intestine. *Curr. Opin. Cell Biol.* 48, 97–105.  
Gillies, S. (2007). Shapely: manipulation and analysis of geometric objects. <https://github.com/Toblerity/Shapely>.  
Guo, X., Yin, C., Yang, F., Zhang, Y., Huang, H., Wang, J., Deng, B., Cai, T., Rao, Y., and Xi, R. (2019). The cellular diversity and transcription factor code of *Drosophila* enteroendocrine cells. *Cell Rep.* 29, 4172–4185.e5.  
Guo, Z., Lucchetta, E., Rafel, N., and Ohlstein, B. (2016). Maintenance of the adult *Drosophila* intestine: all roads lead to homeostasis. *Curr. Opin. Genet. Dev.* 40, 81–86.  
Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., et al. (2020). Array programming with NumPy. *Nature* 585, 357–362.  
Hudry, B., Khadayate, S., and Miguel-Aliaga, I. (2016). The sexual identity of adult intestinal stem cells controls organ size and plasticity. *Nature* 530, 344–348.  
Hunter, J.D. (2007). Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95.  
Imaris, Imaris (version 9.5.1). (2019). Bitplane Inc. <http://bitplane.com>.  
Jiang, H., and Edgar, B.A. (2009). EGFR signaling regulates the proliferation of *Drosophila* adult midgut progenitors. *Development* 136, 483–493.  
Jiang, H., Patel, P.H., Kohlmaier, A., Grenley, M.O., McEwen, D.G., and Edgar, B.A. (2009). Cytokine/Jak/Stat signaling mediates regeneration and homeostasis in the *Drosophila* midgut. *Cell* 137, 1343–1355.  
Jiang, H., Tian, A., and Jiang, J. (2016). Intestinal stem cell response to injury: lessons from *Drosophila*. *Cell Mol. Life Sci.* 73, 3337–3349.  
Marianes, A., and Spradling, A.C. (2013). Physiological and stem cell compartmentalization within the *Drosophila* midgut. *eLife* 2. <https://doi.org/10.7554/eLife.00886>.  
McKinney, W. (2010). Data structures for statistical computing in python. In *Procedures of the 9th Python in Science Conference (SCIPY 2010)*, pp. 56–61.  
Miguel-Aliaga, I., Jasper, H., and Lemaître, B. (2018). Anatomy and physiology of the digestive tract of *Drosophila melanogaster*. *Genetics* 210, 357–396.  
Missiaglia, E., Jacobs, B., D'Ario, G., Di Narzo, A.F., Sonesson, C., Budinska, E., Popovici, V., Vecchione, L., Gerster, S., Yan, P., et al. (2014). Distal and proximal colon cancers differ in terms of molecular, pathological, and clinical features. *Ann. Oncol.* 25, 1995–2001.  
Mowat, A.M., and Agace, W.W. (2014). Regional specialization within the intestinal immune system. *Nat. Rev. Immunol.* 14, 667–685.  
O'Brien, L.E. (2013). Regional specificity in the *Drosophila* midgut: setting boundaries with stem cells. *Cell Stem Cell* 13, 375–376.  
Piper, M.D.W., Blanc, E., Leitão-Gonçalves, R., Yang, M., He, X., Linford, N.J., Hoddinott, M.P., Hopfen, C., Soultoukis, G.A., Niemeyer, C., et al. (2014). A holistic medium for *Drosophila melanogaster*. *Nat. Methods* 11, 100–105.  
Preibisch, S., Saalfeld, S., and Tomancak, P. (2009). Globally optimal stitching of tiled 3D microscopic image acquisitions. *Bioinformatics* 25, 1463–1465.



- Reiff, T., Jacobson, J., Cognigni, P., Antonello, Z., Ballesta, E., Tan, K.J., Yew, J.Y., Dominguez, M., and Miguel-Aliaga, I. (2015). Endocrine remodelling of the adult intestine sustains reproduction in *Drosophila*. *eLife* 4, e06930.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682.
- SciPy 1.0 Contributors; Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272.
- Schmidt, U., Weigert, M., Broaddus, C., and Myers, G. (2018). In MICCAI 2018. Lecture Notes in Computer Science, 11071, J. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, eds. (Springer).
- Seabold, S., and Perktold, J. (2010). Statsmodels: econometric and statistical modeling with Python. In Procedures of the 9th Python in Science Conference (SCIPY 2010), pp. 92–96.
- Strand, M., and Micchelli, C.A. (2011). Quiescent gastric stem cells maintain the adult *Drosophila* stomach. *Proc. Natl. Acad. Sci. U S A* 108, 17696–17701.
- Tian, A., Benchabane, H., Wang, Z., and Ahmed, Y. (2016). Regulation of stem cell proliferation and cell fate specification by wingless/wnt signaling gradients enriched at adult intestinal compartment boundaries. *PLoS Genet.* 12, e1005822.
- van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., and Yu, T. (2014). scikit-image: image processing in Python. *PeerJ* 2, e453.
- Waskom, M., Gelbart, M., Botvinnik, O., Ostblom, J., Hobson, P., Lukauskas, S., Gemperline, D.C., Augspurger, T., Halchenko, Y., Warmenhoven, J., et al. (2020). mwaskom/seaborn: v0.11.1 (Zenodo). <https://zenodo.org/record/4379347#.YOVh-ugzZPY>.
- Wang, C., Guo, X., and Xi, R. (2014). EGFR and Notch signaling respectively regulate proliferative activity and multiple cell lineage differentiation of *Drosophila* gastric stem cells. *Cell Res.* 24, 610–627.
- Weigert, M., Schmidt, U., Haase, R., Sugawara, K., and Myers, G. (2020). Star-convex polyhedra for 3D object detection and segmentation in microscopy. In IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 3655–3662.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Rabbit IgG fraction against $\beta$ -galactosidase	MP Bio	Cat# 0855976; RRID: AB_2334934
Anti-Prospero antibody	DSHB	Cat# Pros(MR1A); RRID: AB_528440
Vectashield mounting medium with DAPI	Vector Laboratories	Cat# H-1200; RRID: AB_2336790
Dextran sulfate sodium salt	MP Bio	CAS: 9011-18-1
<b>Deposited data</b>		
Segmentation data (figures)	This paper	DOI: <a href="https://doi.org/10.23729/fac30a2e-1de5-45d6-879e-f67129bc822d">https://doi.org/10.23729/fac30a2e-1de5-45d6-879e-f67129bc822d</a> ; URL: <a href="https://etsin.fairdata.fi/dataset/78fc1f12-f4d0-4d5a-b2ba-d72f9998145b">https://etsin.fairdata.fi/dataset/78fc1f12-f4d0-4d5a-b2ba-d72f9998145b</a>
<b>Experimental models: Organisms/strains</b>		
Delta-LacZ	Bloomington	Cat# 11651; RRID: BDSC_11651
Gbe + Su(H)-lacZ	<a href="#">Furriols and Bray, 2001</a>	N/A
w; Esg-Gal4, Tub-Gal80-ts, UAS-GFP (Esg <sup>ts</sup> )	<a href="#">Jiang and Edgar, 2009</a>	N/A
Esg <sup>ts</sup> ; UAS-Flp, Act>CD2>Gal4 (Esg FO)	<a href="#">Jiang et al., 2009</a>	N/A
<b>Software and algorithms</b>		
Fiji/ImageJ	<a href="#">Schindelin et al., 2012</a>	RRID: SCR_003070; URL: <a href="https://imagej.net/Fiji">https://imagej.net/Fiji</a>
ImageJ Grid/Collection plugin	<a href="#">Preibisch et al., 2009</a>	URL: <a href="https://imagej.net/Grid/Collection_Stitching_Plugin">https://imagej.net/Grid/Collection_Stitching_Plugin</a>
Imaris 9.5.1	Bitplane Inc., 2019	RRID: SCR_007370; URL: <a href="https://imaris.oxinst.com/">https://imaris.oxinst.com/</a>
LAM 0.4.2	This paper	Zenodo: <a href="https://doi.org/10.5281/zenodo.5037180">https://doi.org/10.5281/zenodo.5037180</a> ; GitHub: <a href="https://github.com/hietakangas-laboratory/LAM">https://github.com/hietakangas-laboratory/LAM</a>
LAM-helper-modules 1.0.0	This paper	Zenodo: <a href="https://doi.org/10.5281/zenodo.5037000">https://doi.org/10.5281/zenodo.5037000</a> ; GitHub: <a href="https://github.com/hietakangas-laboratory/LAM-helper-modules">https://github.com/hietakangas-laboratory/LAM-helper-modules</a>
predictSD 1.0.0	This paper	Zenodo: <a href="https://doi.org/10.5281/zenodo.5036995">https://doi.org/10.5281/zenodo.5036995</a> ; GitHub: <a href="https://github.com/hietakangas-laboratory/predictSD">https://github.com/hietakangas-laboratory/predictSD</a>
Python 3.7	Python Software Foundation	RRID: SCR_008394; URL: <a href="https://www.python.org/">https://www.python.org/</a>
StarDist 0.6.2	<a href="#">Schmidt et al., 2018</a> ; <a href="#">Weigert et al., 2020</a>	GitHub: <a href="https://github.com/stardist/stardist">https://github.com/stardist/stardist</a>
Stitch 1.0.0	This paper	Zenodo: <a href="https://doi.org/10.5281/zenodo.5037000">https://doi.org/10.5281/zenodo.5037000</a> ; GitHub: <a href="https://github.com/hietakangas-laboratory/Stitch">https://github.com/hietakangas-laboratory/Stitch</a>
<b>Other</b>		
Holidic Drosophila diet	<a href="#">Piper et al., 2014</a>	N/A

## RESOURCE AVAILABILITY

### Lead contact

- Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Ville Hietakangas ([ville.hietakangas@helsinki.fi](mailto:ville.hietakangas@helsinki.fi)).

### Materials availability

- This study did not generate new unique reagents.

### Data and code availability

- The raw image data reported in this study cannot be deposited in a public repository because of file size. To request access, contact Ville Hietakangas ([ville.hietakangas@helsinki.fi](mailto:ville.hietakangas@helsinki.fi)). In addition, segmentation data on cell-like objects from the raw images have been deposited at IDA-database and are publicly available as of the date of publication. URL and DOI are listed in the [key resources table](#).
- All original code has been deposited at GitHub and Zenodo and is publicly available as of the date of publication. LAM is also available on Python Package Index (PyPI). URLs and DOIs are listed in the [key resources table](#).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### *Drosophila* stocks and husbandry

Fly stocks used in this study: w; Esg-Gal4, Tub-Gal80-ts, UAS-GFP ; UAS-Flp, Act>CD2>Gal4 (Esg FO) ([Jiang et al., 2009](#)), w; Esg-Gal4, Tub-Gal80-ts, UAS-GFP (Esg<sup>ts</sup>) ([Jiang and Edgar, 2009](#)), Delta-LacZ (DI-LacZ, Bloomington 11651), Gbe+Su(H)-lacZ (Su(H)-LacZ, [Furriols and Bray, 2001](#)). Flies were maintained at 25°C, on medium containing agar 0.6% (w/v), malt 6.5% (w/v), semolina 3.2% (w/v), baker's yeast 1.8% (w/v), nipagin 2.4%, propionic acid 0.7%.

## METHOD DETAILS

### DSS treatment

36-50 KDa DSS was obtained from Fisher Scientific (cat no. 11424352). Staged Esg FO>UAS-GFP, Delta-LacZ and Esg FO>UAS-GFP, Su(H)-LacZ pupae were collected into vials containing holidic diet ([Piper et al., 2014](#)). After eclosion the flies were kept on the holidic diet for 5 days at 18°C, and then transferred into vials containing 2% sucrose (w/v) in medium containing agar 0.5% (w/v), nipagin 2.4%, and propionic acid 0.7%, in water with or without 3% DSS, and then kept at +29°C for 5 days.

### Immunohistochemistry

For immunofluorescence staining, intestines were dissected in PBS and fixed in 8% paraformaldehyde for 3 hours. Tissues were washed with 0.1% Triton-X 100 in PBS and blocked in 1% bovine serum albumin for 1 h. Subsequently, tissues were stained with anti-β-Galactosidase (1:400) (MP Biomedicals cat no: 0855976-CF) and/or anti-Prospero (1:1000) (MR1A, DSHB) antibodies. The samples were mounted in Vectashield mounting media with DAPI (Vector Laboratories) and imaged using the Aurox clarity confocal system (Aurox).

### Microscopy and image processing

Fixed and immunostained whole midguts were mounted in between a microscope slide with 0.12 μm spacers and a coverslip, followed by tile scan imaging by the Aurox clarity spinning disc confocal microscope from the anterior to posterior end. To reduce the image size and scanning time, stacks of only one side of the flattened midgut epithelium were obtained. For stitching the tiles and image processing in ImageJ ([Schindelin et al. 2012](#)), we generated a python script, “Stitch”, with a graphical user interface (<https://github.com/hietakangas-laboratory/Stitch>). “Stitch” is a programme for stitching together a series of tiff images within a directory, utilizing the ImageJ Grid/Collection plugin ([Preibisch et al., 2009](#)), and performing stitching for multiple directories in a batch process. This programme can stitch together a series of tiff images using only a companion.ome metadata file associated with the tiff series. Alternatively, as in this article, “Stitch” can utilize the tile positions output from the microscope to perform image stitching. Full usage instructions and details are available in the “Stitch” user guide. After stitching and image processing, TIFFs were converted to Imaris (Bitplane) files, and features were obtained by the Imaris spot detection algorithm ([Imaris \(version 9.5.1\) 2019](#)). Raw feature data, including spot surface area measurements, were exported and used as input for LAM for further

analysis (see the LAM user guide for details). Python script for enabling easy export of bulk Imaris .csv files to LAM is available in github (<https://github.com/hietakangas-laboratory/LAM-helper-modules>). The repository also contains python scripts with graphical user interface for exporting manually drawn midgut vectors and anchoring points in Fiji/ImageJ. Notably, LAM input is not restricted to Imaris, but accepts data from any source with at least coordinates and unique object identifiers in wide-format tables. We have included python code for running the deep-learning tool StarDist (Weigert et al. 2020) on 3D midgut images in addition to several pre-trained segmentation models (<https://github.com/hietakangas-laboratory/predictSD>).

### Methods in LAM

Data handling in LAM is performed with NumPy (Harris et al., 2020) and Pandas (McKinney, 2010), while plotting is done using matplotlib (Hunter, 2007) and Seaborn (Waskom et al., 2020). Geometric and image operations are performed with Shapely (Gillies, 2007) and Scikit-image (van der Walt et al., 2014), respectively. Statistics are calculated with scipy.stats (SciPy 1.0 Contributors et al., 2020) and statsmodels (Seabold and Perktold, 2010). The border detection additionally uses scipy.signal (SciPy 1.0 Contributors et al., 2020) for locating regions of high signal. LAM includes an easy-to-use graphical user interface (GUI) with enabling/disabling of related options as well as a default settings file that can be edited at will to control all runs. LAM also supports execution from the command line using a limited scope of arguments. Full description of the usage of LAM and step-by-step instructions can be found in the LAM user guide found in GitHub (<https://github.com/hietakangas-laboratory/LAM>). LAM video tutorials are available at (<https://www.youtube.com/playlist?list=PLjv-8Gzxh3AynUtl3HaahU2oddMbDpgtx>).

### Vector creation

LAM provides two alternative methods for creating piecewise median lines, which we colloquially call vectors, for midgut images: bin-smoothing and skeletonization. The methods provided by LAM require the midguts to be horizontally oriented, but the vectors can alternatively be given as coordinate files without restrictions in orientation. An auxiliary script is provided to rotate data to horizontal orientation. Bin-smoothing of the data is performed by binning the x-axis after which the median of the nuclei co-ordinates is calculated for each bin. Then a piecewise line is created to connect the bin midpoints. The number of bins is a user defined parameter to be adjusted for suitable level of smoothing. In the skeleton vector creation option, the DAPI channel co-ordinate data is first converted into a binary image where each nuclei is resized to one pixel. As a result, a binary matrix is created where pixels of nuclei are marked as one, and empty pixels as zero. The binary image is then processed with resizing, smoothing, binary dilation, as well as hole filling in order to produce a continuous blob (user defined parameters, see user guide for more details). The matrix is then subjected to skeletonization, where pixels of the image are eroded until reduced to pixel-wide structures. The vector starting point is determined as the average of five pixel co-ordinates having the smallest x value. The vector is then drawn from pixel to pixel by scoring pixels within a specified range (find distance in GUI) using the following penalty function:

$$penalty = d_{vector} + d_{point} + |rad| * 10$$

where  $d_{vector}$  is the distance of the pixel to the last co-ordinate of the vector, and  $d_{point}$  is the pixel's distance to the projection point ahead of the last coordinate. The projection point is determined by adding the previous vector progression (distance and direction) into the last vector point. The final scoring component, the modulus of radians, is the difference in direction between the last vector co-ordinate and a pixel compared to a fitted line between the last three vector co-ordinates. The x and y co-ordinates of the pixel with the smallest penalty are then added to the path of the vector, and the next pixels are scored based on these coordinates, and so on until no more pixels are found (Figure 1C).

### Projection and counting

All segmented image objects and their associated data, which we collectively call features, are projected to the vector using linear referencing methods of the *shapely* package. To this end, each feature coordinate is assigned a value based on the normalized distance [0...1] to its nearest coordinate point along the A/P length of the vector. The features can then be counted by dividing the vector into a user-defined number of bins of equal length. The default 62 bins is suitable for standard analysis of midgut cell types, but if studying e.g. cell type subpopulations that are more sparse, the bin number may need to be reduced to avoid number of cells per bin skewing towards zero. In contrast, the number of bins may be increased for better resolution if the data has sufficiently high density of cells of interest. By conserving the bin number between samples, LAM enables building of data matrices for bin-to-bin and windowed statistical comparisons.

Bin-wise comparability between sample groups may be reduced by variation in region length. Consequently, LAM allows the data to be joined using the samples' APs, i.e., linear references of distinguishable points for each midgut, to maximize correspondence of regions within the data matrix. Each sample's data can be centered at a specific index position of the matrix when using individual APs. Alternatively, the vector and data can be cut, re-binned, and recombined at each segment flanked by the APs in the "split and combine" approach. To align the samples region-to-region in the "split and combine" approach, the APs can be obtained using LAM's border detection on the data.



### Feature-to-feature distances

LAM has the option to compute pairwise Euclidean distances between nearest features (Figure 3A). The distances can be calculated between features on one channel, e.g. DAPI, or between two channels, e.g. the distance from each Delta<sup>+</sup>-cell to nearest Pros<sup>+</sup>-cell. The features can additionally be filtered by area, volume, or another user defined variable. The algorithm finds for each feature the shortest distance to another feature in the filtered dataset.

### Feature clustering

LAM also includes an algorithm for cluster analysis that functions in a similar manner to the feature-to-feature distance calculations. Cell clusters in the midgut tend to either take the form of longer strands or a more spherical shape, and consequently defining the clusters by their shapes would be problematic. To overcome this, LAM takes the approach to cluster the cells by their proximity to each other (Figure 3B). For each feature, LAM first finds its neighbors within a user-defined distance inside a constructed k-d tree of the 3D co-ordinate data. Found features are then marked as a “cluster seed”. After all seeds are found, they are merged based on shared feature identification. As a result, unique clusters with no shared features are formed. The clusters can be further filtered by a user-defined number of features, and are finally assigned unique cluster identification numbers.

### Gut width measurement

LAM computes the width of each midgut along its vector. The midgut is binned into segments of equal length, and nuclei with the largest distances to the vector are found. As the vector may not exactly follow the true center of the midgut, the handedness of the nuclei relative to the vector are determined. Average distance of the furthest decile of nuclei is calculated for both hand sides, and the width at each bin is the sum of these averages.

### Automatic border detection

Before running the algorithm, the nuclei area distribution is determined, and only polyploid nuclei are included into the analysis. The borders are detected based on normalized values of (i) polyploid nuclei distance to its nearest neighbor, (ii) midgut width, (iii) midgut width bin-to-bin difference, and (iv) polyploid nuclei area bin-to-bin difference (default setting variables). These variables have region specific variation along the midgut’s A/P-axis, and local changes correspond to the major region borders. In order to find local changes of the variables, a fitted fifth degree Chebyshev polynomial is subtracted from the values as a context adjustment. To this end, for each bin ( $x$ ) in the full range of bins  $[0 \dots a]$ , a total score is calculated by summing the weighted ( $w$ ) deviations of each variable’s ( $v_{1\dots n}$ ) normalized value from the fitted curve ( $c$ ):

$$\{x \in \mathbb{N} | 0 \leq x \leq a\}, x_{score} = \sum_{i=1}^n (v_i^x - c^x) \cdot w_i$$

The resulting score arrays are then smoothed and rescaled to interval  $[0, 1]$ . Peak detection is then performed on context-adjusted group average scores to find signals corresponding to region borders. To increase resolution, the border detection algorithm is run by twice the number of bins set by the user.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Statistics in LAM

LAM includes pairwise statistical testing of control and sample groups (Figure 3C). LAM has two types of in-built statistical testing. Firstly, bin values of the sample group are tested against the respective bin of the control group resulting in a representation of  $p$  values along the A/P axis of the midgut. Secondly, total feature counts of a sample group are tested against the control group. Both tests are performed with Mann-Whitney-Wilcoxon U test using continuity correction. In the bin-by-bin testing, false discovery rate correction due to multiple testing is applied. Additionally, for the bin-by-bin testing, a sliding window option of user-defined size is available. The use of a sliding window has some advantages depending on input data. For example, some cell types of the midgut may be spatially too sparse for bin-to-bin testing as the cell count at each bin would be skewed towards zero. Consequently, using a sliding window to merge bins would increase the number of non-zero values in the test population, and therefore increase the strength of the statistical test.

### Other statistical analysis

Statistical analyses were performed in R/Bioconductor. For parametric data, two-sample t-test or two-way ANOVA in conjunction with Tukey’s HSD test was used. For the non-parametric count data Wilcoxon rank-sum test with multiple testing correction ( $FDR < 0.05$ ) was used.

## ADDITIONAL RESOURCES

LAM tutorial videos: <https://www.youtube.com/playlist?list=PLjv-8Gzxh3AynUtI3HaahU2oddMbDpgtx>.