From heterogenous morphogenetic fields to homogeneous regions as a step towards understanding complex tissue dynamics

Satoshi Yamashita*[1], Boris Guirao[2], and François Graner*[1]

[1]Laboratoire Matière et Systèmes Complexes (CNRS UMR7057), Université Paris-Diderot, Paris, France

[2]Institut Curie, PSL Research University, CNRS UMR 3215, INSERM U934, F-75248 Paris Cedex 05, France

July 8, 2019

*Correspondence should be addressed to S.Y. (satoshiy83@gmail.com) and F.G. (francois.graner@univ-paris-diderot.fr)

1 Summary statement

Tissue morphogenesis is driven by multiple mechanisms. This study proposes a methodology to identify regions in the developing tissue, where each of the regions has distinctive cellular dynamics and deformation.

Abstract

Within developing tissues, cell proliferation, cell motility, and other cell behaviors vary spatially, and this variability gives a complexity to the morphogenesis. Recently, novel formalisms have been developed to quantify tissue deformation and underlying cellular processes. A major challenge for the study of morphogenesis now is to objectively define tissue sub-regions exhibiting different dynamics. Here we propose a method to automatically divide a tissue into regions where the local deformation rate is homogeneous. This was achieved by several approaches including image segmentation, clustering, and cellular Potts model simulation. We illustrate the use of the pipeline using a large dataset obtained during the metamorphosis of the *Drosophila* pupal notum. We also adapt it to determine regions where the time evolution of the local deformation rate is homogeneous. Finally, we generalize its use to find homogeneous regions for the cellular processes such as cell division, cell rearrangement, or cell size and shape changes. This pipeline will contribute substantially to the analysis of complex tissue shaping and the biochemical and bio-mechanical regulations driving tissue morphogenesis.

2 Introduction

During tissue development, morphogenesis is accompanied by cellular processes such as cell division, cell rearrangement, apical constriction, and apoptosis. The cellular processes are coordinated together, yielding collective cell migration and local deformation of each tissue region, resulting in convergent extension or epithelial folding. Furthermore, the local deformations of different tissue regions are coordinated too, resulting in large scale tissue morphogenesis. Coordination between invaginating mesoderm and covering ectoderm [Rauzi et al., 2015, Perez-Mockus et al., 2017], between invaginating

midgut and elongating germ-band [Collinet et al., 2015, Lye et al., 2015, Dicko et al., 2017] of *Drosophila* embryo, between contracting wing hinge and expanding wing blade in *Drosophila* pupa [Etournay et al., 2015, Ray et al., 2015], or between invaginating neural plate and covering epidermal ectoderm of *Xenopus* embryo [Brodland et al., 2010], provide examples of how mechanical force generated in one region can drive large scale deformation in adjacent regions. In these cases, the regions which behave differently are easily distinguished by specific gene expressions.

However, many tissues were found to be heterogeneous but without obvious boundary between such regions, leaving analysis limited to arbitrary regions drawn as a grid parallel to tissue axes (Fig. 1A), or regions expressing already known differentiation maker genes. Measured tissue deformation rate showed a large heterogeneity (accompanied by a heterogeneity in cellular processes such as cell proliferation rate, cell division, cell rearrangement, change of cell shape), and smooth spatial variations across the tissue, in *Drosophila* notum in a developing pupa [Bosveld et al., 2012, Guirao et al., 2015] (Fig. 1B), *Drosophila* wing blade [Etournay et al., 2015], blastopore lip of *Xenopus* gastrula [Feroze et al., 2015], chick gastrula [Rozbicki et al., 2015, Firmino et al., 2016], mouse palatal epithelia [Economou et al., 2013], and mouse limb bud ectoderm [Lau et al., 2015]. Recent formalisms have enabled to measure and link quantitatively cellular processes with tissue deformation [Blanchard et al., 2009, Guirao et al., 2015, Etournay et al., 2015, Merkel et al., 2017]. However, cellular quantities also vary smoothly across the tissue. In addition, the causal relationship between cellular processes and tissue deformation is not always trivial, making it difficult to identify regions those actively drive morphogenesis and those passively deformed by adjacent regions.

To study the spatial regulation of morphogenesis at tissue scale, we developed a new multi-technique pipeline to divide a tissue into sub-regions based on quantitative measurements of static or dynamic properties of cells or tissues. Our tissue segmentation pipeline consists of three steps: a first fast tissue segmentation attempted several times with random seeding, then merging these multiple tissue segmentations into a single one, then smoothing the resulting regions boundaries (Fig. 1C). We apply it to the

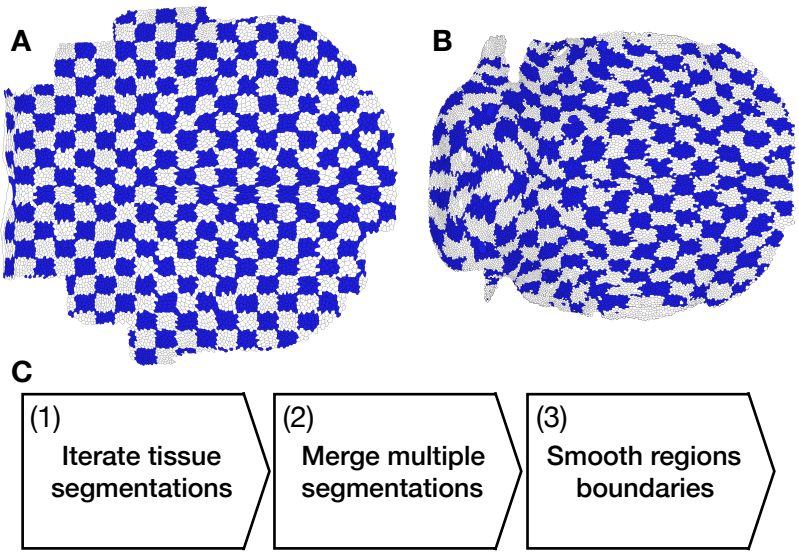

Fig. 1: Morphogenesis of *Drosophila* pupa notum and overview of tissue segmentation pipeline. (**A**, **B**) Heterogeneity of tissue morphogenesis. Here a *Drosophila* notum at 12 hr after pupa formation (APF), with arbitrary regions drawn from a grid (**A**), and at 32 hr APF, showing the heterogeneous deformation of previous regions using cell tracking (**B**). Cell patches are shown with blue and white check pattern. (**C**) Pipeline of the tissue segmentation. (1) Iteration of fast tissue segmentation with random seeding, using region growing algorithm. (2) Merging multiple tissue segmentations of step 1 into a single objective tissue segmentation, using label propagation algorithm on a consensus matrix. (3) Smoothing regions boundaries resulting of step 2, using cellular Potts model.

morphogenesis of *Drosophila* pupa dorsal thorax. The first application is to the tissue deformation rate integrated over the duration of a whole movie; the second application is to the time evolution of this tissue deformation rate; and the third application is to the time evolution of all cellular processes and their contribution to the local tissue deformation. Obtained sub-regions showed distinctive patterns of deformation and cellular processes with higher homogeneity than those along tissue axes. Interestingly, the tissue segmentations based on the local tissue deformation rate and on the cellular processes included some similar regions, suggesting that the cellular processes were regulated similarly inside the regions, therefore resulting in homogeneous tissue deformations inside those regions.

3 Results I : Development of automatic tissue segmentation algorithm

3.1 Image segmentation by region growing algorithm

Finding distinctive and homogeneous regions inside the heterogeneous tissue amounts to segmenting the geometrical space while keeping the points inside each of the regions as similar as possible to each other in the property space. Here, we call *property space* any morphogenesis quantification measured in the tissue, whereas *geometrical space* refers to the two-dimensional space of cell patch positions inside the tissue.

Given a set of objects, collecting similar objects to divide them into groups is generally a task of cluster analysis. However, the cell patches distribute in the property space and geometrical space. On the assumption that expression patterns of genes responsible for morphogenesis make connected regions, and to study physical interactions between the regions, we aimed at getting connected regions. The initial tissue segmentation first defines a metric of similarity between cells, and then a tissue is divided into regions containing similar cells. The image segmentation tool, called region growing [Adams and Bischof, 1994, Ma et al., 2010] (Fig. 2A), was inspired by a study segmenting mouse heart based on cell polarity [Le Garrec et al., 2013].

To validate the algorithm, we first tested segmentation on a simple example, namely the change in cell patch areas from 12 to 32 hr APF (Fig. 2B). The overall change in cell patch areas defines the total tissue growth, while spatially heterogeneous changes in cell patch areas result in local deformation, changes in tissue region proportions, and overall tissue shape change. Technically speaking, the change in cell patch areas is a scalar field, defined as the trace of the tissue deformation rate tensor. The region growing succeeded in finding expanding regions in posterior, lateral posterior, and lateral parts and a shrinking region in anterior part.

However, the results varied dependent on the initial seeds. In contrast to a segmentation in histology or of immuno-stained image, where a true segmentation is well defined, the morphogenetic properties vary continuously with space, making it difficult to determine and validate the resultant segmentations. The silhouette, a measurement of region homogeneity (the silhouette of an object would be 1 if it was similar to all objects in the same cluster, and -1 if it was more similar to objects in other clusters), differed from one segmentation to the other (Fig. 2C). To assess the significancy of the homogeneity, we compared it with the average silhouette of randomly made control segmentations. Some of the region growing results had a low silhouette, even lower than that of half of the control segmentations (Fig. 2C), which means they were lacking any signification.

A tissue segmentation method should return a unique result so that it could be compared with gene expression patterns or fed forward to a study of mechanical interactions between regions, but the region growing algorithm returned multiple results.

3.2 Defining a single tissue segmentation using label propagation on a consensus matrix

To obtain a single tissue segmentation, we turned to consensus clusterings. In fact, since resultant segmentations of the region growing were dependent on randomly given initial values, we ran multiple trials and merged multiple segmentation results into a single one. Given multiple partitions, the consensus clustering returns the partition which is the most similar to all of the initial partitions. We tried several consensus clustering

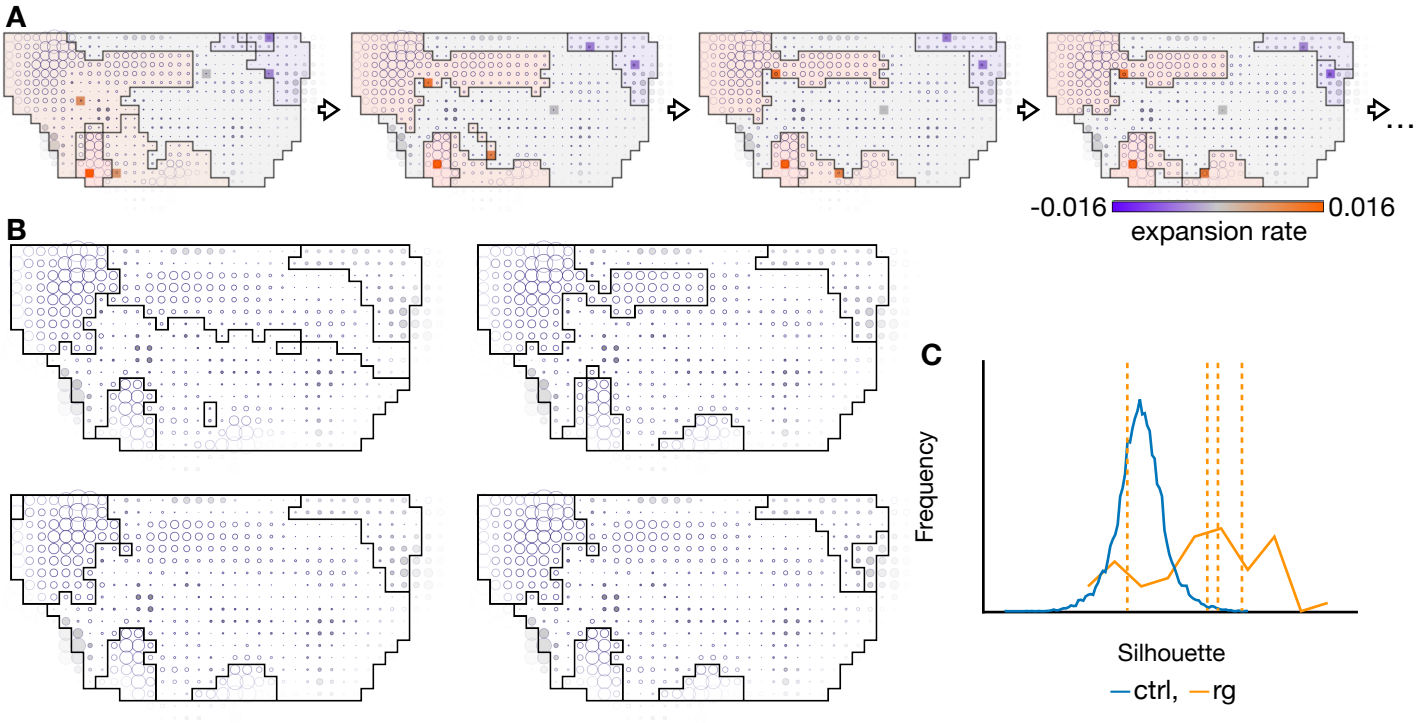

Fig. 2: Tissue segmentation by region growing algorithm. Tissue local expansion/contraction rate is represented by size of white/gray circles. (**A**) Process of region growing algorithm. Points of given number (6 in the shown example) are chosen randomly as initial seeds of regions, and the regions are expanded by collecting points similar to the seeds from their neighbors. Once the field is segmented, the seeds are updated to region centroids in the geometrical space and means in the property space, and the expansion of the regions are performed again from the new seeds. The seeds are shown with colored square, where the color represents an expansion rate of the regions. The regions are colored lighter for visibility. This update of the seeds and the regions are iterated until it reaches a convergence. (**B**) Four example results of region growing. (**C**) Histogram of silhouette value: blue for control segmentations, orange for region growing. Dotted vertical orange lines show silhouette values of the four examples shown in **B**.

algorithms, and found the *label propagation on a consensus matrix* [Lancichinetti and Fortunato, 2012, Raghavan et al., 2007] returning the best segmentation, while others

returned many disconnected regions or single point regions (data not shown).

The label propagation on a consensus matrix converted multiple tissue segmentations into a weighted graph where weight of an edge represented a frequency of segmentations in which incident vertices (points) belonged to the same region (Fig. 3A). Then labels on the vertices were propagated according to the weight so that the same label was assigned to points which were frequently included in the same region among the given multiple region growing segmentations.

The label propagation returned results similar to the region growing segmentations (Fig. 2B, 3B). Also, the label propagation results were more similar to each other than results of region growing, assessed with adjusted Rand indices (ARI), a measurement of similarity between two partitions (ARI of identical partitions would be 1). ARI were 0.50 ± 0.21 among the results of the region growing and 0.97 ± 0.02 among the results of the label propagation. They showed similar average silhouette values, similar to median of those of region growing results, but smaller than the highest value of those of region growing (Fig. 3C). The average silhouette of the label propagation result was higher than those of 99.95% of the randomly made control segmentations.

However, the label propagation algorithm does not assure the connectedness of regions (Fig. 3C lower panels). Also, contour of regions may be complex (data not shown).

3.3 Smoothing of tissue segmentation results by cellular Potts model

The consensus region boundaries were then smoothed by cellular Potts model, which balances the decrease in the length of region boundaries with the preservation of region area and homogeneity (Fig. 4A). It smoothed boundaries and removed disconnected cell patches (Fig. 4B, C) while keeping the average silhouette higher than those of random segmentations (Fig. 4D). Since the cellular Potts model simulation includes the Metropolis update, i.e., choosing a pixel randomly and updating the pixel by probability according to a change of the energy, resultant smoothed segmentations varied among different trials even with the same parameters and initial segmentation (data not shown).

Now we have a pipeline of the region growing, the label propagation, and the cellular

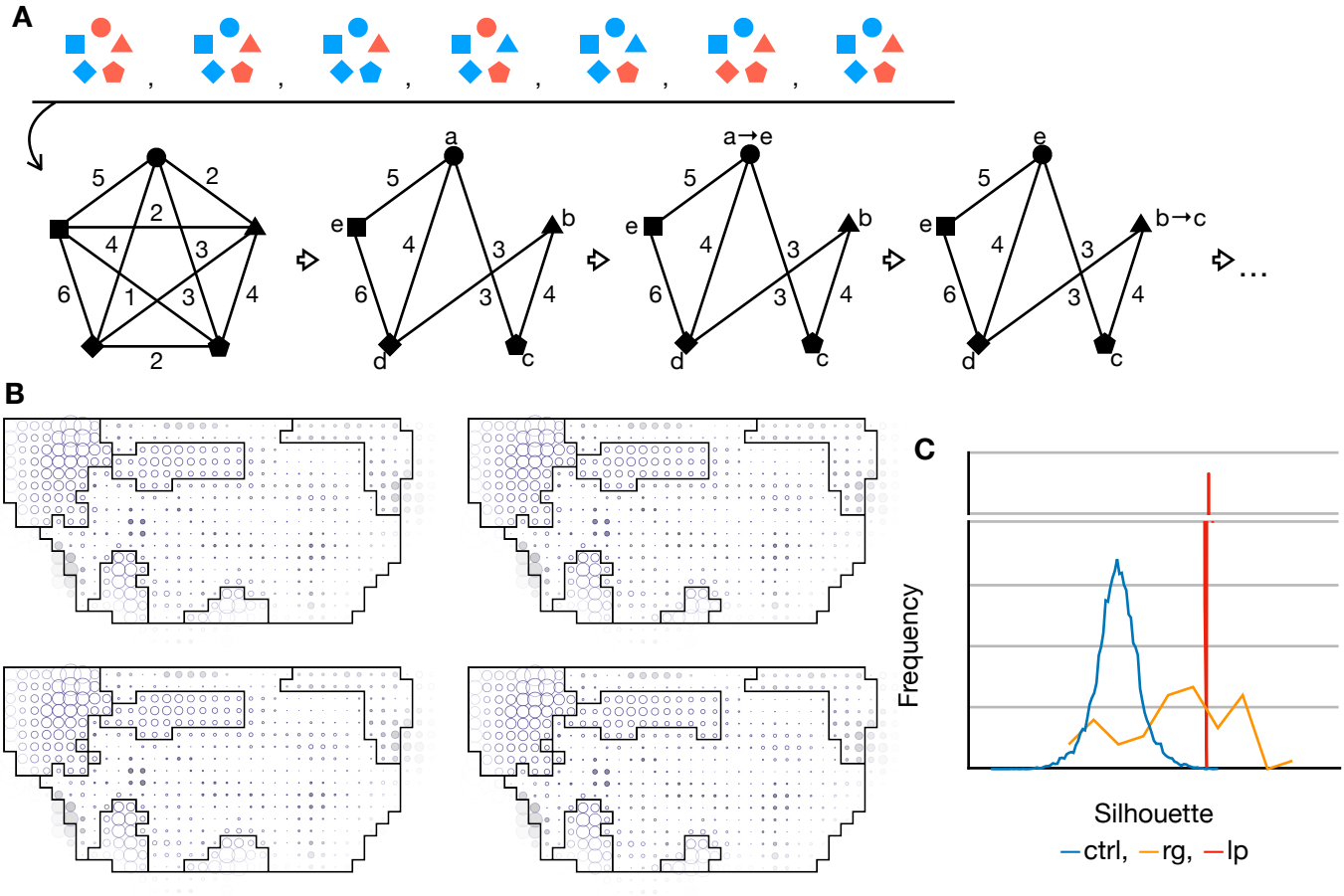

Fig. 3: Tissue segmentation by label propagation on a consensus matrix. (**A**) Process of label propagation algorithm. Multiple clusterings (upper) are converted to a consensus matrix, which gives weights to a complete graph on the objects being clustered (lower left). Edges with weights less than a given threshold are removed. All objects are initially assigned labels different to each other. And then, one by one in random order, each label is updated to the most frequent one weighted by edges incident to the object until it reaches a convergence. (**B**) Four example results of label propagation on the same consensus matrix. (**C**) Histogram of silhouette value: blue for control segmentations, orange for region growing, red for label propagation.

Potts model to divide a field of property (scalar, tensor, or any kind of value with metric) into regions. The resultant regions are homogeneous, where points in each region are

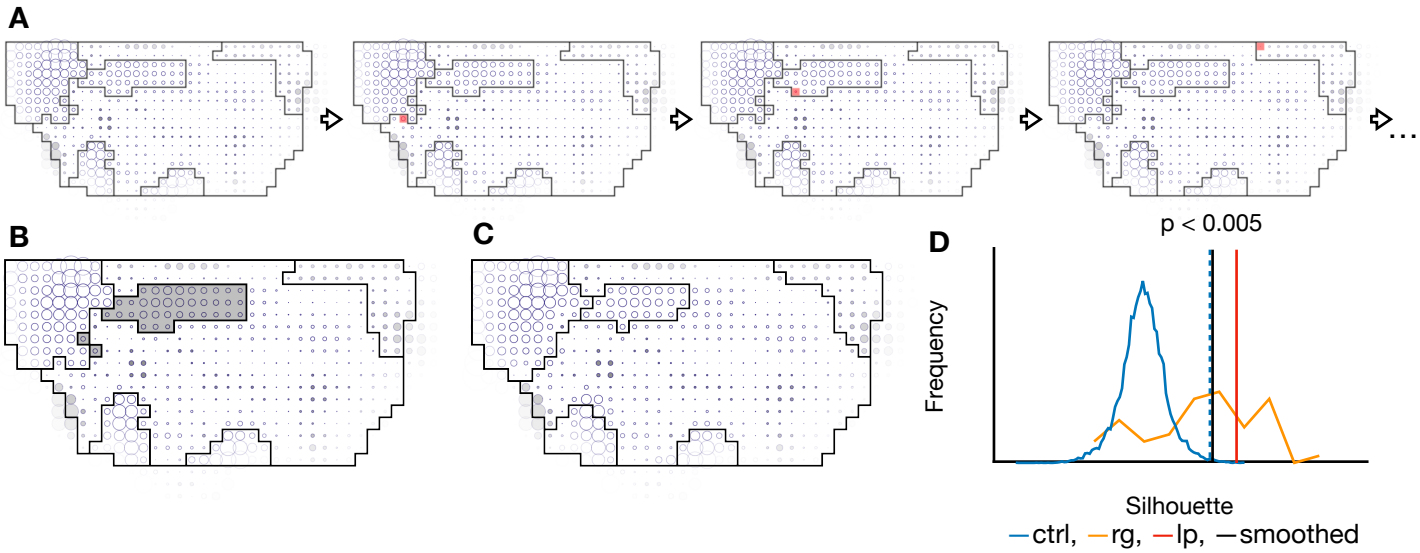

Fig. 4: Boundary smoothing by cellular Potts model. (**A**) Process of cellular Potts model. A pixel is randomly chosen and changes its belonging region if it decreases boundary length and/or increases homogeneity (marked by red). (**B**) Result of label propagation with a disconnected region shown by gray color. (**C**) Result of boundary smoothing by cellular Potts model. (**D**) Histogram of silhouette value: blue for control segmentations, orange for region growing, red vertical line for label propagation, and black vertical line for regions smoothed by cellular Potts model. Dotted blue line shows threshold for the highest 0.5% of the control segmentations.

more similar to each other than to points in other regions.

4 Results II : Tissue segmentation based on tissue morphogenesis

We now turn to property spaces better representing tissue morphogenesis. In Guirao et al. [2015], tissue deformation rate (**G**) and underlying cellular processes, cell division (**D**), cell rearrangement (**R**), cell shape change (**S**), and cell delamination (**A**) were quantified into tensors. The tensors were obtained from change of the texture averaged

over 20 hr from 12 hr APF to 32 hr APF or over 2 hr at each time point. By projecting each tensor on the deformation rate, one can measure an amplitude of deformation rate and an *effective contribution* of cellular process, which represents how much the cellular process contributes to the tissue deformation. They are scalar value and denoted by $\mathbf{G}_{/\!/}$ for the deformation rate, $\mathbf{D}_{/\!/}$ for cell division, $\mathbf{R}_{/\!/}$ for cell rearrangement, $\mathbf{S}_{/\!/}$ for cell shape change, and $\mathbf{A}_{/\!/}$ for cell delamination (see appendix). For the sake of clarity, we call the tissue deformation rate and the cellular processes averaged over the whole 20 hr from 12 to 32 hr APF *time-average* tissue deformation rate and cellular processes.

The effective contributions averaged over the whole tissue showed dynamic time evolution (Fig. S1), with a large peak of cell division and cell shape change around 16 hr APF, second small wave of cell division around 22 hr APF, and gradual increase of cell shape change and cell rearrangement. The effective contributions also showed large variance across the tissue at each time point.

4.1 Tissue segmentations based on time-evolution of deformation rate and cellular processes effective contributions

We first divided the tissue based on time-average and time-evolution of tissue deformation rate. A distance between the tensor-valued functions was given by L^2-norm. The notum was divided into anterior-middle-posterior and medial-lateral regions (Fig. 5A-D).

Next, we divided the tissue based on time-evolution of cellular processes. The effective contributions of the cellular processes were combined in a vector. For the time-evolution of the cellular processes, a distance between the vector-valued function was also given by L^2-norm. In contrast to segmentations based on the time-average and time-evolution of deformation rate, the segmentations based on time-average and time-evolution of the cellular processes effective contribution vectors differed from each other (Fig. 5E-H). The resultant segmentation based on time-evolution included a posterior region, a large middle region, a neck-notum boundary region, lateral posterior region, a scutum-scutellum boundary region, and a lateral region (Fig. 5H).

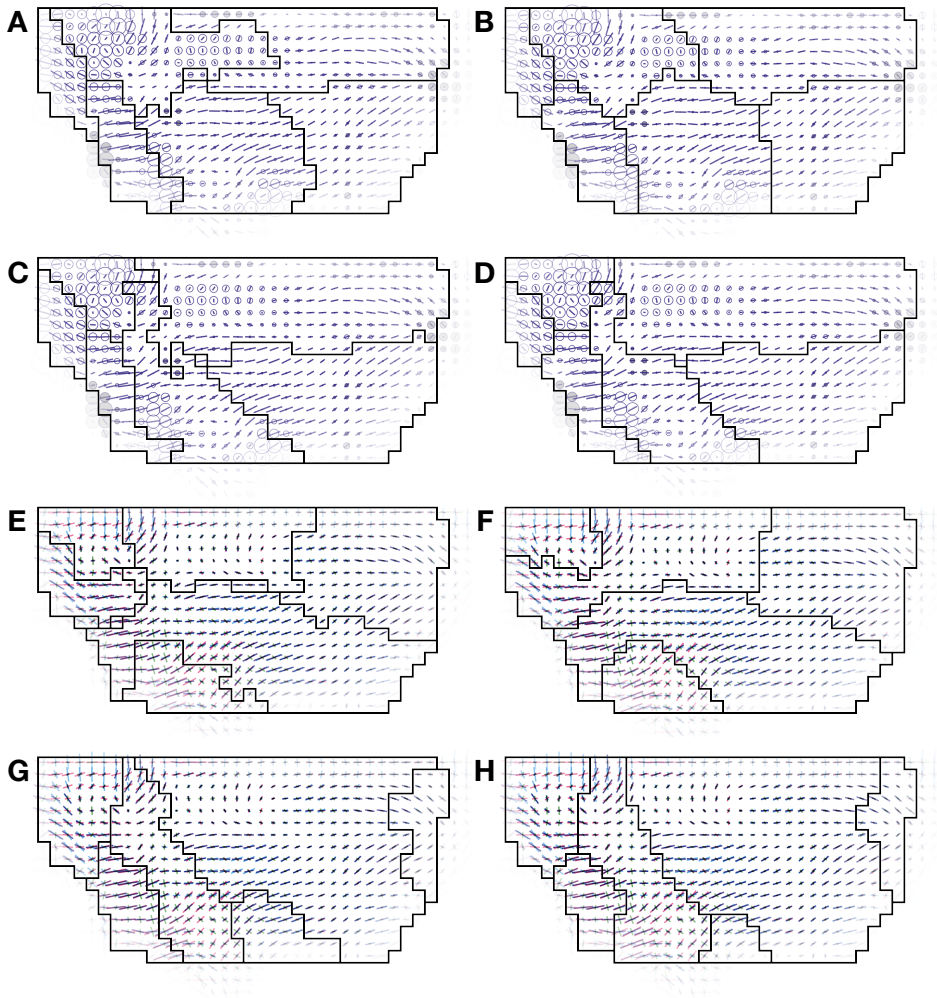

Fig. 5: Segmentations based on tissue deformation and underlying cellular processes. Direction of elongation is represented by bars. For quantification and representation of tissue deformation rate and cellular processes, see methodology and Guirao et al. [2015]. (**A**-**H**) Segmentations based on time-average tissue deformation rate (**A**, **B**), time-evolution of deformation rate (**C**, **D**), time-average cellular processes effective contributions (**E**. **F**), and time-evolution of cellular processes (**G**, **H**). First column shows results of label propagation (**A**, **C**, **E**, **G**) and second column shows results of boundary smoothing (**B**, **D**, **F**, **H**).

4.2 Correspondence between segmentations based on cellular processes and deformation rate

Both of the segmentations based on time-evolution of deformation rate and cellular processes effective contributions included the large anterior region, the middle boundary region, the lateral posterior region, and the posterior region, although the anterior and posterior regions were divided into medial and lateral subregions in the segmentation based on the deformation rate. Figure 6A-D show overlap between segmentation based on time-evolution of deformation rate (Fig. 5D) and the others (Fig. 5B, F, H) or a conventional large grid parallel to tissue axes. Despite the overlap of the middle boundary regions and the lateral posterior regions, ARI between the segmentations based on time-evolution of deformation rate and cellular processes was not high (0.391) because of the medial and lateral subregions in the anterior and posterior regions. The ARI value was similar to those between others (Fig. 6E).

4.3 Homogeneity of the obtained regions

Next, we evaluated the homogeneity of the obtained regions. The time-evolution of deformation rate was similar among cells inside regions of the segmentations based on time-average and time-evolution of deformation rate except the middle-lateral region of the former (Fig. 7A, B). On the other hand, the large grid segmentation showed large heterogeneity in the posterior regions (Fig: 7C). The average silhouette value of the segmentation based on the time-evolution of deformation rate was higher than those of 99.5% of the control segmentations (average silhouette for label propagation: 0.0568, for smoothed regions: 0.0527, the maximum of the smallest 99.5% of controls average silhouettes: 0.0425) (Fig. 7D). The average silhouette of the segmentation based on time-averaged deformation rate was also higher than 95% of the control segmentations (average silhouette for label propagation: 0.0336, for smoothed regions: 0.0107, the maximum of the smallest 95% of controls average silhouettes: -0.0098). On the other hand, that of the conventional grid segmentation was close to median of the control segmentations (average silhouette: -0.0694).

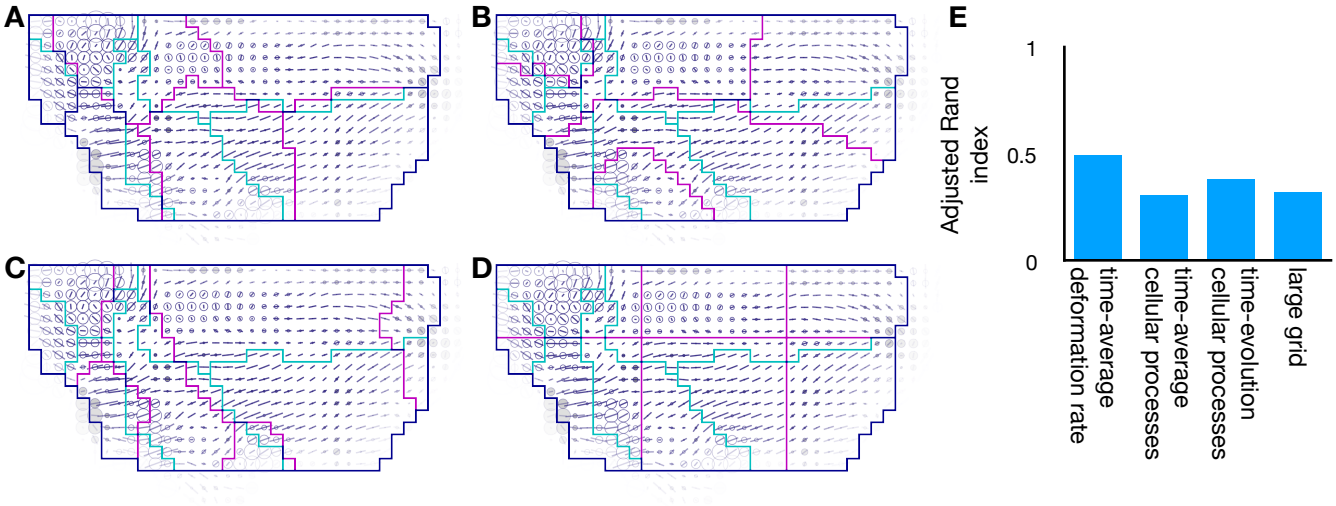

Fig. 6: Correspondence between segmentations based on cellular processes and deformation rate. (**A**-**D**) Overlays of segmentations, where segmentation based on time-evolution of deformation rate is shown by cyan line, while segmentations based on time-average deformation rate (**A**), time-average cellular processes (**B**), time-evolution of cellular processes (**C**), and large grid (**D**) are shown by magenta line. (**E**) Adjusted Rand indices of **A**-**D**.

Also, the time-evolution of cellular processes was homogeneous inside the regions of the segmentation based on time-evolution of cellular processes, but not in segmentation based on time-average of cellular processes nor in the grid (Fig. 7E-G). The average silhouette value of segmentation based on time-evolution was higher than 99.995% of control segmentations (average silhouette for label propagation: 0.147, for smoothed regions: 0.152, the maximum of the smallest 99.995% of controls average silhouettes: 0.124), while that of segmentation based on time-average was smaller than 5% of control segmentations (average silhouette for label propagation: 0.012, for smoothed regions: 0.015, the maximum of the smallest 99.995% of controls average silhouettes: 0.024) (Fig. 7H).

Our tissue segmentation is designed to divide a tissue into regions homogeneous in a given property space, and the homogeneity of deformation rate in the segmentation

based on the deformation rate demonstrated that the pipeline worked fine. Therefore, it does not assure the homogeneity of regions in a property space different from one based on which segmentation was performed. Figure S2 shows heat maps of silhouette values measured in different property spaces. Homogeneity in the regions differed among the property spaces. However, the large anterior region in the segmentation based on time-evolution of cellular processes was homogeneous in all property spaces. On the other hand, the posterior region in the segmentation based on time-evolution of cellular processes showed heterogeneity of deformation rate, while the two posterior regions in the segmentation based on deformation rate showed homogeneity of cellular processes.

4.4 Cellular processes effective contributions inside the regions

Figure 8 shows plots of cellular processes effective contributions averaged in each region of the segmentation based on time-evolution of deformation rate, where the medial and lateral anterior regions were merged. The anterior region did not have the second peak of cell divisions, while the first peak was small in the lateral posterior region. The two posterior regions had different cell divisions and cell shape changes after the second peak of cell division. The lateral regions showed cell shape changes between the two peaks of cell division and increasing cell rearrangements.

Distances between the plots in Figure 8 (0.66 ± 0.16) were close to those between plots in the segmentation based on time-evolution of cellular processes (0.67 ± 0.15) and larger than those between plots in the large grid (0.44 ± 0.14) (Fig. S3). This result demonstrates that cellular processes in the obtained segmentations were more distinctive than those in the conventional grid.

5 Discussion

This study demonstrates that the pipeline of the region growing, the label propagation on the consensus matrix, and the boundary smoothing by cellular Potts model could divide a deforming heterogeneous tissue into homogeneous regions based on the morphogenetic

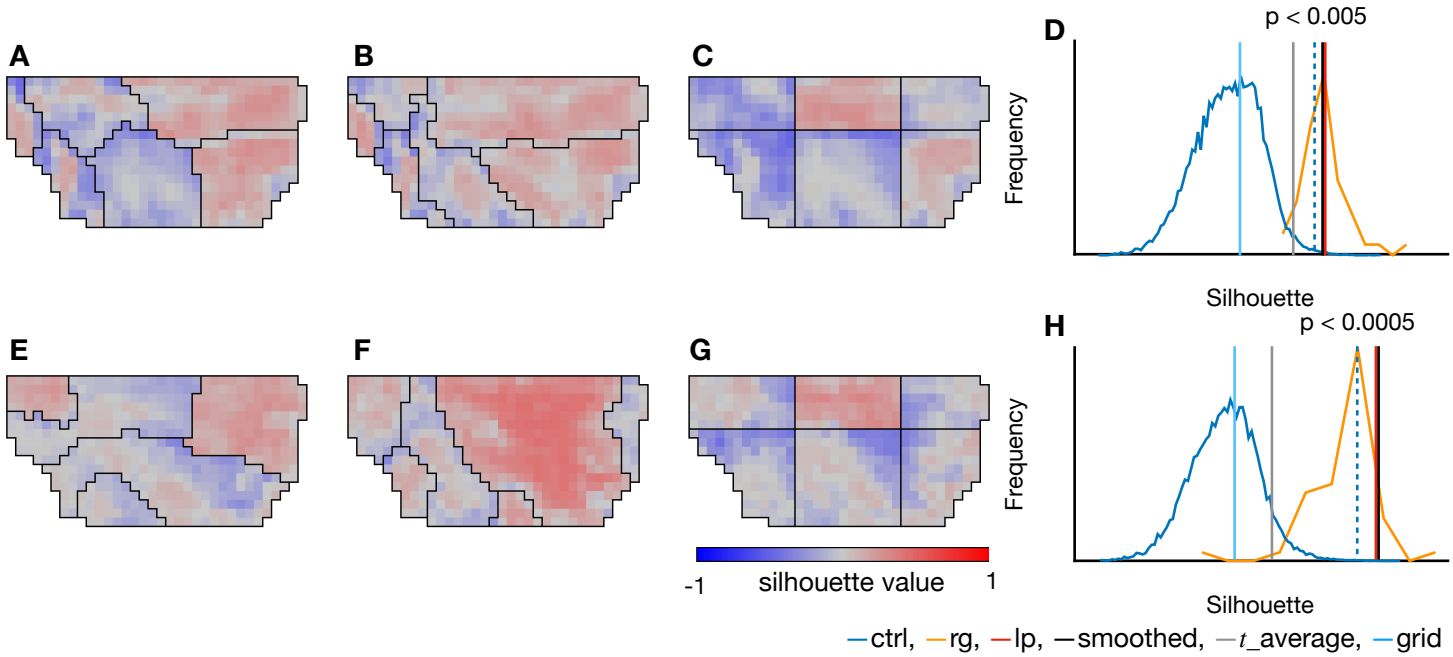

Fig. 7: Homogeneity in the obtained regions. (**A**-**C**) Heat map of silhouette value measured with time-evolution of tissue deformation rate in segmentations based on time-average deformation rate (**A**), time-evolution of deformation rate (**B**), and large grid (**C**). (**D**) Histogram of silhouette value, : blue for control segmentations, orange for region growing. Red vertical lines show silhouette value of label propagation results. Black vertical lines show silhouette value of regions smoothed by cellular Potts model. Dotted blue lines show threshold for the highest 0.5% of the control segmentations. Gray and cyan vertical line shows silhouette value of segmentation based on time-averaged deformation rate and large grid. (**E**-**G**) Heat map of silhouette value measured with time-evolution of cellular processes effective contributions in segmentations based on time-average cellular processes (**E**), time-evolution of cellular processes (**F**), and large grid (**G**). (**H**) Histogram of silhouette value, : blue for control segmentations, orange for region growing. Red vertical lines show silhouette value of label propagation results. Black vertical lines show silhouette value of regions smoothed by cellular Potts model. Dotted blue lines show threshold for the highest 0.05% of the control segmentations. Gray and cyan vertical line shows silhouette value of segmentation based on time-averaged cellular processes and large grid.

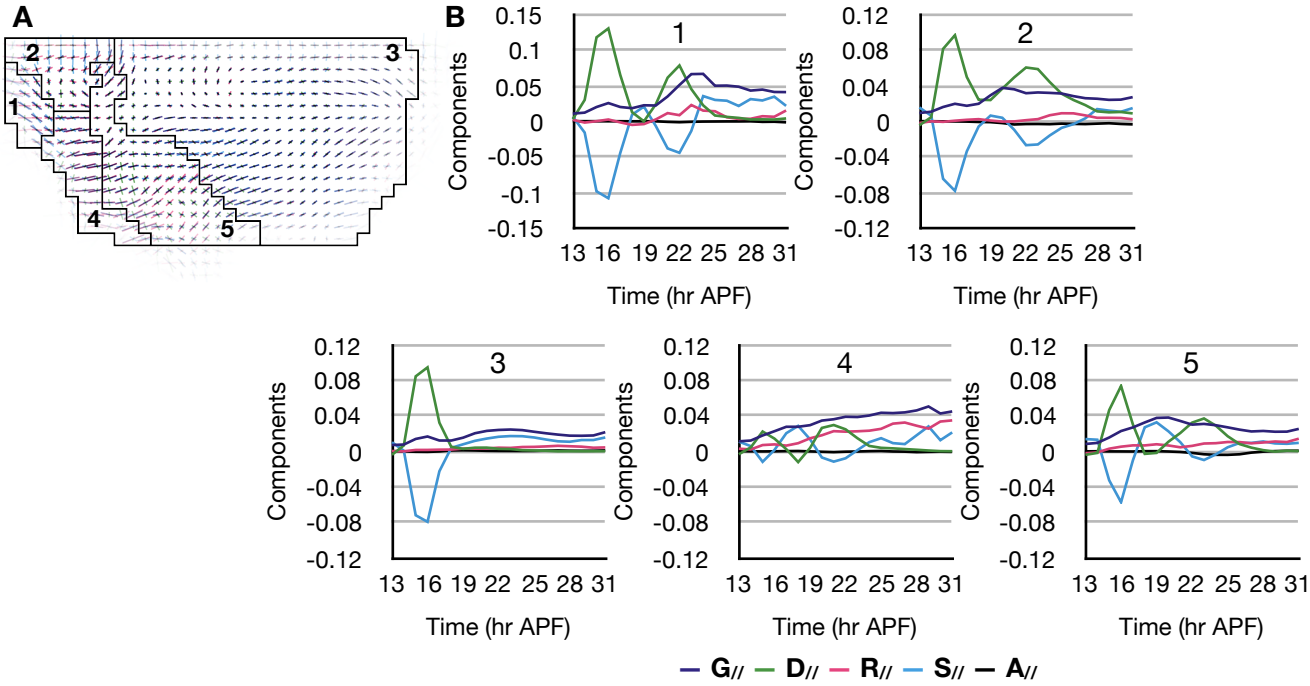

Fig. 8: Cellular processes effective contributions inside the regions. (**A**) Tissue segmentation based on time-evolution of deformation rate, where two anterior subregions were merged. (**B**) Plots of cellular processes effective contributions averaged in each region of 1-5 in **A**.

properties. Using this segmentation, we found a large anterior region, a middle boundary region, a lateral posterior region, and posterior regions that had distinctive deformation rates and cellular processes time-evolutions.

The tissue segmentation based on morphogenesis differs from conventional image segmentations, cell segmentations, and other tissue segmentations those are an automation of manual segmentation and can be corrected manually. First, the morphogenesis was quantified as multiple tensor fields with time evolution, and thus it is hard to visualize them in a 2D image for manual segmentation. Second, it is not easy to evaluate whether given a region actually corresponds to genetical/mechanical regulation of morphogenesis. Therefore we looked for a method which divides a tissue based on any kind of quantity and returns regions with smooth boundaries. Region growing is a conventional and simple method of image segmentation, and requires a property space only to be metric. The varying results of the region growing were given to the label propagation and cellular Potts model to produce a single tissue segmentation, and the result was evaluated by homogeneity of the regions and smoothness of the boundary.

The tissue segmentations based on time evolution of deformation rate and effective contributions of cellular processes returned similar results with large anterior regions, posterior regions, lateral posterior regions, and middle boundary regions (Fig. 6). By the deformation rate, the anterior and posterior regions were divided into medial and lateral subregions. Since the vector of effective contributions ignores the direction of deformation, the two subregions could be interpreted as regions of similar underlying cellular processes but deforming in different directions. On the other hand, the middle boundary region and the lateral region given by the cellular processes, both overlapped with the middle boundary region given by deformation rate, could be interpreted as regions with different cellular processes but of similar deformations.

Silhouette analysis showed that the segmentations based on time-evolution of deformation rate and cellular processes included regions homogeneous in various property spaces, whereas the conventional grid segmentation included heterogeneous regions. Especially, the large anterior region showed high homogeneity of deformation rate and the

two posterior subregions showed high homogeneity of cellular processes.

In conclusion, we built a method to divide a tissue based on any kind of property space. This allows an application to a study of spatial regulation of various processes, where the property space should be chosen for the process of interest. For example, to study a spatial regulation of cell division orientation, the property space may be prepared from, instead of the local deformation rate and cellular processes, the tensor field of cell division and known regulating factors such as cell shape, localization of planar cell polarity proteins, and tension on cell-cell interface, and then resultant regions can be compared with genes expression patterns.

6 Methods

6.1 Quantification tools

Morphogenesis data result from the quantification of local tissue deformation rate and underlying cellular processes as described in [Guirao et al., 2015]. The similarity between two tensors is quantified by the standard Euclidian metric. The homogeneity of a quantity within a given region, i.e. the similarity between measurements of this quantity within a region, is measured by silhouette, a standard tool of cluster analysis. For a measurement of similarity between tissue segmentations, we use the Rand index, which indicates how well two data clusterings agree.

6.1.1 Quantification of tissue deformation and cellular processes

Quantification of local tissue deformation and underlying cellular processes was performed in [Guirao et al., 2015]. Briefly, *Drosophila* nota expressing GFP-tagged E-cadherin were imaged. The notum movies were split in a grid (with patches about 20 μm width) at the first frame (Fig. 1C), 12 hr after pupa formation (APF). The local deformation rate and the cellular processes were measured in each cell patch through the development, as follows.

Epithelial cells contours were detected automatically using watershed algorithm, the

cells were tracked, adjacencies between cells were listed, and relative positions of centers of adjacent cells were recorded. The tissue deformation rate, denoted by the symmetric tensor $\mathbf{G}$, was obtained from change of relative positions between neighbor cells over 20 hr from 12 hr APF to 32 hr APF, or over 2 hr at each time point when recording the time evolution. The tissue deformation rate $\mathbf{G}$ was then decomposed into cell shape change $\mathbf{S}$ and deformation accompanied by change of cell adjacency, which was further divided into cell division $\mathbf{D}$, cell rearrangement $\mathbf{R}$, and cell delamination $\mathbf{A}$, which are symmetric tensors too.

In a collection of cells where the total deformation is driven completely by the four fundamental cellular processes, the tensors are in a balance equation,

$$\mathbf{G} = \mathbf{D} + \mathbf{R} + \mathbf{S} + \mathbf{A}. \tag{1}$$

The scalar product of two tensors $\mathbf{Q}$ and $\mathbf{Q}'$ in dimension d is defined as:

$$\mathbf{Q}.\mathbf{Q}' = \frac{1}{d}\mathrm{Tr}(\mathbf{Q}\mathbf{Q}'^T), \tag{2}$$

and the unitary tensor $\mathbf{u_G}$ that is aligned with $\mathbf{G}$ is given by

$$\mathbf{u_G} = \frac{\mathbf{G}}{(\mathbf{G}.\mathbf{G})^{1/2}}. \tag{3}$$

Since the scalar product (Eqn. 2) is a linear transformation, multiplying $\mathbf{u_G}$ by a tensor, the operation $.\mathbf{u_G} : \mathbf{Q} \rightarrow \mathbf{Q}_{/\!/}$, retains the balance between the tissue deformation rate and the cellular processes in Equation 1 while converting them to magnitudes:

$$\begin{aligned}\mathbf{G}_{/\!/} =& \mathbf{G}.\mathbf{u_G} \\ =& (\mathbf{D} + \mathbf{R} + \mathbf{S} + \mathbf{A}).\mathbf{u_G} \\ =& \mathbf{D}.\mathbf{u_G} + \mathbf{R}.\mathbf{u_G} + \mathbf{S}.\mathbf{u_G} + \mathbf{A}.\mathbf{u_G} \\ =& \mathbf{D}_{/\!/} + \mathbf{R}_{/\!/} + \mathbf{S}_{/\!/} + \mathbf{A}_{/\!/}.\end{aligned} \tag{4}$$

The scalar $\mathbf{G}_{/\!/}$ represents the local magnitude of tissue morphogenesis, and $\mathbf{D}_{/\!/}$, $\mathbf{R}_{/\!/}$, $\mathbf{S}_{/\!/}$, and $\mathbf{A}_{/\!/}$ represent the effective contributions of the cellular processes to the tissue morphogenesis. When a cellular process produces an anisotropic deformation in the

same direction with that of tissue, e.g. cells divided in the same direction with tissue elongation, the scalar product between the them returns a positive value, while it returns negative value when a cellular process counteractes tissue deformation.

6.1.2 Metric

Similarity of morphogenesis between the cell patches was defined as follows.

For expansion/contraction of area (isotropic deformation), similarity was given by difference in expansion/contraction rates.

Similarity of anisotropic deformation was given by a distance between two tensors $\mathbf{Q}$ and $\mathbf{Q'}$,

$$d(\mathbf{Q}, \mathbf{Q'}) = \left\{ \left(\frac{\mathbf{Q}_{xx} - \mathbf{Q}_{yy}}{2} - \frac{\mathbf{Q'}_{xx} - \mathbf{Q'}_{yy}}{2} \right)^2 + (\mathbf{Q}_{xy} - \mathbf{Q'}_{xy})^2 \right\}^{1/2}. \tag{5}$$

For tensors with time-evolution $\mathbf{Q}(t)$ and $\mathbf{Q'}(t)$, distance was given by a sum of distance at each time point,

$$|\mathbf{Q} - \mathbf{Q'}| = \int d(\mathbf{Q}(t), \mathbf{Q'}(t))dt, \tag{6}$$

as an analogy to L^1-norm between functions.

For the composition of cellular processes, the tensors of cellular processes were converted to effective contributions and combined into a vector ($\mathbf{G}_{/\!/}, \mathbf{D}_{/\!/}$, $\mathbf{R}_{/\!/}$, $\mathbf{S}_{/\!/}$, $\mathbf{A}_{/\!/}$). A distance between two vectors was given by L^2-norm, and a distance between vector-valued functions $v(t)$ and $v'(t)$ was given by an integral of the L^2-norm between $v(t)$ and $v'(t)$,

$$|v - v'| = \int ||v(t) - v'(t)||dt. \tag{7}$$

6.1.3 Silhouette and bootstrap

Silhouette quantifies clustering results, indicating how well an object resembles other objects inside its own cluster [Rousseeuw, 1987]. For object p, the silhouette value of p is defined by

$$s(p) = \frac{b(p) - a(p)}{\max\{a(p), b(p)\}}, \tag{8}$$

where $a(p)$ is an average distance from p to other objects in the same cluster, and $b(p)$ is the smallest average distance from p to all objects in any other clusters. By this definition, $-1 \leq s(p) \leq 1$, where $s(p)$ large and close to 1 indicates that p is similar to other objects in the same cluster, while negative $s(p)$ indicates that there is another cluster whose objects are more similar to p than objects in the cluster containing p.

We took the average silhouette value over all points (cell patches) as a measurement of homogeneity of a given segmentation. For significance test, tissue was segmented randomly 20,000 times into a given number, and we got thresholds above which the highest 5%, 0.5%, or 0.005% of the average silhouettes were found. The average silhouette of given regions were compared to those of the control segmentations with the same number of regions.

6.1.4 Adjusted Rand index

For a measurement of similarity between tissue segmentations, we use the Rand index, which indicates how well two data clusterings agree ; its value is 1 if the clusterings entirely disagree and 1 if they entirely agree. Its corrected-for-chance version is a more meaningful quantity, called the adjusted Rand index (ARI): it is the Rand index minus its value expected for the random case, and its value can be negative.

We compute the adjusted Rand index with the permutation model [Hubert and Arabie, 1985]. Given two clusterings $A = \{A_1, \ldots, A_k\}$ and $B = \{B_1, \ldots, B_m\}$ of N elements, the contingency table $\tau = (n_{ij})_{k \times m}$ is made where $n_{ij} = |A_i \cap B_j|$. The Rand index between A and B, $\mathrm{RI}(A, B)$ is given by the function

$$\mathrm{RI}(A, B) = \frac{2\sum_{ij}\binom{n_{ij}}{2} - \sum_i \binom{a_i}{2} - \sum_j \binom{b_j}{2} + \binom{N}{2}}{\binom{N}{2}}, \tag{9}$$

where $a_i = \sum_j n_{ij}$ and $b_j = \sum_i n_{ij}$, and an expected Rand index $\mathbb{E}[\mathrm{RI}(A, B)]$ is given

by the function

$$\mathbb{E}[\mathrm{RI}(A,B)] = \frac{\sum_i \binom{a_i}{2}}{\binom{N}{2}} \frac{\sum_j \binom{b_j}{2}}{\binom{N}{2}} + \left(1 - \frac{\sum_i \binom{a_i}{2}}{\binom{N}{2}}\right)\left(1 - \frac{\sum_j \binom{b_j}{2}}{\binom{N}{2}}\right). \tag{10}$$

The adjusted Rand index $\mathrm{ARI}(A,B)$ is given by the function

$$\mathrm{ARI}(A,B) = \frac{\mathrm{RI}(A,B) - \mathbb{E}[\mathrm{RI}(A,B)]}{1 - \mathbb{E}[\mathrm{RI}(A,B)]}. \tag{11}$$

6.2 Tissue segmentation pipeline

The pipeline was implemented by custom Matlab scripts, in three steps (Fig. 1C).

6.2.1 Region growing tissue segmentation

The initial tissue segmentation first defines a metric of similarity between cells, and then a tissue is divided into regions containing similar cells. This approach was inspired by a study segmenting mouse heart based on cell polarity [Le Garrec et al., 2013]. Testing k-means clustering in the property space returned disconnected regions in the geometrical space (data not shown). On the assumption that expression patterns of genes responsible for morphogenesis make connected regions, and to study physical interactions between the regions, we aimed at getting connected regions.

The algorithm *Region growing* [Adams and Bischof, 1994, Ma et al., 2010] is an image segmentation method using a process similar to k-means clustering, starting from randomly given seeds (corresponding to "means" in k-means clustering), segmenting an image with the seeds followed by update of the seeds within the regions, and iterating this process until convergence (Fig. 2A). The tissue segmentation is done by growing regions from the seeds collecting pixels adjacent to the growing regions, and so the resultant regions are connected.

Initial seeds were randomly chosen from data, and regions were expanded by adding a pixel (cell patch) adjacent to a region and the most similar to the seed of the region

in the property space one by one until all pixels were assigned to one of the regions. The seeds were updated to pixels the closest to centroids of the regions, averages of the regions in the property space were given as property of the seeds, and then regions were expanded again from the seeds. These region expansions and seeds updates were iterated until convergence was reached.

6.2.2 Label propagation on a consensus matrix

To merge multiple segmentation results into a single one independent on the metric, we use label propagation algorithm on a consensus matrix, which takes multiple partitions and returns a consensus partition which is the most similar to all partitions [Lancichinetti and Fortunato, 2012, Raghavan et al., 2007].

For a division of n points, independent 50 trials of region growing were converted to a consensus matrix, whose entry at i-th row and j-th column indicates a frequency of partitions in which i-th point and j-th point were in the same cluster. The entries lower than a given threshold were set to 0. The label propagation started by assigning to each point a different label. Then the label of randomly chosen i-th point was updated to one that was the most weighted by the consensus matrix, where ij element gave the weight to a label of j-th point. The label update was iterated until convergence. The threshold for the consensus matrix was scanned between 20-80% so that a resultant partition contained the same number of regions as the initial partitions.

6.2.3 Cellular Potts model

To smooth the consensus region boundaries while preserving region area and homogeneity, we use the cellular Potts model, in which a cellular structure is numerically simulated in a square lattice, where each cell is a set of pixels ; the pattern is updated to decrease the energy, with some fluctuation allowance [Graner and Glazier, 1992]. In the simplest and common two-dimensional form, the energy $\mathcal{H}$ arises from total perimeter length P (with line energy J) and constraint on each region area A (with compressibility λ); decreasing it results in smoother region with preserved area A_0, removing small protrusions and

disconnected regions. In this study, we also included the silhouette s to account for the region homogeneity, with a weight coefficient h: $\mathcal{H} = \sum_{\text{regions}} \left[JP + \lambda(A - A_0)^2 - hs \right]$. The coefficients J, λ, and h were adjusted manually on a case by case basis.

When updating the label for a randomly selected pixel a, a target label was randomly selected from neighbors of a, and then change of the Hamiltonian was calculated and updated label of a to the target label with probability $min(1, e^{-\Delta\mathcal{H}/T})$, where $\Delta\mathcal{H}$ denotes change of $\mathcal{H}$ by the change of label of a, and T is the fluctuation allowance. Simulations were iterated 100 times, then the result was visually inspected to check whether boundaries were enough smooth, and if needed were iterated up to 400 times.

7 Acknowledgement

The authors thank Floris Bosveld for imaging, Stéphane U. Rigaud for cell segmentation and tracking, Yohanns Bellaïche for directing and advises, and all team members for insightful discussions.

8 Competing interests

The authors declare no competing nor financial interests.

9 Author contributions

Quantification of morphogenesis: B.G.; Methodology: S.Y., F.G.; Definition of metric: S.Y., F.G.; Programming and implementation: S.Y.; Writing and editing: S.Y., F.G.; Supervision: F.G.

10 Funding

This research was supported by grants from Uehara Memorial Foundation (201630059), Japan Society for the Promotion of Science (201860572).

References

R. Adams and L. Bischof. Seeded region growing. *IEEE Trans Pattern Anal Mach Intell*, 16:641–647, Jun 1994.

G. B. Blanchard, A. J. Kabla, N. L. Schultz, L. C. Butler, B. Sanson, N. Gorfinkiel, L. Mahadevan, and R. J. Adams. Tissue tectonics: morphogenetic strain rates, cell shape change and intercalation. *Nat Methods*, 6:458–464, 2009.

F. Bosveld, I. Bonnet, B. Guirao, S. Tlili, Z. Wang, A. Petitalot, R. Marchand, P.-L. Bardet, P. Marcq, F. Graner, and Y. Bellaïche. Mechanical control of morphogenesis by fat/dachsous/four-jointed planar cell polarity pathway. *Science*, 336:724–727, 2012.

G. W. Brodland, X. Chen, P. Lee, and M. Marsden. From genes to neural tube defects (ntds): insights from multiscale computational modeling. *HFSP J*, 4:142–152, 2010.

C. Collinet, M. Rauzi, P.-F. Lenne, and T. Lecuit. Local and tissue-scale forces drive oriented junction growth during tissue extension. *Nat Cell Biol*, 17:1247–1258, 2015.

M. Dicko, P. Saramito, G. B. Blanchard, C. M. Lye, B. Sanson, and J. Etienne. Geometry can provide long-range mechanical guidance for embryogenesis. *PLoS Comput. Biol.*, 13(3):e1005443, Mar 2017.

A. D. Economou, L. J. Brock, M. T. Cobourne, and J. B. Green. Whole population cell analysis of a landmark-rich mammalian epithelium reveals multiple elongation mechanisms. *Development*, 140(23):4740–4750, Dec 2013.

R. Etournay, M. Popović, M. Merkel, A. Nandi, C. Blasse, B. Aigouy, H. Brandl, G. Myers, G. Salbreux, F. Julicher, and S. Eaton. Interplay of cell dynamics and epithelial tension during morphogenesis of the Drosophila pupal wing. *Elife*, 4:e07090, Jun 2015.

R. Feroze, J. H. Shawky, M. von Dassow, and L. A. Davidson. Mechanics of blastopore closure during amphibian gastrulation. *Dev. Biol.*, 398(1):57–67, Feb 2015.

J. Firmino, D. Rocancourt, M. Saadaoui, C. Moreau, and J. Gros. Cell Division Drives Epithelial Cell Rearrangements during Gastrulation in Chick. *Dev. Cell*, 36(3):249–261, Feb 2016.

F. Graner and J. A. Glazier. Simulation of biological cell sorting using a two-dimensional extended Potts model. *Phys. Rev. Lett.*, 69(13):2013–2016, Sep 1992.

B. Guirao, S. U. Rigaud, F. Bosveld, A. Bailles, J. Lopez-Gay, S. Ishihara, K. Sugimura, F. Graner, and Y. Bellaïche. Unified quantitative characterization of epithelial tissue development. *Elife*, 4, Dec 2015.

L. Hubert and P. Arabie. Comparing partitions. *Journal of Classification*, 2:193–218, 1985.

A. Lancichinetti and S. Fortunato. Consensus clustering in complex networks. *Sci Rep*, 2:336, 2012.

K. Lau, H. Tao, H. Liu, J. Wen, K. Sturgeon, N. Sorfazlian, S. Lazic, J. T. A. Burrows, M. D. Wong, D. Li, S. Deimling, B. Ciruna, I. Scott, C. Simmons, R. M. Henkelman, T. Williams, A.-K. Hadjantonakis, R. Fernandez-Gonzalez, Y. Sun, and S. Hopyan. Anisotropic stress orients remodelling of mammalian limb bud ectoderm. *Nat Cell Biol*, 17:569–579, 2015.

J. F. Le Garrec, C. V. Ragni, S. Pop, A. Dufour, J. C. Olivo-Marin, M. E. Buckingham, and S. M. Meilhac. Quantitative analysis of polarity in 3D reveals local cell coordination in the embryonic mouse heart. *Development*, 140(2):395–404, Jan 2013.

C. M. Lye, G. B. Blanchard, H. W. Naylor, L. Muresan, J. Huisken, R. J. Adams, and B. Sanson. Mechanical Coupling between Endoderm Invagination and Axis Extension in Drosophila. *PLoS Biol.*, 13(11):e1002292, 2015.

Z. Ma, J. M. Tavares, R. N. Jorge, and T. Mascarenhas. A review of algorithms for medical image segmentation and their applications to the female pelvic cavity. *Comput Methods Biomech Biomed Engin*, 13(2):235–246, 2010.

M. Merkel, R. Etournay, M. Popović, G. Salbreux, S. Eaton, and F. Julicher. Triangles bridge the scales: Quantifying cellular contributions to tissue deformation. *Phys Rev E*, 95(3-1):032401, Mar 2017.

G. Perez-Mockus, K. Mazouni, V. Roca, G. Corradi, V. Conte, and F. Schweisguth. Spatial regulation of contractility by Neuralized and Bearded during furrow invagination in Drosophila. *Nat Commun*, 8(1):1594, Nov 2017.

U. N. Raghavan, R. Albert, and S. Kumara. Near linear time algorithm to detect community structures in large-scale networks. *Phys Rev E Stat Nonlin Soft Matter Phys*, 76(3 Pt 2):036106, Sep 2007.

M. Rauzi, U. Krzic, T. E. Saunders, M. Krajnc, P. Ziherl, L. Hufnagel, and M. Leptin. Embryo-scale tissue mechanics during Drosophila gastrulation movements. *Nat Commun*, 6:8677, 2015.

R. P. Ray, A. Matamoro-Vidal, P. S. Ribeiro, N. Tapon, D. Houle, I. Salazar-Ciudad, and B. J. Thompson. Patterned Anchorage to the Apical Extracellular Matrix Defines Tissue Shape in the Developing Appendages of Drosophila. *Dev. Cell*, 34(3):310–322, Aug 2015.

P. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Computational and Applied Mathematics*, 20:53–65, 1987.

E. Rozbicki, M. Chuai, A. I. Karjalainen, F. Song, H. M. Sang, R. Martin, H.-J. Knölker, M. P. MacDonald, and C. J. Weijer. Myosin-ii-mediated cell shape changes and cell intercalation contribute to primitive streak formation. *Nat Cell Biol*, 17:397–408, 2015.

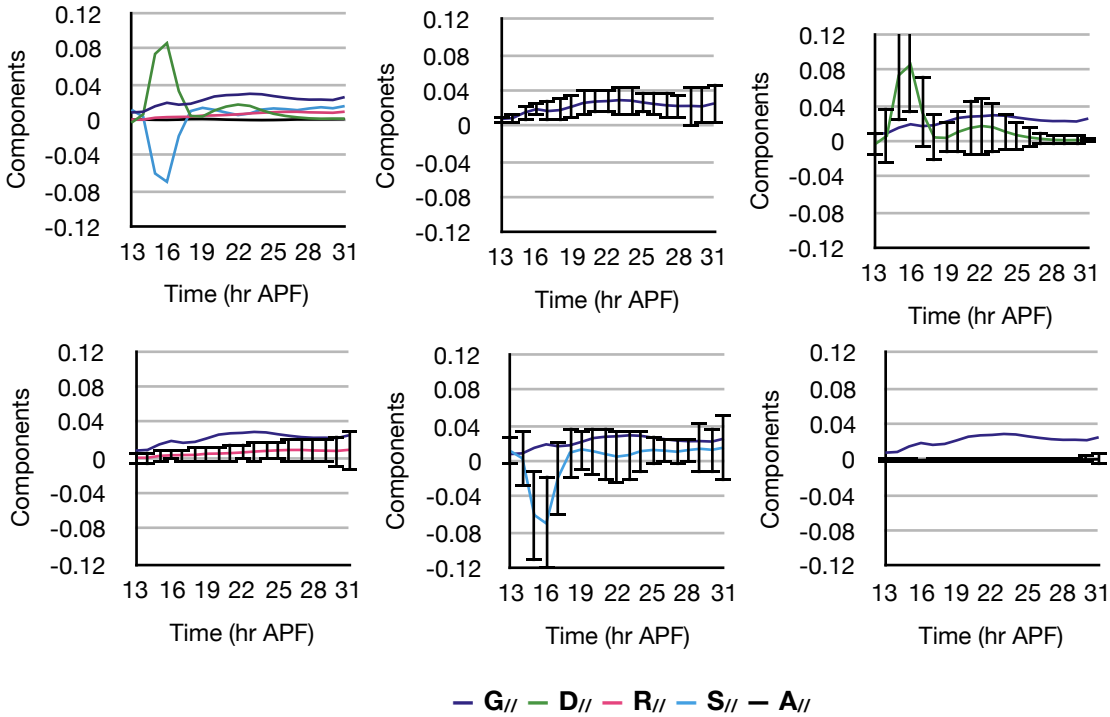

Fig. S1: Variance of effective contribution of cellular processes at each time point. Plots show time evolution of effective contribution of cellular processes in the *Drosophila* notum and standard deviation of them.

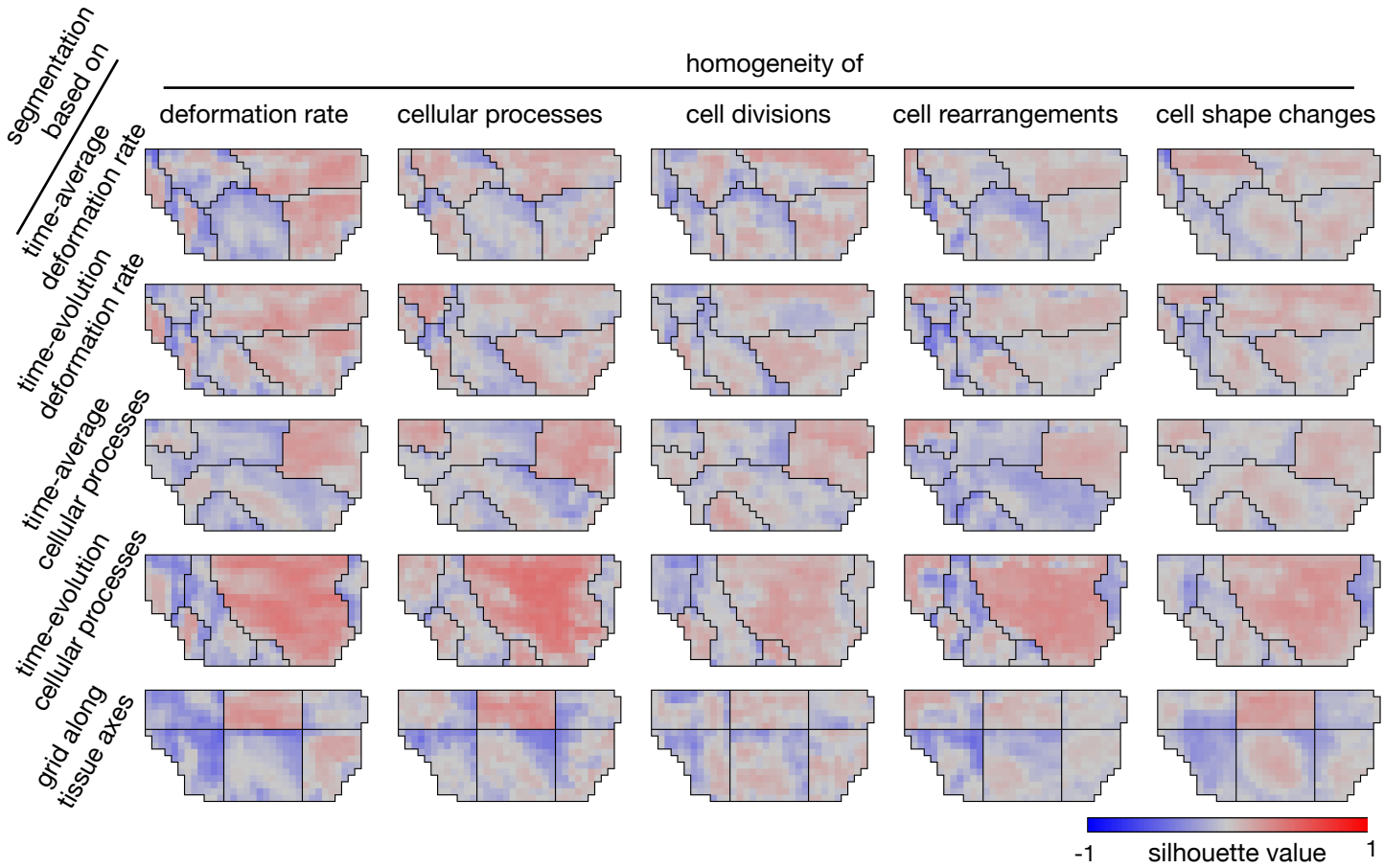

Fig. S2: Heat maps of silhouette value. First row: segmentation based on time-average tissue deformation rate. Second row: segmentation based on time-evolution of deformation rate. Third row: segmentation based on time-average cellular processes effective contributions. Fourth row: segmentation based on time-evolution of cellular processes. Fifth row: conventional segmentation of large grid parallel to tissue axes. First column: silhouette values measured in the property space of time-evolution of deformation rate. Second column: silhouette values measured by time-evolution of cellular processes. Third column: silhouette values measured by time-evolution of cell divisions. Fourth column: silhouette values measured by time-evolution of cell rearrangements. Fifth column: silhouette values measured by time-evolution of cell shape changes.

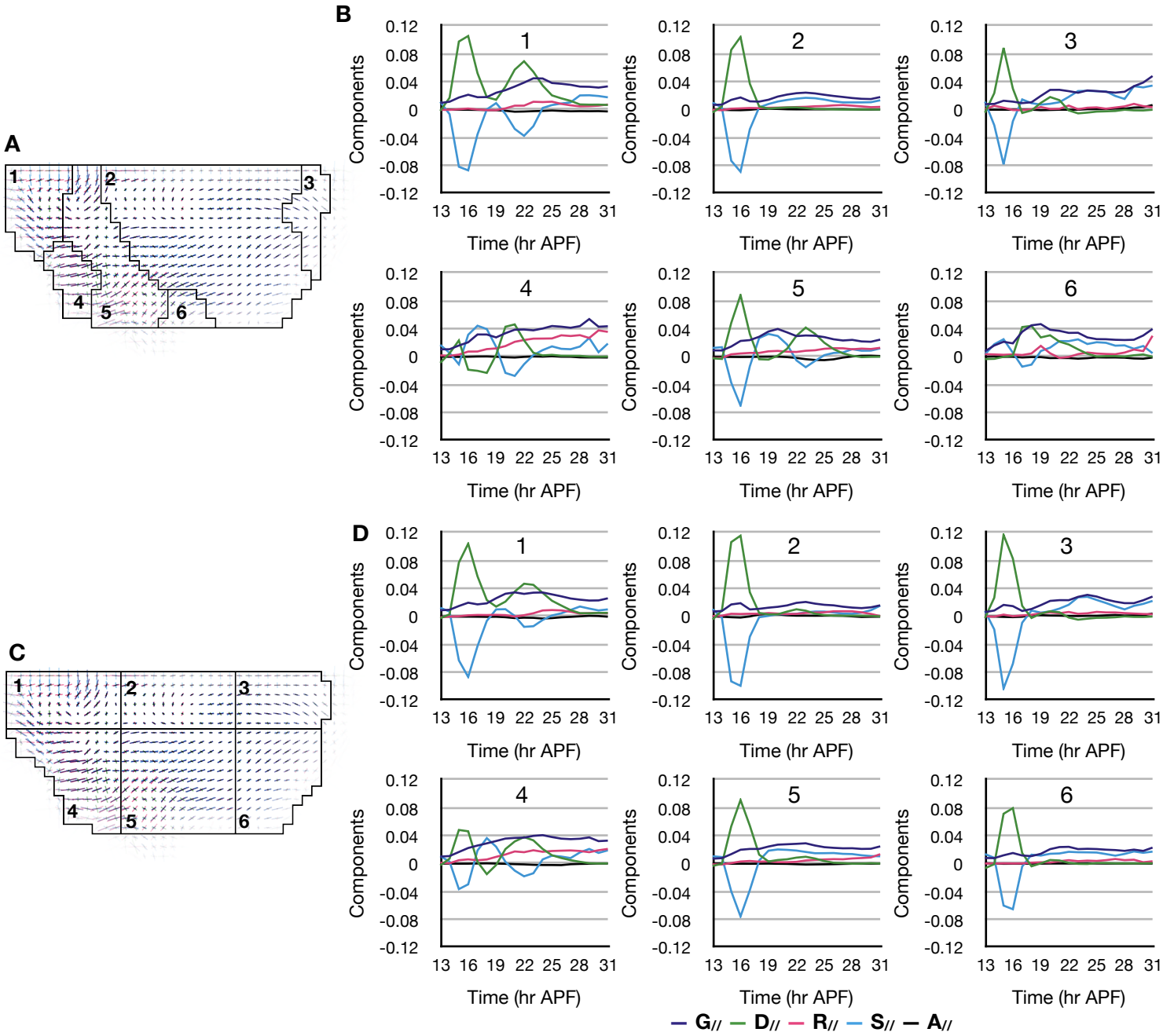

Fig. S3: Plots of cellular processes in the segmentations based on time evolution of cellular processes and the conventional large grid. (**A**, **B**) The tissue segmentation based on time-evolution of cellular processes (**A**) and plots of cellular processes effective contributions averaged in each region (**B**). The numbers indicate the regions. (**C**, **D**) The large grid (**C**) and plots of cellular processes in each region (**D**).