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

Satoshi Yamashita^{*1}, Boris Guirao², and François Graner^{*1}

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

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

May 7, 2020

*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. We also illustrate it on wing blade morphogenesis. 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

8 covering ectoderm [Rauzi et al., 2015, Perez-Mockus et al., 2017], between invaginating
9 midgut and elongating germ-band [Collinet et al., 2015, Lye et al., 2015, Dicko et al.,
10 2017] of *Drosophila* embryo, between contracting wing hinge and expanding wing blade
11 in *Drosophila* pupa [Etournay et al., 2015, Ray et al., 2015], or between invaginating neu-
12 ral plate and covering epidermal ectoderm of *Xenopus* embryo [Brodland et al., 2010],
13 provide examples of how mechanical force generated in one region can drive large scale
14 deformation in adjacent regions. In these cases, the regions which behave differently are
15 easily distinguished by specific gene expressions.

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

32 To study the spatial regulation of morphogenesis at tissue scale, we developed a new
33 multi-technique pipeline to divide a tissue into sub-regions based on quantitative mea-
34 surements of static or dynamic properties of cells or tissues. Our tissue segmentation
35 pipeline consists of three steps: a first fast tissue segmentation attempted several times
36 with random seeding, then merging these multiple tissue segmentations into a single

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.

one, then smoothing the resulting regions boundaries (Fig. 1C). We apply it to the 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

64 heart based on cell polarity [Le Garrec et al., 2013].

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

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

83 Among the various results because of the random initial seeding, we don't know
 84 which one should be compared with gene expression patterns or fed forward to a study
 85 of mechanical interactions between the regions. For practical applications, we need a
 86 single segmentation result for a given morphogenetic property.

87 **3.2 Defining a single tissue segmentation using label propagation on a** 88 **consensus matrix**

89 To obtain a single tissue segmentation, we turned to consensus clusterings. In fact, since
 90 resultant segmentations of the region growing were dependent on randomly given initial
 91 values, we ran multiple trials and merged multiple segmentation results into a single

Fig. 2: Tissue segmentation by region growing algorithm. Cell patches expansion/contraction rates are 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.

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

94 algorithms, and found the *label propagation on a consensus matrix* [Lancichinetti and
95 Fortunato, 2012, Raghavan et al., 2007] returning regions similar to the results of region
96 growing.

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

103 The label propagation returned results similar to the region growing segmentations
104 (Fig. 2B, 3B). Also, the label propagation results were more similar to each other than
105 results of region growing, assessed with adjusted Rand indices (ARI), a measurement
106 of similarity between two partitions (ARI of identical partitions would be 1). ARI were
107 0.50 ± 0.21 among the results of the region growing and 0.97 ± 0.02 among the results of
108 the label propagation. They showed similar average silhouette values, similar to median
109 of those of region growing results, but smaller than the highest value of those of region
110 growing (Fig. 3C). The average silhouette of the label propagation result was higher
111 than those of 99.95% of the randomly made control segmentations.

112 However, a consensus clustering algorithm ignores original properties of objects in
113 principle and divides the objects only based on how they were divided among given
114 partitions, and thus it might return disconnected regions and zigzag boundary between
115 them. Some segmentations in Figure 3B also included disconnected regions as marked
116 by gray color.

117 3.3 Smoothing of tissue segmentation results by cellular Potts model

118 To smooth the consensus regions boundaries, we employed cellular Potts model, which
119 simulates dynamics of a cellular tissue by calculating energies of cells from their geometry,
120 trying to decrease the total energy. In our application, the energy was lower when the
121 region boundary was shorter and the homogeneity was higher (Fig. 4A). The boundary

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.

length and homogeneity were balanced so that all regions had enough smooth boundary,
evaluated by a circularity [Bosveld et al., 2016], and was as homogeneous as possible.

124 It smoothed boundaries and removed disconnected cell patches (Fig. 4B, C) while
 125 keeping the average silhouette higher than those of random segmentations (Fig. 4D).
 126 Since the cellular Potts model simulation includes the Metropolis update, i.e., choosing
 127 a pixel randomly and updating the pixel by probability according to a change of the
 128 energy, resultant smoothed segmentations varied among different trials even with the
 129 same parameters and initial segmentation. Therefore we iterated the cellular Potts
 130 model smoothing 50 times and integrated its results by the label propagation algorithm
 131 again.

132 Now we have a pipeline of the region growing, the label propagation, and the cellular
 133 Potts model to divide a field of property (scalar, tensor, or any kind of value with metric)
 134 into regions. The resultant regions are homogeneous, where points in each region are
 135 more similar to each other than to points in other regions.

136 4 Results II : Tissue segmentation based on tissue mor- 137 phogenesis

138 We now turn to property spaces better representing tissue morphogenesis. In Guirao
 139 et al. [2015], tissue deformation rate ($\mathbf{G}$) and underlying cellular processes, cell division
 140 ($\mathbf{D}$), cell rearrangement ($\mathbf{R}$), cell shape change ($\mathbf{S}$), and cell delamination ($\mathbf{A}$) were
 141 quantified into tensors. The tensors were obtained from change of the texture averaged
 142 over 20 hr from 12 hr APF to 32 hr APF or over 2 hr at each time point. By comparing
 143 the tensors, for example, one can check whether cell divisions and cell rearrangements
 144 elongated tissue in the same direction or attenuated each other. In the same way,
 145 by comparing the tensors of deformation rates with a unit tensor which has the same
 146 direction of elongation with tissue deformation rate, one can estimate an amplitude
 147 of the tissue deformation rate and how much the cellular processes contribute to the
 148 tissue deformation in both terms of contraction/expansion (isotropic deformation) and
 149 narrowing/elongation (anisotropic deformation) [Guirao et al., 2015]. They are scalar
 150 value and denoted by $\mathbf{G}_{//}$ for the tissue deformation rate, $\mathbf{D}_{//}$ for cell division, $\mathbf{R}_{//}$ for cell

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.

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. Therefore we included the time evolution in the property space. Assume that there are two regions where one expands during 14-17 hr APF while another expands during 25-28 hr APF resulting in similar size changes, then the two regions cannot be distinguished by the time-average expansion rate. To distinguish them, we compared a property at each time point and summed up its difference through the whole time. When two cell patches always behaved similarly, then the difference at each time point is small and so the total difference is also small, whereas cell patches with deformations occurring in different timing are separated at each time point and thus the total difference get large. In contrast with time-average, we call the sum of difference at each time point *time-evolution*.

4.1 Tissue segmentations based on tissue deformation rate and cellular processes effective contributions

We first divided the tissue based on time-average and time-evolution of tissue deformation rate. The similarity was given by Euclidean distance of tensors (see methods for detail). The notum was divided into anterior-middle-posterior and medial-lateral regions by both of the time-average and time-evolution, while the middle regions were smaller and the middle lateral region extended medially in the segmentation based on time-evolution (Fig. 5A-D).

Next, we divided the tissue based on time-average and time-evolution of cellular processes. The amplitude of tissue deformation rate and cellular processes effective

179 contributions were combined in a vector, and their similarity was given by Euclidean
180 distance of vector. In contrast to the segmentations based on the time-average and
181 time-evolution of tissue deformation rate, the segmentations based on time-average and
182 time-evolution of the cellular processes were dissimilar to each other (Fig. 5E-H). The
183 segmentation based on time-evolution of cellular processes included a posterior region, a
184 large anterior region, a neck-notum boundary region, lateral posterior region, a scutum-
185 scutellum boundary region, and a lateral region (Fig. 5H).

186 4.2 Correspondence between segmentations based on cellular processes 187 and tissue deformation rate

188 Both of the segmentations based on time-evolution of tissue deformation rate and cellular
189 processes effective contributions included the large anterior region, the middle boundary
190 region, the lateral posterior region, and the posterior region, although the anterior and
191 posterior regions were divided into medial and lateral subregions in the segmentation
192 based on the tissue deformation rate. Figure 6A-D show overlap between segmentation
193 based on time-evolution of cellular processes (Fig. 5H) and the others (Fig. 5B, D, F) or
194 a conventional large grid parallel to tissue axes. The middle lateral and posterior lateral
195 regions in the segmentation based on time-evolution of tissue deformation rate and the
196 middle scutum-scutellum boundary region and lateral posterior regions in the segmen-
197 tation based on time-evolution of cellular processes overlapped each other (Fig. 6C).
198 We also evaluated the overlap between the segmentations by ARI (Fig. 6E). Despite the
199 difference between the anterior subregions, the segmentations based on time-evolution
200 of tissue deformation rate and cellular processes overlapped each other more than the
201 others.

202 4.3 Homogeneity of the obtained regions

203 Next, we evaluated the homogeneity of the obtained regions. The time-evolution of
204 tissue deformation rate was similar among cells inside regions of the segmentations based
205 on time-average and time-evolution of tissue deformation rate except the middle-lateral

Fig. 5: Segmentations based on tissue deformation and underlying cellular processes. For each cell patch, direction of elongation is represented by a bar, and the effective contributions of cellular processes are indicated by relative directions of deformation rate between the tissue and each cellular process. 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).

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.

206 region of the former (Fig. 7A, B). On the other hand, the large grid segmentation showed
 207 large heterogeneity in the posterior regions (Fig: 7C). The average silhouette value of
 208 the segmentation based on the time-evolution of deformation rate was higher than those
 209 of 99.5% of the control segmentations (average silhouette for label propagation: 0.0568,
 210 for smoothed regions: 0.0600, the maximum of the smallest 99.5% of controls average
 211 silhouettes: 0.0425) (Fig. 7D). The average silhouette of the segmentation based on time-
 212 averaged tissue deformation rate was also higher than 95% of the control segmentations
 213 (average silhouette for label propagation: 0.0336, for smoothed regions: 0.0260, the
 214 maximum of the smallest 95% of controls average silhouettes: -0.0098). On the other
 215 hand, that of the conventional grid segmentation was close to median of the control
 216 segmentations (average silhouette: -0.0694).

217 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: 0.1448, 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: 0.0125, for smoothed regions: -0.0385, the maximum of the smallest 95% of controls average silhouettes: 0.0241) (Fig. 7H).

Our tissue segmentation is designed to divide a tissue into regions homogeneous in a given property space, and the homogeneity of either tissue deformation rate or cellular processes in the segmentations based on each property demonstrated that the pipeline worked fine (Fig. 7B, F). However, 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. Even though the homogeneity in the regions differed among the different property spaces, the segmentations based on time-evolution of tissue deformation rate and cellular processes showed higher homogeneity than the others also in the property spaces of deformation rates due to cell divisions, cell rearrangements, and cell shape changes.

We projected the regions divided based on the time-evolution of cellular processes onto the cells, and found that the anterior and posterior regions corresponded to scutum and scutellum, and the middle boundary and lateral posterior regions corresponded to the scutum-scutellum boundary (Fig. S4). This result demonstrates that the obtained regions corresponded to the anatomical features.

4.4 Cellular processes effective contributions inside the regions

Figure 8 shows plots of cellular processes effective contributions average in each region of the segmentation based on time-evolution of cellular processes. The second peak of cell division was observed only in the posterior regions and the scutum-scutellum bound-

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.

ary region, consistent with the preceding studies with maps of number of cell divisions [Bosveld et al., 2012, Guirao et al., 2015], while we also found the first peak of cell division small in the lateral posterior region. Plots of average cellular processes differed from each other also among regions in the segmentation based on tissue deformation rate but less distinctive in the large grid segmentation (Fig. S3). Distances between the plots in Figure 8 were 0.65 ± 0.16 and those for the segmentation based on tissue deformation rate were 0.63 ± 0.20 , larger than those for large grid segmentation (0.44 ± 0.14). This result demonstrates that cellular processes in the obtained segmentations were more distinctive than those in the conventional grid.

4.5 Application to the morphogenesis of wing blade

To demonstrate the generality of our method to divide a tissue, we performed the same segmentation and analysis in the *Drosophila* pupa wing blade. During 15-32 hr APF, the wing blade is elongated in proximal-distal direction by a contracting wing hinge connected with the wing blade proximal side while its distal side is anchored to a cuticle via Dumpy [Etournay et al., 2015, Ray et al., 2015]. The wing hinge contraction also narrows it in the anterior-posterior direction and induces shear strain in wing blade proximal anterior and posterior regions (Fig. 9A, B). We performed the tissue segmentation for the wing blade based on time-evolution of tissue deformation rate (Fig. 9C) and cellular processes (Fig. 9D) dividing into four regions. In both cases, the wing blade was divided into anterior, middle, posterior, and distal regions. All regions showed positive silhouette values (Fig. 9E, F), and their averages were significantly higher than the average silhouette values of control segmentations (Fig. 9G, H). Plots of the cellular processes effective contributions also showed distinctive patterns between the regions, where the cell division showed small contribution in the anterior region, the cell rearrangements dominated the tissue deformation around 26 hr APF in the anterior and posterior regions, and the cell shape changes showed two peaks around 17 and 22 hr APF in the distal region (Fig. 9I). Projection of the four regions onto the cells showed a difference between the regions and interveins, whereas the posterior region roughly corresponded

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.

274 to the proximal posterior intervein and the boundary between the anterior and distal
275 regions corresponded L3 vein (Fig. S5).

276 5 Discussion

277 This study demonstrates that the pipeline of the region growing, the label propagation on
278 the consensus matrix, and the boundary smoothing by cellular Potts model could divide
279 a deforming heterogeneous tissue into homogeneous regions based on the morphogenetic
280 properties. Using this segmentation method, we divided the developing dorsal thorax
281 and wing of *Drosophila* pupa, and found regions with distinctive tissue deformation rate
282 and underlying cellular processes.

283 The tissue segmentation based on morphogenesis differs from conventional image seg-
284 mentations, cell segmentations, and other tissue segmentations those are an automation
285 of manual segmentation and can be corrected manually. First, the morphogenesis was
286 quantified as multiple tensor fields with time evolution, and thus it is hard to visualize
287 them in a 2D image for manual segmentation. Second, it is not easy to evaluate whether
288 given a region actually corresponds to genetical/mechanical regulation of morphogenesis.

Fig. 9 (following page): Segmentation of *Drosophila* wing blade into four regions. (**A**, **B**) Deformation of a wing blade from 15 hr (**A**) to 32 hr APF (**B**). Cell patches are shown with blue and white check pattern. (**C**, **D**) Segmentation based on time-evolution of tissue deformation rate (**C**) and its heat map of silhouette value (**D**). (**E**, **F**) Segmentation based on time-evolution of cellular processes (**E**) and its heat map of silhouette value (**F**). (**G**, **H**) Histogram of average silhouette value of control segmentations for the property spaces of tissue deformation rate (**G**) and cellular processes (**H**). Red vertical lines show average silhouette values of label propagation results. Black vertical lines show average silhouette values of smoothed regions. Dotted blue lines show threshold for the highest 0.05% (**G**) and 0.005% (**H**) of the control segmentations. (**I**) Plots of cellular processes effective contributions averaged in each region of 1-4 in **D**.

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 notum segmentations based on time-evolution of tissue deformation rate and effective contributions cellular processes returned similar regions corresponding to the scutum, scutellum, and the boundary between them. By the tissue deformation rate, the scutum and scutellum 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 tissue deformation rate, could be interpreted as regions with different cellular processes but of similar tissue deformations. The wing blade was divided into anterior, middle, posterior, and distal regions based on both of the tissue deformation rate and cellular processes, but the regions did not matched the wing veins pattern.

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.

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. Also, this method is not dependent on

318 how the morphogenesis was quantified, and one can include rotational movement by
319 anti-symmetric strain rate tensors, or 3D deformation by using voxels instead of pixels.

320 6 Methods

321 6.1 Quantification tools

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

329 6.1.1 Quantification of tissue deformation and cellular processes

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

336 Epithelial cells contours were detected automatically using watershed algorithm, the
337 cells were tracked, adjacencies between cells were listed, and relative positions of centers
338 of adjacent cells were recorded. The tissue deformation rate, denoted by the symmetric
339 tensor $\mathbf{G}$, was obtained from change of relative positions between neighbor cells over
340 20 hr from 12 hr APF to 32 hr APF, or over 2 hr at each time point when recording
341 the time evolution. The tissue deformation rate $\mathbf{G}$ was then decomposed into cell shape
342 change $\mathbf{S}$ and deformation accompanied by change of cell adjacency, which was further
343 divided into cell division $\mathbf{D}$, cell rearrangement $\mathbf{R}$, and cell delamination $\mathbf{A}$, which are

344 symmetric tensors too.

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

$$\mathbf{G} = \mathbf{D} + \mathbf{R} + \mathbf{S} + \mathbf{A}. \quad (1)$$

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

$$\mathbf{Q} \cdot \mathbf{Q}' = \frac{1}{d} \text{Tr}(\mathbf{Q} \mathbf{Q}'^T), \quad (2)$$

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

$$\mathbf{u}_\mathbf{G} = \frac{\mathbf{G}}{(\mathbf{G} \cdot \mathbf{G})^{1/2}}. \quad (3)$$

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

$$\begin{aligned} \mathbf{G}_\parallel &= \mathbf{G} \cdot \mathbf{u}_\mathbf{G} \\ &= (\mathbf{D} + \mathbf{R} + \mathbf{S} + \mathbf{A}) \cdot \mathbf{u}_\mathbf{G} \\ &= \mathbf{D} \cdot \mathbf{u}_\mathbf{G} + \mathbf{R} \cdot \mathbf{u}_\mathbf{G} + \mathbf{S} \cdot \mathbf{u}_\mathbf{G} + \mathbf{A} \cdot \mathbf{u}_\mathbf{G} \\ &= \mathbf{D}_\parallel + \mathbf{R}_\parallel + \mathbf{S}_\parallel + \mathbf{A}_\parallel. \end{aligned} \quad (4)$$

349 The scalar $\mathbf{G}_\parallel$ represents the local magnitude of tissue morphogenesis, and $\mathbf{D}_\parallel$, $\mathbf{R}_\parallel$,
350 $\mathbf{S}_\parallel$, and $\mathbf{A}_\parallel$ represent the effective contributions of the cellular processes to the tissue
351 morphogenesis. When a cellular process produces an anisotropic deformation in the
352 same direction with that of tissue, e.g. cells divided in the same direction with tissue
353 elongation, the scalar product between them returns a positive value, while it returns
354 negative value when a cellular process counteracts tissue deformation.

355 6.1.2 Metric

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

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

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

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

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

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

363 as an analogy to [distance between functions](#).

364 For the composition of cellular processes, the tensors of cellular processes were con-
365 verted to effective contributions and combined into a vector $(\mathbf{G}_{//}, \mathbf{D}_{//}, \mathbf{R}_{//}, \mathbf{S}_{//}, \mathbf{A}_{//})$. A
366 distance between two vectors was given by [Euclidean distance, the square root of the](#)
367 [sum of the square of the differences between corresponding elements](#), and a distance
368 between [vectors with time-evolution \$v\(t\)\$ and \$v'\(t\)\$](#) was given by a sum of the distance
369 at each time point,

$$|v - v'| = \int \|v(t) - v'(t)\| dt. \quad (7)$$

370 6.1.3 Silhouette and bootstrap

371 Silhouette quantifies clustering results, indicating how well an object resembles other
372 objects inside its own cluster [Rousseeuw, 1987]. [Assume that \$n\$ objects \$\{p_1, p_2, \dots, p_n\}\$](#)
373 [are partitioned into \$k\$ clusters \$\{C_1, C_2, \dots, C_k\}\$. For an object \$p_i \in C_I\$, we can compute](#)
374 [average distance \$a\(p_i\)\$ from \$p_i\$ to all other objects in \$C_I\$. For \$J \neq I\$, we can also compute](#)
375 [average distance \$d\(p_i, C_J\)\$ from \$p_i\$ to all objects in \$C_J\$, and can select the smallest of](#)
376 [those, denoted by \$b\(p_i\) = \min_{J \neq I} d\(p_i, C_J\)\$. The silhouette value \$s\(p_i\)\$ is obtained by](#)
377 [combining \$a\(p_i\)\$ and \$b\(p_i\)\$ as follow:](#)

$$s(p_i) = \frac{b(p_i) - a(p_i)}{\max\{a(p_i), b(p_i)\}}. \quad (8)$$

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 0 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 $A = \{A_1, \dots, A_k\}$ and $B = \{B_1, \dots, B_m\}$ 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 , $\text{RI}(A, B)$ is given by the function

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

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

399 by the function

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

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

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

401 6.2 Tissue segmentation pipeline

402 The pipeline was implemented by custom Matlab scripts, in three steps (Fig. 1C). The
403 Matlab scripts are available at GitHub (<https://doi.org/10.5281/zenodo.3626111>).

404 6.2.1 Region growing tissue segmentation

405 The initial tissue segmentation first defines a metric of similarity between cells, and then
406 a tissue is divided into regions containing similar cells. This approach was inspired by
407 a study segmenting mouse heart based on cell polarity [Le Garrec et al., 2013]. On
408 the assumption that expression patterns of genes responsible for morphogenesis make
409 connected regions, and to study physical interactions between the regions, we aimed at
410 getting connected regions.

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

418 Initial seeds were randomly chosen from data, and regions were expanded by adding
419 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. An energy of system depends on the shape of cells, and the pattern is updated in an iteration 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 : $\mathcal{H} = \sum_{\text{regions}} [JP + \lambda(A - A_0)^2 - hs]$. 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. In the simulations, the updates of labels were iterated 50 times. For resultant regions, a circularity C was calculated, where it was defined as $C = 4\pi \times \text{area}/\text{perimeter}^2$ in Bosveld et al. [2016]. The parameters J , λ , h , and T were screened for resultant regions with the highest homogeneity and circularity larger than 0.45. With the screened parameters, the boundary smoothing was iterated for 50 times, and the results were integrated again by the label propagation on a consensus matrix algorithm.

7 Acknowledgement

The authors thank Floris Bosveld for imaging, Stéphane U. Rigaud for cell segmentation and tracking, Yohanns Bellaïche for useful 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.

471 10 Funding

472 This research was supported by grants to S.Y. from Uehara Memorial Foundation
473 (201630059), Japan Society for the Promotion of Science (201860572).

474 References

- 475 R. Adams and L. Bischof. Seeded region growing. *IEEE Trans Pattern Anal Mach Intell*,
476 16:641–647, Jun 1994.
- 477 G. B. Blanchard, A. J. Kabla, N. L. Schultz, L. C. Butler, B. Sanson, N. Gorfinkel,
478 L. Mahadevan, and R. J. Adams. Tissue tectonics: morphogenetic strain rates, cell
479 shape change and intercalation. *Nat Methods*, 6:458–464, 2009.
- 480 F. Bosveld, I. Bonnet, B. Guirao, S. Tlili, Z. Wang, A. Petitalot, R. Marchand, P.-L.
481 Bardet, P. Marcq, F. Graner, and Y. Bellaïche. Mechanical control of morphogenesis
482 by fat/dachsous/four-jointed planar cell polarity pathway. *Science*, 336:724–727, 2012.
- 483 F. Bosveld, B. Guirao, Z. Wang, M. Riviere, I. Bonnet, F. Graner, and Y. Bellaïche.
484 Modulation of junction tension by tumor suppressors and proto-oncogenes regulates
485 cell-cell contacts. *Development*, 143(4):623–634, Feb 2016.
- 486 G. W. Brodland, X. Chen, P. Lee, and M. Marsden. From genes to neural tube defects
487 (ntds): insights from multiscale computational modeling. *HFSP J*, 4:142–152, 2010.
- 488 C. Collinet, M. Rauzi, P.-F. Lenne, and T. Lecuit. Local and tissue-scale forces drive
489 oriented junction growth during tissue extension. *Nat Cell Biol*, 17:1247–1258, 2015.
- 490 M. Dicko, P. Saramito, G. B. Blanchard, C. M. Lye, B. Sanson, and J. Etienne. Geometry
491 can provide long-range mechanical guidance for embryogenesis. *PLoS Comput. Biol.*,
492 13(3):e1005443, Mar 2017.
- 493 A. D. Economou, L. J. Brock, M. T. Cobourne, and J. B. Green. Whole population

494 cell analysis of a landmark-rich mammalian epithelium reveals multiple elongation
495 mechanisms. *Development*, 140(23):4740–4750, Dec 2013.

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

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

501 J. Firmino, D. Rocancourt, M. Saadaoui, C. Moreau, and J. Gros. Cell Division Drives
502 Epithelial Cell Rearrangements during Gastrulation in Chick. *Dev. Cell*, 36(3):249–
503 261, Feb 2016.

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

506 B. Guirao, S. U. Rigaud, F. Bosveld, A. Bailles, J. Lopez-Gay, S. Ishihara, K. Sugimura,
507 F. Graner, and Y. Bellaïche. Unified quantitative characterization of epithelial tissue
508 development. *Elife*, 4:e08519, Dec 2015.

509 L. Hubert and P. Arabie. Comparing partitions. *Journal of Classification*, 2:193–218,
510 1985.

511 A. Lancichinetti and S. Fortunato. Consensus clustering in complex networks. *Sci Rep*,
512 2:336, 2012.

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

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

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

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

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

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

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

536 M. Rauzi, U. Krzic, T. E. Saunders, M. Krajnc, P. Zihler, L. Hufnagel, and M. Lep-
537 tin. Embryo-scale tissue mechanics during *Drosophila* gastrulation movements. *Nat*
538 *Commun*, 6:8677, 2015.

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

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

545 E. Rozbicki, M. Chuai, A. I. Karjalainen, F. Song, H. M. Sang, R. Martin, H.-J. Knölker,
546 M. P. MacDonald, and C. J. Weijer. Myosin-ii-mediated cell shape changes and cell
547 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 tissue deformation rate and the conventional large grid. (**A**, **B**) The tissue segmentation based on time-evolution of tissue deformation rate (**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**). Scale bars in **A** and **C** indicate deformation rate 0.02 h^{-1} with colors for tissue and cellular processes.

Fig. S4: Projection of the segmentation onto the notum cells. The segmentations based on time evolution of cellular processes were projected. (A, B) The segmentation was projected onto the notum cells at 12 hr (A) and 32 hr APF (B), where regions were indicated by colors. The regions corresponded to scutum (pale blue and green), scutellum (red), and boundary between them (dark blue and purple). Scale bars indicate 50 μm .

Fig. S5: Projection of the segmentation onto the wing blade cells. The segmentations based on time evolution of cellular processes were projected. (A, B) The segmentation was projected onto the wing blade cells at 15 hr (A) and 32 hr APF (B), where the regions were indicated by colors. Scale bars indicate 50 μm .

1 Pseudo codes for tissue segmentation algorithms

In below pseudo codes show algorithms of the automatic tissue segmentation. Matlab custom functions and framework developed for this study are available at GitHub (<http://doi.org/10.5281/zenodo.3626111>). For details of the functions and framework, see its README file and comments in the codes.

1.1 Region growing algorithm

Algorithm 1 shows a pseudo code of the region growing image segmentation in Matlab-like syntax. It divides a bitmap image stored in a data object *dataMap*. In the algorithm, a number of regions, a limit to update the seeds, and a metric are given as parameters. With the parameters, supporting objects *seedList*, *meanList*, *regionsList*, *meter*, and *seeder* are allocated and initialized. The *seedList*, *meanList*, and *regionsList* are instances of data object with a property *var* representing seeds and means of regions and regions, shared among the supporting objects. The *meter* is an object measuring distance between the mean of region and a point adjacent to the region. A method *measure* returns the distance measured by the given metric. The *seeder* is an object choosing seeds of regions. Methods *initSeeds* and *initialMeans* return indices of randomly chosen points and their values. Once the *dataMap* was divided into regions, methods *newSeeds* and *newMeans* return indices of points at center of the regions and mean values of the regions. A method *initQueue* returns an array where its element represents a point adjacent to one of the seeds and holds the region and distance to the region's mean value. Inside a loop, a point in the queue with the smallest distance to the region's mean value is added to the region, and points adjacent to the point, returned by a method *neighborsOfPoint* of *dataMap*, are added to the queue.

In our tissue segmentation, a Matlab custom function *run_region_growing()* iterates this algorithm for given time, returning a stack of resultant partitions.

1.2 Label propagation on a consensus matrix

Algorithm 2 shows a pseudo code of the label propagation. It divides N objects into clusters based on an $N \times N$ consensus matrix M whose rows and columns correspond to the objects, and an element m_{ij} represents frequency the i -th and j -th objects were included in a cluster among given clustering results. A parameter t_M indicates a threshold value, where elements in M smaller than t_M is ignored in the label propagation.

In the tissue segmentation, 50 results of region growing were converted to the consensus matrix and given to a Matlab custom function *run_label_propagation()* implementing the label propagation. Number of resultant regions is influenced by t_M , and thus a Matlab custom function *run_cm_thresholding_lp()* screens t_M value so that *run_label_propagation()* returns the same number of regions with the given partitions.

Algorithm 1: Region growing algorithm

```

input : dataMap to be segmented and parameters.
% seedList, meanList, regionsList, allocatedList, meter, and seeder are
  supporting objects and variable initialized with the parameters.

seedList.var = seeder.initalSeeds;
meanList.var = seeder.initalMeans;
while loop counter is smaller than limit do
  % Initialize partition, allocated list, queue.
  regionsList.var(:) = false;
  allocatedList(:) = false;
  queue = seeder.initalQueue;
  while queue is not empty do
    point = queue(1);
    if allocatedList.var(point.index) == false then
      % Grow region to the point.
      regionsList.var(point.index, point.region) = true;
      allocatedList(point.index) = true;
      % Enqueue neighbors of the point.
      array = dataMap.neighborsOfPoint(point.index);
      for neighbor in array do
        neighbor.region = point.region;
        neighbor.distance = meter.measure(neighbor);
        queue = cat(1, queue, neighbor);
      % Remove the allocated point from queue.
      queue(1) = [];
      % Sort queue.
      [values, indices] = sort([queue.distance]);
      queue = queue(indices);
    else
      queue(1) = [];
  % Check convergnence.
  lastMeanList = meanList.var;
  seedList.var = seeder.newSeeds;
  meanList.var = seeder.newMeans;
  if isequal(lastMeanList, meanList.var) then
    break;
return regionsList.var

```

Algorithm 2: Label propagation

```

input : Matrix  $M$  and threshold  $t_M$ .

% Cut elements in  $M$  smaller than the  $t_M$ .
 $M(M < t_M) = 0$ ;
% Make labelArray representing labels on  $N$  vertices.
labelArray = (1:N)';

flag = true;
while flag do
    flag = false;
    % Enumerate vertices in random order and update their label.
    for  $i = \text{randperm}(N)$  do
        % Make labelMatrix representing labels on vertices.
        labelMatrix = labelArray == 1:N;
        % Choose label most weighted by edges incident to the  $i$ -th vertex.
        ,indices= max(sum( $M(:,i) .* \text{labelMatrix}$ ),1));
        if  $\text{labels}(i) \neq \text{indices}(1)$  then
            % Update label of the  $i$ -th vertex.
            labelArray( $i$ ) = indices(1);
            flag = true;
        end if
    end for

% Convert labelArray to a matrix.
labelMatrix = labelArray == 1:N;
indices = any(labelMatrix,1);
partition = labelMatrix(:,indices);

return partition

```

1.3 Cellular Potts model

Algorithm 3 shows a pseudo code of the cellular Potts model. It simulates a deformation of regions (partition of dataMap) by giving small fluctuations. In the algorithm, an array of function handles, coefficients to combine the functions results, the system temperature, and number of label updates are given as parameters. With the regions and parameters, supporting objects *analyser* and *dict* are allocated and initialized. The functions in the array calculate system energy with the analyser and dict. For each fluctuation, one of points at regions rim returned by analyser *rim_points* is selected randomly, and a label of neighboring points is also selected randomly and copied. Connectedness of a region is checked locally, with a coordinate of neighboring points returned by dataMap *coordinates*.

In the tissue segmentation, a Matlab custom function *run_CPM_smoothing()* implement this algorithm with energy functions consist of area constraint, surface tension, and total silhouette value. The coefficients and temperature influence resultant regions, and thus a Matlab custom function *run_CPM_fitting()* screens the parameters so that *run_CPM_smoothin()* returns smoothed regions with a circularity larger than given value and the total silhouette value as large as possible.

Algorithm 3: Cellular Potts model with region homogeneity

```

input : Partition, dataMap, and parameters
% regionList, analyser, dict, H_functions, coefficients, T, and counter are
% supporting objects and variables initialized with the parameters.

% Calculate the system energy.
H = 0;
for k = 1:length(H_functions) do
    fh = H_functions(k);
    H = H + fh(analyser,dict) * coefficients(k);

% Update labels for given times.
while true do
    % Select a point randomly.
    rim = analyser.rim_points;
    rim = find(rim);
    if isempty(rim) then
        % There is only one region.
        break;
    i = ceil(rand() * length(rim));
    i = rim(i);

    % Select a label from neighbors of the point.
    neighbors = dataMap.neighborsOfPoint(i);
    j = ceil(rand() * length(neighbors));
    j = neighbors(j);
    if any(regionsList.var(i,:) & regionsList.var(j,:)) then
        % The i-th and j-th points are in a region.
        continue;

    % Check connectedness.
    m = zeros(3,'logical');
    x0 = dataMap.coordinates(i).x - 2;
    y0 = dataMap.coordinates(i).y - 2;
    for k = neighbors do
        x = dataMap.coordinates(k).x - x0;
        y = dataMap.coordinates(k).y - y0;
        m(y,x) = any(regionsList.var(i,:) & regionsList.var(k,:));
    array = m([1,2,3,6,9,8,7,4]);
    brray = m([2,3,6,9,8,7,4,1]);
    if sum(array ~= brray) > 2 then
        continue;

    % Get a change of energy.
    oldLabel = regionsList.var(i,:);

```

```

regionsList.var(i,:) = regionsList.var(j,:);
newH = 0;
k = 1:length(H_functions)
    fh = H_functions(k);
    newH = newH + fh(analyser,dict) * coefficients(k);
dH = newH - H;
% Adapt the change when possible.
p = exp(-dH / T);
if p > rand() then
    H = newH;
    counter = counter - 1;
    if counter < 1 then
        break;
else
    regionsList.var(i,:) = oldLabel;
return regionsList.var

```