EpiGraph: an open-source platform to quantify epithelial organization.

Pablo Vicente-Munuera[1,†], Pedro Gómez-Gálvez[1,†], Antonio Tagua[1], Marta Letrán[1], Yanlan Mao[2,3] and Luis M. Escudero[1,*].

1: Departamento de Biología Celular, Universidad de Sevilla and Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla. 41013 Seville, Spain.

2: MRC Laboratory for Molecular Cell Biology, University College London, Gower Street, London WC1E 6BT, UK.

3: College of Information and Control, Nanjing University of Information Science and Technology, Nanjing, Jiangsu 210044, China

†: These authors contributed equally to this work.

*: Corresponding author: L.M. Escudero (lmescudero-ibis@us.es).

SUMMARY

During development, cells must coordinate their differentiation with their growth and organization to form complex multicellular structures such as tissues and organs. Epithelia are packed assembles of cells that will originate the different tissues of the organism during embryogenesis. A big caveat for the analysis of the morphogenetic changes in epithelia is the lack of simple tools that enable the quantification of cell rearrangements. Here we present EpiGraph, an image analysis tool that quantifies epithelial organization. Our method combines computational geometry and graph theory to measure the degree of order of any packed tissue. EpiGraph goes beyond the traditional polygon distribution analysis, capturing other organizational traits that improve the characterization of epithelia. EpiGraph can objectively compare the rearrangements of epithelial cells during development and quantify how the global assemble is affected. Importantly, it has been implemented in the open-access platform FIJI. This makes EpiGraph very simple to use without any programming skills being required.

KEYWORDS

Morphogenesis, Epithelial organization, Graph Theory.

HIGHLIGTHS

Epithelial images can be converted into graphs to quantify their topology.

EpiGraph is a new simple method to capture epithelial organization.

EpiGraph has been designed as a plugin of FIJI to be accessible to any biology lab.

INTRODUCTION

The development of any multicellular organism is based on coordinated changes that transform the embryo into the adult individual. During morphogenesis and growth, patterning, cell divisions and architectural changes must perfectly fit together for the correct development of the body plan. Any morphogenetic movement such as migration, extension or invagination of epithelial cells is coupled with dramatic changes in the organization of cells (Bertet et al., 2004; Blankenship et al., 2006; Escudero et al., 2007; Farhadifar et al., 2007; Girdler and Roper, 2014; Lecuit and Lenne, 2007; Pilot and Lecuit, 2005).

The control of cell proliferation and its role in morphogenesis has been deeply studied and the main signalling pathways that control organ size and shape have been revealed (Lecuit and Le Goff, 2007; Mao et al., 2013; Yu and Guan, 2013). In contrast, how tissues modulate their organization during development is an important question that remains unsolved. The lack of simple and general methods that can capture and quantify the arrangement of cells is responsible for this gap in the understanding of morphogenesis. It is known for almost a hundred years that epithelial tissues exhibit a degree of order (Lewis, 1928). The analysis of epithelial organization has been mainly based on the number of neighbours of the epithelial cells, considering the apical surface of these cells as convex polygons with the same number of sides as neighbours (Blankenship et al., 2006; Classen et al., 2005; Escudero et al., 2011; Farhadifar et al., 2007; Gibson et al., 2006, 2011; Heller et al., 2016; Kram et al., 2010; Lewis, 1928; Mao et al., 2011; Nagpal et al., 2008; Patel et al., 2009; Sanchez-Gutierrez et al., 2016; Zallen and Zallen, 2004). In previous works we have investigated several aspects of the organization of packed tissues using Voronoi tessellations to compare the polygon distributions of natural and mathematical tessellations (Sanchez-Gutierrez et al., 2016). We have described that the polygon distribution of natural tessellations is restricted to a series of frequencies of polygons that match the Voronoi diagrams that conform the Centroidal Voronoi tessellation (CVT). This is what we call a

"CVT path" and was used as a scale to compare the organization of different packed tissues. However, polygon distribution is not sufficient to completely characterize tissue organization. Tissues with a clearly different appearance can present very similar polygon distribution (Sanchez-Gutierrez et al., 2016).

As an alternative approach, we have proposed that Graph Theory could capture differences in the topology of tissues (Escudero et al., 2011; Sanchez-Gutierrez et al., 2013; Sánchez-Gutiérrez et al., 2017). This is based on the idea of converting the epithelium in a network of cell-to-cell contacts (Escudero et al., 2011). The resulting "epithelial graph" can be analysed by combining the tools of network theory and multivariable statistical analysis (Escudero et al., 2011; Kursawe et al., 2016; Sanchez-Gutierrez et al., 2013; Yamashita and Michiue, 2014). Finding features and patterns that can describe the graphs is key for very diverse fields including biology (Benson et al., 2016; Costa et al., 2007; Hayes et al., 2013). A network can be split up into different subgraphs named graphlets. The graphlets composition of a network has been used to quantify differences between complex systems (Hayes et al., 2013; Ho et al., 2010; Kuchaiev et al., 2011; Pržulj et al., 2004). These measurements are based in on the comparison of the number of these subgraphs in different networks and were able to provide an index of distance between them. This feature has the advantage of integrating in a single value the differences between diverse networks, simplifying the analyses and allowing multiple comparisons.

In summary, there is a clear need for a method to specifically quantify organization and aid the interpretation of biophysical and mechanical aspects of morphogenesis (Blankenship et al., 2006; Canela-Xandri et al., 2011; Farhadifar et al., 2007; Gibson et al., 2006, 2011; Heisenberg and Bellaiche, 2013; Heller et al., 2016; Lecuit and Lenne, 2007; Mao et al., 2013; Sanchez-Gutierrez et al., 2016; Zallen and Zallen, 2004). The advances in imaging techniques, together with the appearance of powerful methods for automated image analysis (Heller et al., 2016; Khan et al.,

2014; Kursawe et al., 2016) and new simulation resources (Blanchard et al., 2009; Etournay et al., 2016; Fletcher et al., 2014; Guirao et al., 2015; Mirams et al., 2013; Tanaka et al., 2015) provide a large amount of source data that require analysis in terms of organization. Here we present the open source platform EpiGraph, a new image analysis method that uses segmented images from real epithelia or simulations to easily quantify and compare the organization of packed tissues.

RESULTS

Graphlet measurements as an approach to capture organization of packed tissues.

In previous studies, a set of 29 graphlets has been used to distinguish between different types of networks (Pržulj et al., 2004) (**Fig. S1**). This method calculated the Graphlet degree Distribution agreement Distance (GDD) between two networks (Pržulj, 2007). Epithelial images can be seen as natural tessellations and converted into networks of cell-to-cell contacts (Escudero et al., 2011). We have used the "graphlet" approach to capture the topology of epithelial tissues. Tessellations give rise to "geographic networks" (Albert and Barabasi, 2002) that only makes sense in a planar surface. For this reason, when we translated the set of graphlets to cellular patterns, some of them were redundant (G5 and G27) or not possible (G20, G22, G25). Therefore, in this study we have used a total of 26 graphlets corresponding to 29 different cellular motifs that account for the organization of groups of up to 5 cells (**Fig. 1A, Fig. S1**). Most of the analyses performed in this work were completed with only 17 motifs (Mo17, **Fig. 1A, mauve**) since we found that excluding the motifs capturing three, four and more than seven sided cells, the graphlets comparisons were less dependent on the polygon distributions of the samples. However, it would be possible to use other combinations such as all the motifs (Mo29) or cellular motifs that account for the organization of groups of up to 4 cells (Mo10) (**Fig. 1A**).

Graphlet measurements capture differences that polygon distributions do not.

We tested the power of graphlet-based measurements in quantifying differences between sets of images with very similar polygon distributions (**Fig. 1B**). During the third instar larva of *Drosophila* the photoreceptors are being recruited giving rise to a particular repetitive arrangement of the presumptive eye cells (Eye, **Fig. 1C**). This arrangement is very different to the irregular distribution in a Voronoi tessellation where the initial seeds were placed in a random way (Sanchez-Gutierrez et al., 2016), (Diagram 1, **Fig. 1C**). We previously showed that it was not possible to discriminate between the polygon distributions of these two tessellations (Sanchez-Gutierrez et al., 2016), however we obtained a GDD value of 0.086 when comparing these two sets of images (Mo17, **Table S1, Fig. S2**). To set a baseline we compared other images with very similar polygon distribution that also presented, apparently, an arrangement alike. This was the case for Diagram 4 of the CVT vs. the *Drosophila* wing imaginal disc in larva (dWL) and the Diagram 5 of the CVT vs. the *Drosophila* wing imaginal disc in prepupa (dWP) (**Fig. 1B-C**). Both results were on the same range with a GDD value of 0.042 for Diagram 4 vs. dWL and 0.049 for Diagram 5 vs. dWP (**Fig. 1-C**). Similar results were obtained when comparing Diagram 4 vs. Diagram 5 and dWL vs. dWP (**Fig. 1-C**). We interpreted that there was a baseline in the range of 0.04-0.05 values that correspond to similar cellular arrangements that cannot be well distinguished using the graphlets distribution. This contrasted with the Eye vs. Diagram 1 comparison where the use of graphlets can indeed capture differences beyond the polygon distribution of the samples. The results obtained using Mo17 and Mo29 were equivalent (**Table S1**).

EpiGraph quantitatively compares the organization of multiple sets of images.

The GDD had the limitation of comparing only 2 samples each time. However, one of our objectives was to evaluate different types of images together. Therefore, we designed EpiGraph, a method that calculates the GDD of any epithelial tissue with other tessellation that serves as a reference. As an output we obtained an index of distance for each image

with respect to this reference. We used three different references: i) a tessellation formed by regular hexagons, representing the most ordered way to pave the space (**Fig. 2A**, GDDH). ii) the network motifs emerging from a random Voronoi tessellation (**Fig. 2B**, GDDRV). iii) a Voronoi Diagram 5 from the CVT path (**Fig. 2C**, GDDV5) that present a polygon distribution similar to the one from multiple examples in nature (Gibson et al., 2006; Sanchez-Gutierrez et al., 2016).

To have a scale and facilitate fast comparisons we used the concept of the CVT path (Sanchez-Gutierrez et al., 2016). We calculated the values for GDDH, GDDRV and GDDV5 for all the Voronoi diagrams and used this property to visualize the results and plot the GDDH, GDDRV and GDDV5 with respect to the percentage of hexagons of the corresponding diagram (**Fig. S3**). The CVT does not progress beyond the 70% of hexagons limiting the possibilities of analysis. We extended the Voronoi scale spanning a wider range of polygon distributions. The algorithm that devises the CVT was modified introducing “noise” in the positioning of the seed that produce the subsequent diagram. In this way, we obtained a “CVT noise” (CVTn) whose last diagrams reached 90% of hexagons (**Fig. 2D-G** and **Experimental Procedures**). We decided to compare GDDH, GDDRV and GDDV5 values with the percentage of hexagons since it is indicative of the proportion of the other types of polygons along the CVTn (**Table S2**). Interestingly, the plot obtained using CVT and CVTn diagrams was an optimum way to easily visualize these geometric scales as continuous “CVT path” and “CVTn path”. Therefore, we used this framework to analyse the values of GDDH, GDDRV and GDDV5 for each diagram in the scale. As expected, the GDDH values were higher in the initial diagrams and were progressively decreasing with the increase in the percentage of hexagons of the Voronoi diagrams (**Fig. 2D**). The opposite happened in the case of the GDDRV values (**Fig. 2E**). In the plot of percentage of hexagons vs GDDV5, the CVTn path presented the shape of a walking stick. The GDDV5 values of Voronoi Diagrams 1, 2, 3, and 4 were decreasing progressively, being the

Diagram 5 the closest to the zero value. The values for the rest of the diagrams were gradually increasing similarly to the case of the GDDRV.

We tested the method with epithelial images that have been previously shown to present polygon distributions that match the "CVT path": chicken neural tube (cNT), dWL, dWP and Eye (Sanchez-Gutierrez et al., 2016). We calculated the GDDH, GDDRV and GDDV5 of these images and compared them with the CVTn (**Fig. 2D-G**). We found that for cNT, dWL and dWP the GDDH, GDDRV and GDDV5 values were similar to the CVTn at the same percentage of hexagons of the polygon distribution. In agreement with our previous results using the GDD, the Eye images presented a higher GDDRV and GDDV5 values than the expected for a 30% of hexagons (**Fig. 2E-G**). Therefore, these data points appeared displaced from the CVTn. This result suggested that EpiGraph was able to distinguish between different tessellations with a similar polygon distribution.

Quantification of organization in altered tissues and simulations.

We further investigated the possible applications of EpiGraph and performed a series of experiments aiming to understand what traits of the tissue organization are being captured and quantified by the graphlet measurements. First, we used the basis for the formation of the CVT path but including perturbations in subsets of cells. In a weighted Voronoi tessellation it is possible to define different distance functions for the formation of the Voronoi cells. In contrast to the diagrams from the CVT or CVTn (where all the points of a Voronoi cell are closer to the seed that has generated it) in the case of a weighted Voronoi tessellation the size of a cell is going to be proportional to the weight defined for its seed and the surrounding seeds. The construction of a CVT weighted (CVTw) allowed the study of the effects in the organization of the appearance of cells with an increased weight (**Experimental Procedures**). We placed the weighted cell into a circle in the middle of the tessellation (**Fig. 3A**). Then, we calculated how the interaction of the positive weighted cells with the surrounding control cells altered the organization of the tissue for every step (4-80

iterations). We plotted the values for the positive weighted cells and the control neighbour separately, but taking into account the percentage of hexagons of the whole image (**Experimental Procedures** and **Fig. 3B, C**). GDDH helped to distinguish the organization traits between both types of cells. While positive weighted cells were changing as the CVTn diagrams, in the case of the control neighbours there was an increment of the percentage of hexagons for the same GDDH value (**Fig. 3B**). The use of GDDRV and GDDV5 showed clear differences between both types of cells and the CVTn path (**Fig. 3B and Fig. S4**). Altogether, our results suggest that GDDH, GDDRV and GDDV5 can capture different organizational traits making EpiGraph flexible to analyse and compare diverse tissues or groups of cells.

In a second approach we used images of different vertex model simulations of proliferating tissues with normal and altered parameters that change the epithelial organization (Sanchez-Gutierrez et al., 2016), **Experimental Procedures** and **Fig. 3D-F**). These alterations mimic the effect of the reduction of myosin II in the *Drosophila* third instar wing disc epithelium: dMWP (**Fig. 3G**). In the control simulation, cells grow to double the original area and then divide into two cells. In the case 3 and case 4 simulations there was a random reduction of the tension parameter together with a requirement of a minimum tension threshold to be able to divide. If the cells do not reach this threshold they continue to grow, but are not able to divide the cell body. When this happen, the cells will be stuck in mitotic phase and will not start a second round of cell division. This was described to be similar to what happened in the dMWP samples when examining the polygon distributions (Sanchez-Gutierrez et al., 2016). The control simulation gave similar values to the CVTn, while case 3, case 4 and dMWP images presented a clear deviation in the GDDRV plot (**Fig. 3H**) and the GDDRV vs GDDV5 graph (**Fig. 3I**). Interestingly, these data-points distributed in the same region of the graph. These differences did not appear in the GDDH one (**Fig. S4**). These data suggested that GDD calculation was a useful method to identify differences in organization due to changes in cell ideal volume.

The different results comparing GDDH, GDDRV and GDDV5 values also suggested that GDDH had better resolution for images with a higher percentage of hexagons while GDDRV and GDDV5 were more sensitive to the differences between images with less than 40% of hexagons.

EpiGraph: a method to capture epithelial organization implemented in FIJI.

Aiming to enhance the accessibility to the biology community for the analysis of tissue organization we have designed EpiGraph, a plugin for FIJI (Schindelin et al., 2012). EpiGraph consists in a pipeline of 5 very simple steps. First, the skeleton of an epithelial image is uploaded and the individual cells are identified. Second, there is a step where the user selects the distance threshold to identify two cells as neighbours. Here it is possible to select different thresholds and to check the number of neighbours of every cell in each case. Third, a ROI is selected. There are several possibilities such as a default ROI from the image or the selection of individual cells. Fourth, the graphlet information for the selected cells is calculated. These data are used to obtain the GDDH, GDDRV and GDDV5. These values are incorporated to a table. The fifth step includes the classification and labelling of different images in order to represent them in a new window. This final phase allows exporting the representation of the data in a three-dimensional graph. **Movie S1** shows an example of EpiGraph usage. A detailed description of EpiGraph can be found in the **Supplemental Experimental Procedures** section. A full set of tutorials explaining how to install and use EpiGraph is available at EpiGraph's wiki (https://imagej.net/EpiGraph).

As a proof of principle of the usability of EpiGraph we used images obtained from the literature, not obtained by our group and not directly related with our work. We chose a tissue where the organization changes are clear during time: the appearance of rosettes during germ band extension of the embryonic development of *Drosophila* (Blankenship et al., 2006). We selected eight representative time points from "Movie S1" of this

article and quantified the graphlets content (**Fig. 4A**). After segmentation (**Experimental Procedures**), the images were uploaded to EpiGraph and plotted in the 3D graph (**Fig. 4B-C**). We found that the early frames from the original movie (001, 035, 070) were located close to the CVT. Progressively, the appearance of rosettes (100, 135) correlated with the separation of these samples from the CVT. This distance was greater in the case of the image corresponding with the frame 170. In contrast, the last two frames did not increase the distance with the CVTn. We observed a reversion in the measurement of the organization, with the image 235 located very close to the time point 135. Interestingly, this was captured using different combinations of the four measured parameters (percentage of hexagons, GDDH, GDDRV and GDDV5, **Fig. 4B-C**). The ability of compare different steps of a developmental process suggested that EpiGraph was a versatile method able to capture and analyse the dynamic changes in organization of developmental processes.

DISCUSSION

Textbook definitions of morphogenesis include the term “organization” as key to explain this fundamental developmental process (Dai and Gilbert, 1991). The authors wonder, “How can matter organize itself so as to create a complex structure such as a limb or an eye?” Here we have provided a tool that can help the developmental biology community to solve that question.

The analysis of the polygon sides of the epithelial cells has been shown to be insufficient to completely understand tissue organization. Some tessellations can present very different arrangements and the same frequencies of number of neighbours. A second problem is the lack of a simple value as an indicator of epithelial organization. This feature complicates the analysis of the morphogenesis of normal development and after genetics alterations. Our previous attempts to go beyond this caveat were based on multi-statistical analyses of graph features (Sanchez-Gutierrez et al., 2013) and the creation of a Voronoi scale to statistically

compare groups of images with the CVT reference (Sánchez-Gutiérrez et al., 2016). However, we are aware that these methods are difficult to incorporate into the average developmental biology lab.

We have developed EpiGraph aiming to solve both questions. EpiGraph transforms the image into a graph of cell-to-cell contacts and extracts their graphlet content to later compare with other images. These complex algorithms are hidden behind the friendly user window of FIJI. This is the most popular open-source biological image analysis platform. In addition, the output data options of EpiGraph facilitate a fast and clear representation and interpretation of the results. Therefore, our method can be easily used in any developmental biology lab.

One of the strengths of EpiGraph is the comparison of any tessellation with the hexagonal lattice, the "random" Voronoi tessellation and the Voronoi tessellation that presents the "conserved polygon distribution" (Gibson et al., 2006; Sánchez-Gutiérrez et al., 2016) (**Fig. 2A-C**). We have tested EpiGraph with different types of samples: as expected, the average of the natural tessellations such as dWL, dWP and cNT matched the CVTn path position (**Fig. 2D-G**). We interpret that these three natural samples present similar polygon distribution and graphlet composition that some Voronoi Diagrams from the CVTn. This was independent of the combination of GDD values plotted. On the other hand, the average of the Eye samples appeared far from the CVTn when GDDV5 or GDDRV values were plotted. These two GDD were capturing differences in organization between the Eye and any Voronoi Diagram (including Diagram 1, the one that present a similar polygon distribution than the Eye). This result supports the utility of EpiGraph to quantify organizational traits that were not accessible until now. The same idea is reinforced by the results obtained with the mutant samples for myosin II (dMWP) and the two vertex simulations that try to reproduce the same phenomena by increasing the value of the "ideal area" term in some cells (case III and case IV, **Fig. 3F-G** and **Experimental Procedures**). In a previous work we showed that these three set of samples slightly deviated the CVT scale in terms of polygon distribution (Sánchez-

Gutiérrez et al., 2016). Here we show very clear differences in terms of the values of GDDV5 and GDDRV (**Fig. 3F-G**), suggesting a higher sensitivity of the new method when capturing differences in organization.

Eye or myosin II simulations appear more different when plot GDDV5 and GDDRV values. But differences between two populations of CVTw cells are clearer using GDDH (**Fig. 3C-D**). We have found that depending on the samples some GDD work better than others to find differences. We think that this is related to the total content of hexagons in the images, being GDDH more appropriate when it is higher and GDDRV and GDDV5 better when it is lower. For this reason, we have designed the visualization step of the program to easily change the three axes and check the different results using any combination of GDDs and the “percentage of hexagons”.

As a proof of principle of the capabilities of our method we have chosen a previously published movie to analyse the changes of organization during morphogenesis. We have chosen the emergence and resolution of rosettes during the germ band extension since it is a well-known event with a dynamic and complex arrangement of cells (Blankenship et al., 2006). The selected time lapses included the formation and resolution of the rosettes. The GDD values of these images correlate with the developmental process and the 3D visualization tool of EpiGraph captures the complete morphogenetic event (**Fig. 4B-C**). The selected frames 001, 035 and 070, appear near the CVTn diagrams since they share a similar composition of motifs. This starts to change with the appearance and increase of the rosettes (frames 100, 135 and 170) that progressively separate them from the CVTn. The rosettes contain motifs that illustrate this developmental biology process and they are captured by the graphlet analysis. Indeed, the two last frames (200 and 235) show that this tendency is being reversed, going back closer to the CVTn. We think that this is reflecting the return to an epithelial organization without so many motifs formed by groups of cells contacting in a vertex: the resolution of rosettes that allows tissue elongation.

EpiGraph limitations.

Although Epigraph accepts a wide range of images as inputs, we have specified some minimum requirements. It does not accept images bigger than 3000 pixels of width or 3000 pixels of height, since process them could be computationally intensive. In addition, EpiGraph only accepts single images. Stack of images should be adapted to single frames before uploading it to EpiGraph.

Computers with little RAM memory (less than 16gb) will work but with a series of restrictions. For ensuring the usability, it is not recommended computing images with a high number of cells (more than 1000) due to a possible lack of memory. In the same way, we suggest skeletonizing the edges of the images and using a small radius (lower than 10 pixels) to calculate the cells neighbourhood. Choosing a high radius value could slow down the work queue, increasing the use of RAM memory.

If any of these requirements are not satisfied, the program alert the user allowing him/her to change the image provided. Importantly, the images and the ROIs require a minimum number of cells in order to get coherent graphlets. Therefore, to get any result, EpiGraph must detect at least a 4' valid cell in the case of Mo7 or Mo10 or a 5' valid cell in the case of Mo17 and Mo29. In any case, we strongly recommend having a greater number of 4' and 5' valid cells to get results that can be trusted in terms of capturing the organization of a tissue. Regarding the 3D visualization tool, it allows the user to see the position of the samples from different angles. However, the resolution of the exported file is only 72 pixels per inch (dpi). This could be too low for publications and therefore EpiGraph provides an excel table with all the information needed to represent it with other programs.

In summary, we have generated a very accessible open source method to produce a quantitative description of epithelial tissue arrangements. We anticipate that this tool will improve the study of tissue morphogenesis by incorporating the epithelial organization as result of the effect of mutations or diseases.

EXPERIMENTAL PROCEDURES

Source images used in the study.

Centroidal Voronoi Tessellation (CVT) diagrams and variations

For the generation of this set of paths we have used the software Matlab R2014b to iteratively apply the Lloyd´s algorithm to a random Voronoi tessellation (Lloyd, 1982). This implies that the centroid of a cell in a Voronoi diagram is the seed for the same cell in the next iteration.

- Centroidal Voronoi Tessellation (CVT) diagrams

Centroidal Voronoi Tessellation diagrams were obtained as described previously by our group (Sanchez-Gutierrez et al., 2016). The original 20 original Voronoi diagrams were created placing 500 seed randomly in an image of 1024x1024 pixels. A total of 700 iterations were generated for each initial image.

- Centroidal Voronoi Tessellation noise (CVTn) diagrams

We have developed a variation of CVT path, named CVT noise path. We started from the same 20 initial random diagrams described above. The development process of CVTn path was modified so the new seeds were not strictly the centroid regions. In even iterations, we selected a region of 5 pixels of radius from the centroid position where seeds could be placed randomly. In odd iterations the system was stabilized, applying the original Lloyd algorithm. A total of 700 iterations were generated for each initial image.

- Centroidal Voronoi Tessellation weighted (CVTw) path

We computed a variant of the generalized CVT path, in which, a subset of seeds presented a weight higher than the rest. We used 4 images with a size of 2048x2048 pixels and with 2000 seeds located randomly over these images. A total of 80 iterations were performed from the initial seeds. We applied the same algorithm used for the CVT path generation until the iteration number 3. After this step, we made use of an external function of Matlab: *powerDiagramWrapper* (McCollum, 2014). This function took the

position of the seeds and their weight and returned the cells edges of the weighted Voronoi Diagram. In this way, we calculated the centroids position of the Voronoi cells as seeds and applied the *powerDiagramWrapper* function for each iteration. So, we have replaced the generation of the conventional Voronoi diagram as dictate the Lloyd algorithm, by the weighted Voronoi Diagram generation.

From the initial Voronoi diagram, all the cells located in a radius of 300 pixels from the centre point in the image were considered as weighted cells. The weighted cells were endowed with a weight of 2000, and the rest with a null weight. This weight (w) expressed the capability of the seed (p) to influence its neighbourhood, being then the corresponding cell larger. Thus, nearby points (x) are "closer" to its seed with respect the power of the point: $pow(x,p) = |x-p|^2 - w(p)$.

Natural packed tissues and vertex model simulations

The details of the obtaining and processing of the epithelial images were described in (Escudero et al., 2011). Control vertex model simulation includes cell proliferation and is the base for the other two cases. Case III corresponds to a vertex model simulation with heterogeneous reduction of line tension and an impairment of the cell division when tension value is under the 30 percentage of the initial value. Case IV is a similar simulation than Case III with a threshold of 40 percent. The exact conditions for the vertex model simulations were described in (Sanchez-Gutierrez et al., 2016).

Multicellular Rosette Formation segmentation

We have segmented a set of images taken from a movie that represents the evolution of the multicellular rosette formation. This is "Movie S1" from Blankenship and cols., 2006 article (url: https://ars.els-cdn.com/content/image/1-s2.0-S153458070600400X-mmc2.mov) (Blankenship et al., 2006). We do not show these panels in this version of the manuscript since we have not been able to obtain the editorial permission yet. The segmentation procedure consisted in the following

steps: 1) We used Trainable Weka Segmentation (Arganda-Carreras et al., 2017) a plugin developed for FIJI (Yoon et al., 1995). This plugin uses machine learning to correctly segment images. We trained the program with a representative image of the whole set, classifying two categories: the outlines and the body of the cells. After the training step, we applied the plugin classifier to all the chosen images from the film, getting a first set of segmented images. 2) Manual curation: This step was necessary to reflect rigorously the real cell outlines in the segmented images, and discard the zones with bad visibility or artefacts. The process was carried out by using Adobe Photoshop CS6. 3) Image processing: We developed a script in Matlab R2014b to clean the noise and the segmentation imperfections. We used several morphological operations for this purpose. Finally, we fixed the cells outlines with a width of 3 pixels to preserve the nature of the image.

Graphlets and motifs selection.

The different images of the previous section were used to create a graph of cell-to-cell contacts (Escudero et al., 2011) and **Supplemental Experimental Procedures**) that served as the source for the graphlet analysis (Pržulj, 2007; Przulj et al., 2004). First, we adapted the graphlet analysis performed by EpiGraph to the nature of our samples (tessellations). Three graphlets were discarded since they were not possible in the context of an epithelial tissue (**Fig. 1** and **Fig. S1**). Second, we used the computer program for graphlet identification and calculation ORCA (Orbit Counting Algorithm) (Hocevar and Demsar, 2014), to extract the different conformations of nodes assembling the graphlets, called orbits (Pržulj, 2007). We computed the Graphlet degree Distribution of the 73 given orbits from the 29 graphlets, and then we removed the non-used ones.

REFERENCES

Albert, R., Barabasi, A.-L., 2002. Statistical mechanics of complex networks. Rev. Mod. Phys. 74, 47–97.

Arganda-Carreras, I., Kaynig, V., Rueden, C., Eliceiri, K.W., Schindelin, J., Cardona, A., Seung, H.S., 2017. Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. Bioinformatics. doi:10.1093/bioinformatics/btx180

Benson, A.R., Gleich, D.F., Leskovec, J., 2016. Higher-order organization of complex networks. Science (80-.). 353, 163–166. doi:10.1126/science.aad9029

Bertet, C., Sulak, L., Lecuit, T., 2004. Myosin-dependent junction remodelling controls planar cell intercalation and axis elongation. Nature 429, 667–671.

Blanchard, G.B., Kabla, A.J., Schultz, N.L., Butler, L.C., Sanson, B., Gorfinkiel, N., Mahadevan, L., Adams, R.J., 2009. Tissue tectonics: morphogenetic strain rates, cell shape change and intercalation. Nat Methods 6, 458–464. doi:nmeth.1327 [pii]10.1038/nmeth.1327

Blankenship, J.T., Backovic, S.T., Sanny, J.S., Weitz, O., Zallen, J.A., 2006. Multicellular rosette formation links planar cell polarity to tissue morphogenesis. Dev Cell 11, 459–470.

Canela-Xandri, O., Sagués, F., Casademunt, J., Buceta, J., 2011. Dynamics and mechanical stability of the developing dorsoventral organizer of the wing imaginal disc. PLoS Comput. Biol. 7. doi:10.1371/journal.pcbi.1002153

Classen, A.K., Anderson, K.I., Marois, E., Eaton, S., 2005. Hexagonal packing of Drosophila wing epithelial cells by the planar cell polarity pathway. Dev Cell 9, 805–817.

Costa, L.D.A.F., Rodrigues, F.A., Travieso, G., Boas, P.R. V, 2007. Characterization of complex networks: a survey of measurements. Adv. Phys. 56, 167–242.

Dai, J.D., Gilbert, L.I., 1991. Metamorphosis of the corpus allatum and degeneration of the prothoracic glands during the larval-pupal-adult transformation of Drosophila melanogaster: a cytophysiological analysis of the ring gland. Dev Biol 144, 309–326.

Escudero, L.M., Bischoff, M., Freeman, M., 2007. Myosin II regulates complex cellular arrangement and epithelial architecture in Drosophila. Dev Cell 13, 717–729.

Escudero, L.M., Costa Lda, F., Kicheva, A., Briscoe, J., Freeman, M., Babu, M.M., 2011. Epithelial organisation revealed by a network of cellular contacts. Nat Commun 2, 526. doi:10.1038/ncomms1536ncomms1536 [pii]

Etournay, R., Merkel, M., Popović, M., Brandl, H., Dye, N.A., Aigouy, B., Salbreux, G., Eaton, S., Jülicher, F., 2016. TissueMiner: A multiscale analysis toolkit to quantify how cellular processes create tissue dynamics. Elife 5, 773–786. doi:10.7554/eLife.14334

Farhadifar, R., Roper, J.C., Aigouy, B., Eaton, S., Julicher, F., 2007. The influence of cell mechanics, cell-cell interactions, and proliferation on epithelial packing. Curr Biol 17, 2095–2104.

Fletcher, A.G., Osterfield, M., Baker, R.E., Shvartsman, S.Y., 2014. Vertex models of epithelial morphogenesis. Biophys. J. 106, 2291–2304. doi:10.1016/j.bpj.2013.11.4498

Gibson, M.C., Patel, A.B., Nagpal, R., Perrimon, N., 2006. The emergence of geometric order in proliferating metazoan epithelia. Nature 442, 1038–

1041.

Gibson, W.T., Veldhuis, J.H., Rubinstein, B., Cartwright, H.N., Perrimon, N., Brodland, G.W., Nagpal, R., Gibson, M.C., 2011. Control of the mitotic cleavage plane by local epithelial topology. Cell 144, 427–438. doi:10.1016/j.cell.2010.12.035S0092-8674(10)01527-8 [pii]

Girdler, G.C., Roper, K., 2014. Controlling cell shape changes during salivary gland tube formation in Drosophila. Semin Cell Dev Biol 31, 74–81. doi:10.1016/j.semcdb.2014.03.020

Guirao, B., Rigaud, S.U., Bosveld, F., Bailles, A., López-Gay, J., Ishihara, S., Sugimura, K., Graner, F., Bellaïche, Y., 2015. Unified quantitative characterization of epithelial tissue development. Elife 4, 773–786. doi:10.7554/eLife.08519

Hayes, W., Sun, K., Pržulj, N., 2013. Graphlet-based measures are suitable for biological network comparison. Bioinformatics 29, 483–491. doi:10.1093/bioinformatics/bts729

Heisenberg, C.P., Bellaiche, Y., 2013. Forces in tissue morphogenesis and patterning. Cell 153, 948–962. doi:10.1016/j.cell.2013.05.008S0092-8674(13)00573-4 [pii]

Heller, D., Hoppe, A., Restrepo, S., Gatti, L., Tournier, A.L., Tapon, N., Basler, K., Mao, Y., 2016. EpiTools: An Open-Source Image Analysis Toolkit for Quantifying Epithelial Growth Dynamics. Dev Cell 36, 103–116. doi:S1534-5807(15)00797-2 [pii]10.1016/j.devcel.2015.12.012

Ho, H., Milenković, T., Memisević, V., Aruri, J., Przulj, N., Ganesan, A.K., 2010. Protein interaction network topology uncovers melanogenesis regulatory network components within functional genomics datasets. BMC Syst. Biol. 4, 84. doi:10.1186/1752-0509-4-84

Hocevar, T., Demsar, J., 2014. A combinatorial approach to graphlet counting. Bioinformatics 30, 559–565. doi:10.1093/bioinformatics/btt717

Khan, Z., Wang, Y.-C., Wieschaus, E.F., Kaschube, M., 2014. Quantitative 4D analyses of epithelial folding during Drosophila gastrulation. Development 141, 2895–2900. doi:10.1242/dev.107730

Kram, Y.A., Mantey, S., Corbo, J.C., 2010. Avian cone photoreceptors tile the retina as five independent, self-organizing mosaics. PLoS One 5, e8992. doi:10.1371/journal.pone.0008992

Kuchaiev, O., Stevanovi, A., Hayes, W., 2011. GraphCrunch 2 : Software tool for network modeling , alignment and clustering. doi:10.1186/1471-2105-12-24

Kursawe, J., Bardenet, R., Zartman, J.J., Baker, R.E., Fletcher, A.G., 2016. Robust cell tracking in epithelial tissues through identification of maximum common subgraphs. bioRxiv. doi:10.1101/049551

Lecuit, T., Le Goff, L., 2007. Orchestrating size and shape during morphogenesis. Nature 450, 189–192.

Lecuit, T., Lenne, P.F., 2007. Cell surface mechanics and the control of cell shape, tissue patterns and morphogenesis. Nat Rev Mol Cell Biol 8, 633–644.

Lewis, F.T., 1928. The correlation between cell division and the shapes and sizes of prismatic cells in the epidermis of cucumis. Anatom. Rec. 38, 341–376.

Lloyd, S., 1982. Least square quantization in PCM. IEEE Trans. Inform. Theory 28, 129–137.

Mao, Y., Tournier, A.L., Bates, P.A., Gale, J.E., Tapon, N., Thompson, B.J., 2011. Planar polarization of the atypical myosin Dachs orients cell divisions in Drosophila. Genes Dev 25, 131–136. doi:10.1101/gad.61051125/2/131 [pii]

Mao, Y., Tournier, A.L., Hoppe, A., Kester, L., Thompson, B.J., Tapon, N., 2013. Differential proliferation rates generate patterns of mechanical tension that orient tissue growth. EMBO J 32, 2790–2803. doi:10.1038/emboj.2013.197emboj2013197 [pii]

McCollum, F., 2014. Power Diagrams - File Exchange - MATLAB Central [WWW Document]. URL https://es.mathworks.com/matlabcentral/fileexchange/44385-power-diagrams (accessed 3.10.16).

Mirams, G.R., Arthurs, C.J., Bernabeu, M.O., Bordas, R., Cooper, J., Corrias, A., Davit, Y., Dunn, S.J., Fletcher, A.G., Harvey, D.G., Marsh, M.E., Osborne, J.M., Pathmanathan, P., Pitt-Francis, J., Southern, J., Zemzemi, N., Gavaghan, D.J., 2013. Chaste: an open source C++ library for computational physiology and biology. PLoS Comput Biol 9, e1002970. doi:10.1371/journal.pcbi.1002970PCOMPBIOL-D-12-01337 [pii]

Nagpal, R., Patel, A., Gibson, M.C., 2008. Epithelial topology. BioEssays 30, 260–266. doi:10.1002/bies.20722

Patel, A.B., Gibson, W.T., Gibson, M.C., Nagpal, R., 2009. Modeling and inferring cleavage patterns in proliferating epithelia. PLoS Comput Biol 5, e1000412. doi:10.1371/journal.pcbi.1000412

Pilot, F., Lecuit, T., 2005. Compartmentalized morphogenesis in epithelia: from cell to tissue shape. Dev Dyn 232, 685–694.

Pržulj, N., 2007. Biological network comparison using graphlet degree distribution. Bioinformatics 23, 177–183. doi:10.1093/bioinformatics/btl301

Przulj, N., Corneil, D.G., Jurisica, I., 2004. Modeling interactome: scale-free or geometric? Bioinformatics 20, 3508–3515. doi:10.1093/bioinformatics/bth436

Pržulj, N., Corneil, D.G., Jurisica, I., 2004. Modeling interactome : scale-free or geometric ? 20, 3508–3515. doi:10.1093/bioinformatics/bth436

Sánchez-Gutiérrez, D., Sáez, A., Gómez-Gálvez, P., Paradas, C., Escudero, L.M., 2017. Rules of tissue packing involving different cell types: human muscle organization. Sci. Rep. 7, 40444. doi:10.1038/srep40444

Sanchez-Gutierrez, D., Saez, A., Pascual, A., Escudero, L.M., 2013. Topological progression in proliferating epithelia is driven by a unique variation in polygon distribution. PLoS One 8, e79227. doi:10.1371/journal.pone.0079227PONE-D-13-27658 [pii]

Sanchez-Gutierrez, D., Tozluoglu, M., Barry, J.D., Pascual, A., Mao, Y.,

Escudero, L.M., 2016. Fundamental physical cellular constraints drive self-organization of tissues. EMBO J 35, 77–88. doi:10.15252/embj.201592374embj.201592374 [pii]

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A., 2012. Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682. doi:10.1038/nmeth.2019

Tanaka, S., Sichau, D., Iber, D., 2015. LBIBCell: a cell-based simulation environment for morphogenetic problems. Bioinformatics 31, 2340–7. doi:10.1093/bioinformatics/btv147

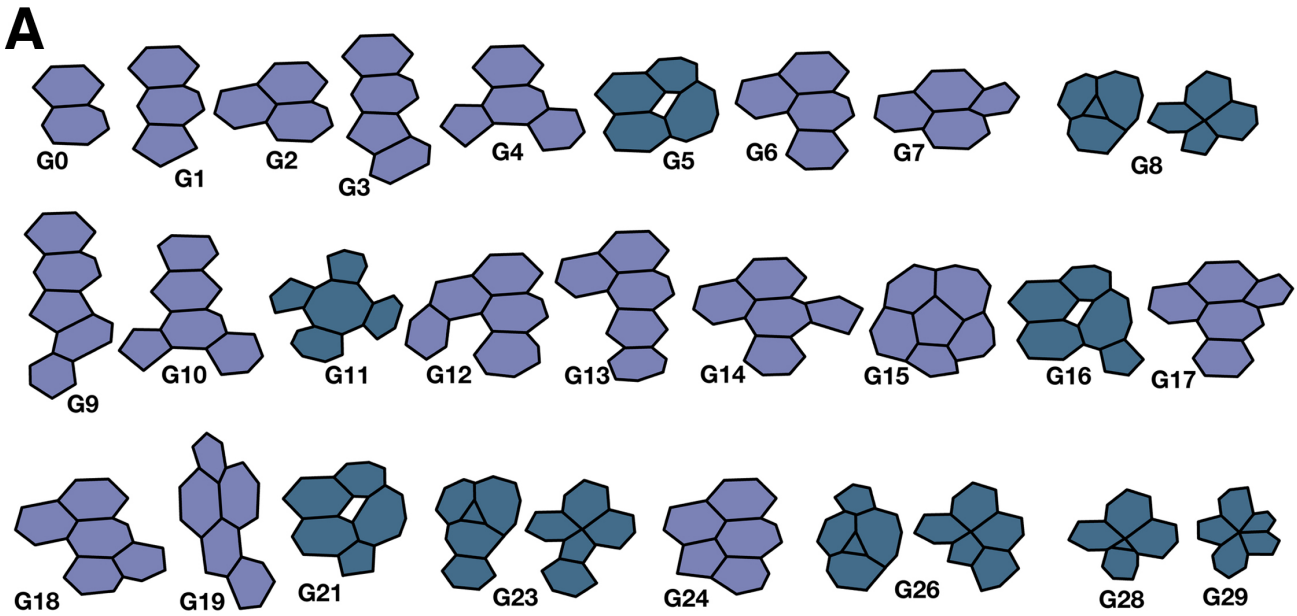
Yamashita, S., Michiue, T., 2014. Quantitative analysis of cell arrangement indicates early differentiation of the neural region during Xenopus gastrulation. J. Theor. Biol. 346, 1–7. doi:10.1016/j.jtbi.2013.12.029

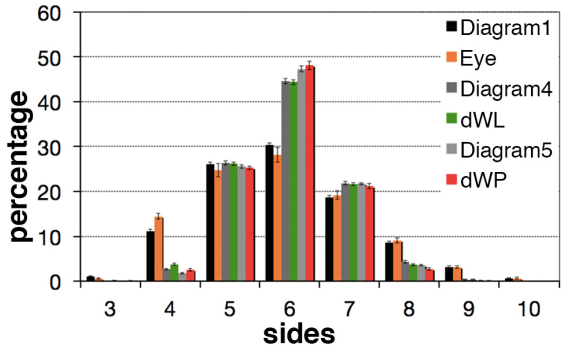
Yoon, C.H., Lee, J., Jongeward, G.D., Sternberg, P.W., 1995. Similarity of sli-1, a regulator of vulval development in C. elegans, to the mammalian proto-oncogene c-cbl. Science (80-.). 269, 1102–1105.

Yu, F.X., Guan, K.L., 2013. The Hippo pathway: regulators and regulations. Genes Dev 27, 355–371. doi:10.1101/gad.210773.11227/4/355 [pii]

Zallen, J.A., Zallen, R., 2004. Cell-pattern disordering during convergent extension in Drosophila. J. Phys. Condens. matter. 16, S5073–S5080.

Figure 1. Graphlets, cellular motifs and characterization of epithelial organization. A) A representation of the cellular motifs that correspond to graphlets of up to five nodes. There are 29 motifs corresponding to 26 different graphlets (**Fig. S1**). Note that one graphlet can represent two cellular motifs (G8, G23 and G26). Mauve motifs form the Mo17 set. Prussian Blue motifs capture three, four and more than seven sided cells. In the first row are the motifs that account for the organization of groups of up to 4 cells (Mo10). Therefore, Mo7 set is formed by the mauve coloured graphlets at the first row. **B)** Polygon distribution comparison of images from: Voronoi diagram 1 (black bar), Eye (orange), Voronoi diagram 4 (gray), dWL(green), Voronoi diagram 5 (light gray), dWP (red). Data shown refer to the mean ± SEM. Diagram 1, 4 and 5: 20 replicates. Eye: 3 samples. dWL: 15 samples. dWP: 16 samples. **C)** GDD value calculation (Mo17) between natural images and Voronoi diagrams with similar polygon distribution. The data shown are the mean of the GDD between each pair of images.





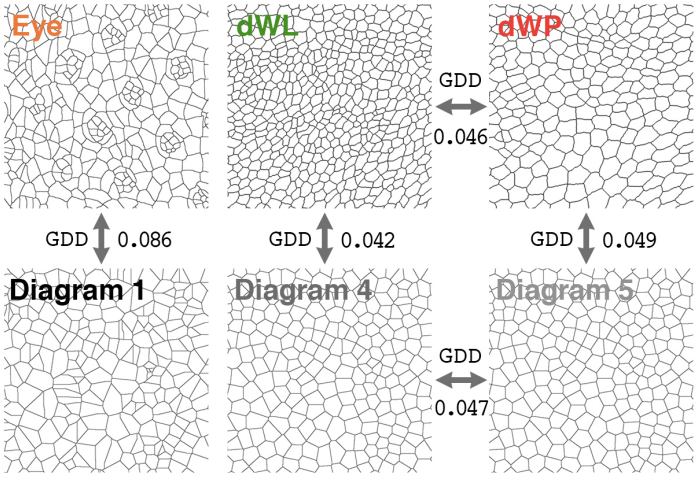


Figure 2. Epithelial organization of biological tissues with respect the CVTn. A-C) Tessellations with the corresponding graph of cell-to-cell contacts for a perfect hexagonal arrangement (**A**) a Voronoi Diagram 1 (**B**) and a Voronoi Diagram 5 (**C**) from a CVTn. These tessellations represent the diagrams used as reference to calculate the GDDH, GDDRV and GDDV5 respectively. The light blue edges in these panels represent the cellular connectivity network. The colourful nodes mark the valid cells that were involved in the cellular motifs to measure graphlets presence. The dark blue and green nodes are the 4' valid cells (cells that did not have a no valid cell within a distance of four cells connexions), which were used to calculate the graphlets for Mo10 and Mo7. The green nodes are the 5' valid cells (cells that did not have a no valid cell within a distance of five cells connexions) that were used to quantify the graphlets for Mo29 and Mo17. Cells without nodes were no valid cells for graphlet calculation. **D–G)** Plots showing the different combinations of the values for Mo17 of GDDH, GDDRV, GDDV5 and percentage hexagons. The diagrams of the CVTn path from the iteration 1 until the iteration 700 are represented as a grayscale beginning in black and reducing its darkness with the increase of the iterations. The natural tessellations are: Eye, *Drosophila* eye disc, 3 replicates, orange. cNT, chicken embryo neural tube epithelium, 16 replicates, light blue. dWL, *Drosophila* larva wing disc, 15 replicates, green. dWP, *Drosophila* prepupal wing imaginal disc epithelium, 16 replicates, red. Circumferences are individual values, circles are the average value obtained from the individual samples from each category.

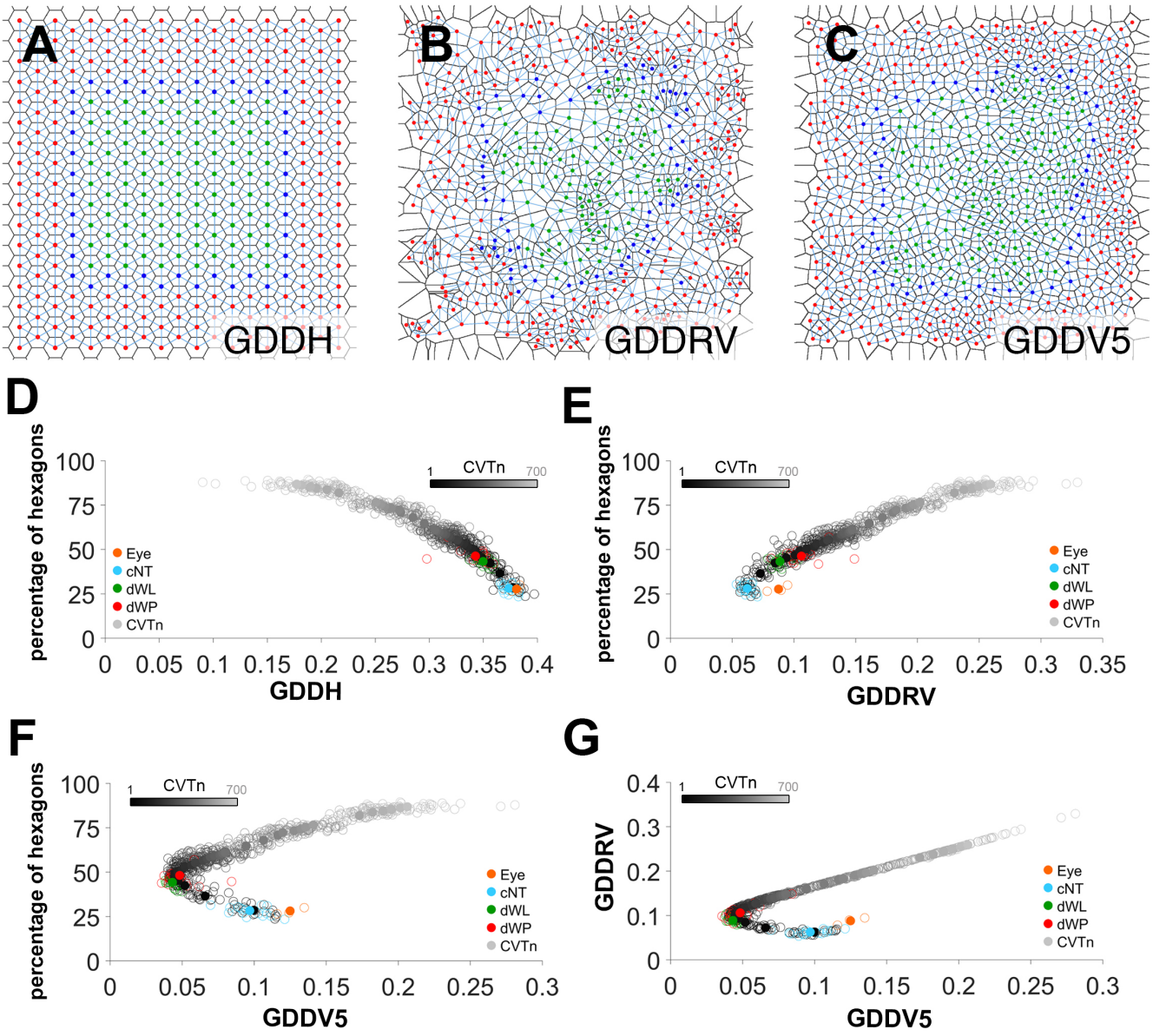


Figure 3. Comparison of different simulations and mutants with the CVTn. A) Several representative diagrams from the CVTw. Aquamarine blue cells incorporate a bigger weight than the rest of the cells in the tessellation in each iteration. The dark blue colour labels the adjacent cells to the group with higher weight. In (B), (C), (H) and (I) the diagrams of the CVTn path from the iteration 1 until the iteration 100 are represented as a grayscale beginning in black and reducing its darkness with the increase of the iterations; circumferences are individual values, circles are the average value obtained from the individual samples from each category. **B-C**) Plots showing the values of the percentage of hexagons vs GDDH (Mo17) (B) and GDDRV vs GDDV5 (C) for CVTn, Eye, cNT, dWL, dWP, CVTw1 and CVTw2. CVTw1 represent the cells with weight in (A) from iterations 4, 10, 20, 30, 40, 50, 60, 70 and 80 (increase darkness of aquamarine blue from 4 to 80). CVTw2 represent the adjacent cells to the group with higher weight in (A) from iterations 4, 10, 20, 30, 40, 50, 60, 70 and 80 (increase darkness of dark blue from 4 to 80). **D**) Cell arrangement resulting from the control simulation that includes cell proliferation **F-G**) Diagrams resulting from a vertex model simulation with an increase of the ideal area value, with respect the control, in some cells. Case III and Case IV slightly differ in the line-tension parameter conditions (see **Experimental Procedures**). **H-I**) Plots showing the values of the percentage of hexagons vs GDDRV (Mo17) (H) and GDDRV vs GDDV5 (I) for CVTn, Eye, cNT, dWL, dWP Control, Case III and Case IV.

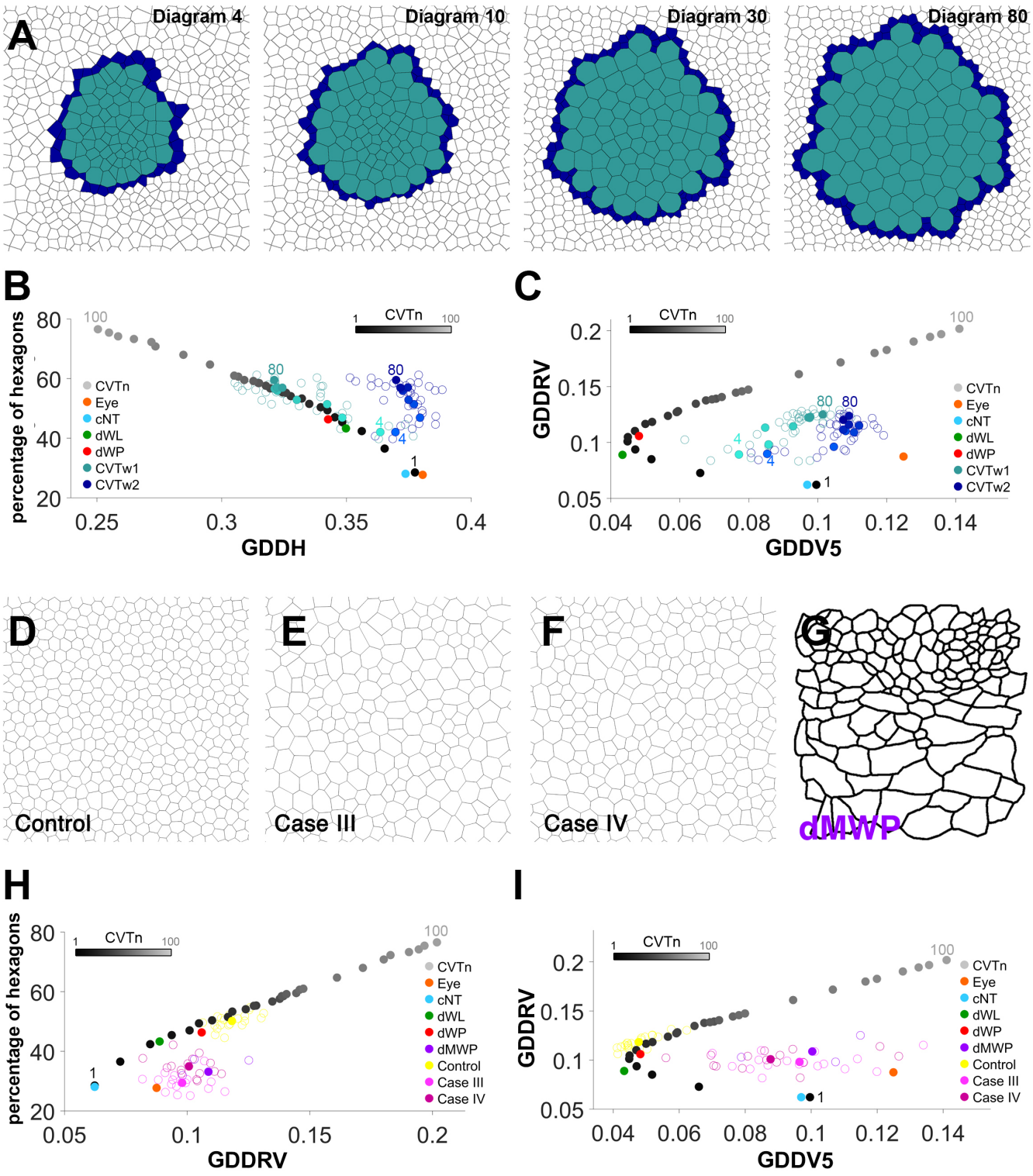


Figure 4. EpiGraph can evaluate the changes in epithelial organization of a morphogenetic process. A) Time points selected from "Movie S1" from Blankenship and cols., 2006 article (url: https://ars.els-cdn.com/content/image/1-s2.0-S153458070600400X-mmc2.mov) (Blankenship et al., 2006). We do not show these panels in this version of the manuscript since we have not been able to obtain the editorial permission yet. **B-C**) Captures from the 3D visualization tool of EpiGraph showing the position of CVTn (1-700), Eye, cNT, dWL, dWP, and the eight frames showed in (A). GDDH, GDDRV and percentage of hexagons (Mo17) are the axes in (B). GDDH, GDDRV and GDDV5 (Mo17) are the axes in (C).