Epithelial organisation revealed by a network of cellular contacts

Luis M. Escudero[1,+,*], Luciano da F. Costa[2], Anna Kicheva[3], James Briscoe[3], Matthew Freeman[1] and M. Madan Babu[1,*]

[1]MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, United Kingdom
[2]Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, SP, Brazil
[3]Developmental Neurobiology, National Institute for Medical Research, London, UK
[+]Current address: Instituto Biomedicina Sevilla (IBiS), Universidad de Sevilla/ CSIC/ Hospital Virgen del Rocío. Seville, Spain.

[*]To whom correspondence should be addressed.
Emails: MMB (madanm@mrc-lmb.cam.ac.uk) or LME (lmescudero-ibis@us.es)

Abstract

The emergence of differences in arrangement of cells is the first step towards establishment of many organs. Understanding this process is limited by the lack of systematic characterization of epithelial organisation. Here we apply network theory at the scale of individual cells to uncover patterns in cell-to-cell contacts that govern epithelial organisation. We provide an objective characterization of epithelia using network representation where cells are nodes and cell contacts are links. The features of individual cells, together with attributes of the cellular network produce a defining signature that distinguishes epithelia from different organs, species, developmental stages and genetic conditions. The approach permits characterization, quantification and classification of normal and perturbed epithelia and establishes a framework for understanding molecular mechanisms that underpin the architecture of complex tissues.

Introduction

Investigating the interactions between components as a network provides a common platform to uncover signatures of complex systems[1-5]. While this approach has been recently exploited to investigate biological systems of different scales, ranging from interactions between molecules to interactions between species, application of network theory at the level of individual cells has been rather limited. In this study, we present a network based approach to understand general principles in the organisation of cells in an organism (*i.e.*, epithelia). Early in animal development, cells in an epithelia begin to divide and alter their position, shape and size in stereotypical ways[6-15]. Despite being a highly dynamic process, this results in ordered, robust structure that ultimately leads to the formation of mature organs with cellular organisation suited to their specialised functions. Though we have an understanding of the contribution of genetic mechanisms (*e.g.*, external signals and the associated gene regulatory pathways[6, 14]) and cellular mechanics (*e.g.*, intrinsic patterns arising due to the rate of cell division[8, 9, 11, 15] or cell re-arrangement due to anisotropy of cortical forces within individual cells[6, 16-20]) to the development of epithelial architecture in various model systems, we lack the means to objectively characterise and quantify the similarities and differences in the organisation of epithelia.

Previous studies have offered insights into epithelial organisation by focusing primarily on geometric characteristics of individual cells such as the cell area and the number of contacts[8, 9, 18, 21] and have led to the formulation of empirical relationships such as Aboav-Weaire's law and Lewis' law[22, 23]. This has largely emphasized the similarity between different epithelia. However, a more comprehensive view of epithelial organisation can be achieved if one considers the higher order organisation of cells such as the patterns in the network of interactions between cells that typically characterises an epithelium. The ability to do this would provide a way to describe objectively an epithelium, facilitate the investigation of fundamental questions about its organisation and dynamics, and establish an objective basis for comparative studies of epithelia from different sources. Importantly, the network characteristics of epithelial organisation (as opposed to geometric features) are not readily assessed by eye. This implies that higher-order organisation may not be accounted for in our current understanding of epithelial architecture.

In this work we present an approach, we term GNEO (Geometric and Network representation of Epithelial Organisation) which, by combining network and geometric measures of epithelial organisation, addresses these issues. We show that our approach is able to capture a defining signature that distinguishes epithelia from different organs, species, developmental stages and genetic conditions. In this way, GNEO permits characterization, quantification and classification of normal and perturbed epithelia in an objective manner.

Results

The GNEO method for characterising epithelial organisation. In order to capture information about the spatial organisation of cells and the global features of an epithelium, we generated network representations of confocal images of epithelia based on the cell-cell contacts. This allows principles from graph theory and complex networks[2, 3, 5, 24] to be used to investigate short- and long-range patterns in epithelial organisation. In such a network, the centre of each cell is treated as a node and two nodes are linked if the two cells are neighbors (*i.e.*, physically contact each other) in the epithelium (*see* **Methods**; **Fig. 1a** and **Supplementary Fig. S1**). For each image, we generated a 'feature vector' consisting of eight features (*see* **Fig. 2**): the means and standard deviations of the cell area, degree (*i.e.*, number of neighbors), clustering coefficient (the amount of interconnectedness among a cell's immediate neighbors) and average degree of neighbors (the average number of neighbors of a cell's neighbor). Thus, the mean values of the features in the feature vector represent information about the cell shape (area and degree) and the pattern of cell-to-cell contacts (degree, clustering coefficient and average degree of neighbors). While the degree

1

is informative of the short-range pattern of contacts (the immediate neighbors of cell), the clustering coefficient and average degree of the neighbors represents the cell's context and surrounding, thus reflecting higher order organization. In turn, the standard deviation values are indicative of the cell-to-cell variability (*i.e.*, heterogeneity) of a feature in an epithelium.

For each type of epithelium, we collected a set of images from several different individuals and extracted the feature vectors in each case (**Supplementary Table S1**). This allows us to compare epithelia, for example from different developmental stages, tissues and species (**Fig. 1b** and **Supplementary Fig. S2-S3**), and the natural variation (*i.e.*, individual-to-individual variability) in epithelial organisation. To compare different epithelia, we used multivariate statistical methods to identify the contribution of the different feature in the feature vector that best separate different epithelial types. We took advantage of an unsupervised and a supervised method, namely Principal Component Analysis (PCA) and Discriminant Analysis (DA)[4, 24, 25] (*see* **Fig. 2** and **Methods**). Both these methods provide information about the relative contribution of the features that distinguish different epithelia, termed "feature weights" (**Fig. 3a** and **Supplementary Table S2**). The statistical significance of the separation of the different epithelia was assessed using the MANOVA test (*see* **Methods**).

Comparison of epithelia from different organisms. To validate the approach, we first used the feature vector to compare visually distinct epithelia. For this we took advantage of the neural tube (samples cNT1 to cNT12) and embryonic ectoderm (cEE1 to cEE14) from chicken embryos, and the *Drosophila* wing imaginal disc from the prepupal stage (dWP1 to dWP16). Discriminant Analysis and PCA revealed that the epithelia from the two different organisms could be clustered into two distinct groups (**Fig. 3b**; cNT and dWP; DA; $p=9.17\times10^{-18}$; **Supplementary Fig. S4**; cEE and dWP; DA; $p=2.41\times10^{-21}$, **Supplementary Fig. S5**) demonstrating the efficacy of the method. The greater importance of the average degree of the neighbours (N) and the standard deviation of the degree (D) in the dWP-cNT separation (**Fig. 3a,b** and **Supplementary Table S1**) suggests that these two network features capture a certain defining signature that is independent of the cell area (which is comparable for these epithelial types).

Comparison of different epithelial types. To verify whether we could discriminate between different types of epithelia from the same organism, we compared the columnar neuroepithelium (cNT) of the chick to the squamous embryonic ectoderm (cEE). PCA and Discriminant Analysis of the feature vectors of these two epithelia revealed that they form two distinct groups (**Fig. 3c**; DA; $p=4.49\times10^{-16}$; **Supplementary Fig. S6**; PCA; $p=4.27\times10^{-10}$). In this case, the cEE dataset was more spread out, reflecting the greater heterogeneity among these samples. However, each sample was clearly separated from cNT epithelia. In addition, the method separated the *Drosophila* wing pouch (dWP), the chick neuroepithelium (cNT) and embryonic ectoderm (cEE) epithelia into three groups (**Fig. 3d**), demonstrating that it is sensitive to different types of epithelial organization. Multiple features contributed to the separation of these epithelia, suggesting that it was the combination of features that allowed the discrimination (**Fig. 3a**). This underscores the importance of global structure of the network. In particular, we found that the standard deviation of the non-geometric (*i.e.*, network) features were more important in separating the different epithelial types, suggesting that, in this case, GNEO is able to capture patterns in epithelial organisation which are not visually apparent and are independent of cell area.

Comparison of epithelia from different developmental stages. We next compared more closely related epithelia. Discriminant Analysis on the feature vectors of the wildtype (WT) epithelia of the wing pouch (which develops into the adult wing blade; dWP1 to dWP16) and the notum (which develops into the adult thorax; dNP1 to dNP12) of the wing imaginal disc from the prepupal stage showed that these samples were only partially separable. This indicates a similarity in the overall organisation of both epithelia at this stage during development (**Fig. 3e**; DA; $p=0.001$), consistent with these regions comprising different areas of the same epithelial sheet. Moreover, a comparison of the prepupal wing pouch (dWP1 to dWP16) with the third instar larva wing pouch (dWL1 to dWL15) produced a discriminant graph with two groups, which also overlapped but were better separated (**Fig. 3f**; DA; $p=3.45\times10^{-5}$). Together these data suggest that the global organization of the wing epithelia change gradually during development and across an epithelial sheet. The GNEO approach provides a way to assess the relatedness of different epithelia through approaches that calculate distances between data points using standard approaches such as hierarchical clustering. Consistent with our observation, a comparison of four epithelial types (Discriminant Analysis and PCA of the *Drosophila* dWP and dNP and the chick cEE and cNT) showed that dWP and dNP forms an overlapping cluster whereas the cEE and cNT form separate clusters (**Fig. 3g**). These observations confirm that the epithelial organisation of the prepupal notum and

wing are similar whereas the neural tube and embryonic ectoderm are not.

Comparison of epithelia from different tissues. While the wing pouch and the notum are closely related, we tested if our approach can separate distantly related epithelia in *Drosophila*, by comparing the third instar larval wing disc epithelia (dWL) with the eye epithelia (dEL1 to dEL5). Unlike the wing disc epithelium, the third instar eye disc contains a gradient of apical constriction induced by myosin II activation[26, 27], leading to apparently different cell shapes and sizes across the epithelial sheet. Discriminant Analysis of the feature vectors clearly demonstrated that the eye and wing disc epithelia are significantly different and that most of the features, except the average clustering coefficient, contribute to this separation (**Supplementary Fig. S8**; DA; $p=1.11 \times 10^{-5}$).

Natural variation in epithelial organisation. The observation that the network features of the feature vectors were important in discriminating different types of epithelial tissues raised the possibility that local cell packing (resulting from cell shape and distribution) gives rise to the characteristic long-range patterns of an epithelium. In this view, the global organization is a property of the epithelium that emerges from the collective behaviour of the cells. Consistent with this idea, the individual-to-individual variation (measured as coefficient of variation (C.V.), *see* **Fig. 3h** and **Methods**) of mean values of the network features between the different individuals is several fold smaller than that of cell area (**Fig. 3h**). In addition, the individual-to-individual variation (C.V.) of the mean network features was much smaller than the C.V. of the cell-to-cell variability (*i.e.*, the standard deviation of the features in the feature vector) across individuals. These observations suggest that there is a reproducible long-range epithelial structure, which is, to a large extent, independent of variations in cell size.

Comparison of wild-type and genetically perturbed epithelia. What regulates this reproducible long-range organization of epithelia? Several factors have been implicated in controlling the behaviour of individual cells and consequently epithelial architecture[6, 8, 9, 11, 15-18]. However, how global epithelial structure is determined by, for example the effect of the cytoskeleton of the cells within the epithelium is not understood. Therefore, we applied GNEO to quantify objectively the effect in the wing disc of removing myosin II heavy chain using RNAi[28, 29], a genetic manipulation that robustly and uniformly disrupts the cytoskeletal organization and epithelial architecture (*see* **Methods**). Both PCA and DA were clearly able to separate wildtype (WT) discs from those in which myosin II had been reduced (**Fig. 4a** and **Supplementary Fig. S9a**). Interestingly, while all WT wings formed one distinct tight cluster, the mutant wings were more broadly spread over the graph (**Fig. 3a**; PCA; $p=4.27 \times 10^{-10}$). This is consistent with a visual inspection of the data, which showed that reducing the levels of myosin II by RNAi knockdown disrupted epithelial organisation to different extents in different wing discs. In order to provide an objective score for the severity of the mutant phenotype, we calculated the Euclidean distance between each mutant wing and the center of mass of the WT wings (**Supplementary Table S3**). The coefficient of variation of the distances was ~26%, which most likely reflects the variability of the RNAi efficiency among the individuals. Individual mutant samples were between 15 and 35 times further from the center of mass than the average of the distances of the dWP samples (**Supplementary Table S3**).

We then investigated how the inclusion of other epithelia during the comparison affected our analysis of myosin II reduction. Addition of the prepupal notum (dNP) showed that both wildtype epithelia cluster together whereas the mutant wing pouch epithelia were still scattered (**Fig. 4b**; PCA; $p=1.57 \times 10^{-13}$ and **Supplementary Fig. S9b**; DA; $p=1.89 \times 10^{-18}$). Thus GNEO can objectively recognise that the *Drosophila* WT samples are similar to each other but distinct from the mutant ones. A PCA that includes the chick embryonic ectoderm and the neural tube, either together or alone, revealed that each group forms distinct clusters (PCA: **Fig. 4c**; $p=2.98 \times 10^{-31}$; **Fig. 4d**; $p=3.40 \times 10^{-22}$ and DA: **Supplementary Fig. S9c**; $p=2.03 \times 10^{-44}$; **Supplementary Fig. S9d**; $p=1.36 \times 10^{-34}$). This suggests that the *Drosophila* mutant epithelia are clearly distinct from the WT *Drosophila* and chick epithelia. Discriminant Analysis indicated that the area, the degree, the degree of neighbours and the standard deviation of the degree are the most important features (**Fig. 3a**) for obtaining this separation (**Supplementary Fig. S9a-d**). This is consistent with an effect of myosin II on regulating cell shape[26, 27, 30, 31], whose levels when altered affects cell area, the number of neighbours and the "regularity" of the pattern of cell contacts. The increased variability of the network features when myosin II is knocked down demonstrates an important biological conclusion of this work: long-range constancy of epithelial packing is regulated by cytoskeletal organisation within individual cells. This is also reflected by the higher variation between discs from different individuals in the myosin II knock-down compared to WT discs (**Fig. 3h**). Together, these data suggest that the long range organization of an epithelium is determined, at least in part, by the cytoskeleton of the cells comprising the tissue affecting

interactions across the field of cells.

Discussion

The network-based approach, GNEO, introduced here, captures epithelial organisation by accounting for patterns of cell-contacts which cannot be quantified either by visual inspection or by using geometric features that describe individual cells alone. In particular, GNEO can objectively quantify differences between epithelia from different tissues and organisms, even when the size and shape of the cells comprising these epithelial appear visually indistinguishable. First, we show that epithelia from different organs and species have distinct, reproducible and quantifiable differences in their structure. Second, a surprising result is that differences in cell area play a relatively minor role in distinguishing wing disc and neural tube epithelia – two epithelia that come from different species which produce very different tissues. This indicates that there is unexpected consistency in both the average size and range of sizes of cells across epithelia from different species. Third, we provide evidence that the non-geometric features of the epithelia are most informative in distinguishing them. This shows that the topological organization of the epithelium differs strongly between tissues and between species. Together with the finding that the structure of the same epithelia from different individuals is highly reproducible, our analysis indicates an unexpected level of genetic control over the long range organization of epithelia. To the best of our knowledge this has not been previously reported.

Lewis's and Aboav-Weaire's laws[23] define relationships between area, degree and neighbour degree, and place the emphasis on the universal connection between these features. By contrast our work examines how measurements of these features across a population of cells can be used to build a quantitative and objective description of an epithelium. The experiment in which we disrupt myosin II provides an indication of the mechanism by which population level features of an epithelium emerge from the collective behaviour of individual cells. Moreover, we provide evidence that GNEO can operate as a reliable classifier to differentiate mutant and WT epithelia, and to quantify precisely the severity of mutant phenotypes.

While there are other ways of constructing networks from images of epithelia, this representation of cell-to-cell contacts offers a simple and readily applicable method to analyze epithelia objectively. Representing epithelia as feature vectors opens up the possibility of applying artificial intelligence (*e.g.*, pattern recognition) algorithms to classify them in an objective manner and can be extended to include more sophisticated network features. Since the method is automatable, adaptations of this approach can be used in high-throughput experiments aimed at identifying pathways and quantifying the effects of mutations in functional genomics screens. Of particular value, GNEO allows characterisation of subtle phenotypes undetectable by visual inspection. This approach can also be adapted to other biological samples such as nerve-cell connections, muscle cells attachments, and tumours.

Methods

Genetic strains and confocal imaging of the epithelia. Flies were grown by employing standard culture techniques. The following lines were used: Arm-GFP (WT), *C765-Gal4* (http://flybase.bio.indiana.edu) and *U-zip-RNAi* (Vienna *Drosophila* RNAi Center collection). Imaginal discs from the prepupa and third instar larvae were dissected in PBS and fixed with 4% paraformaldehyde in PBS for 35 min. The samples were washed six times for 10 min with PBT (PBS, 0.3% triton) and 3 times for 5 min with PBS. Imaginal discs were mounted using Fluoromount-G (Southern Biotech). Images were taken with a BioRad Radiance 2100 laser scanning confocal microscope. All the images were captured using 63x immersions objective with 3 times zoom and exported as a 1024 x 1024 pixel TIFF file. The area of 1 pixel is 3.78×10^{-3} μm^2. The regions of the imaginal discs that appear in the images were selected with the following criteria: For the prepupal and larval wing disc, images from dorsal compartments (leaving out the D/V boundary region) of the wing pouch region were obtained. The notum images were taken from the anterior part of the disc (see **Supplementary Fig. S2**). For the eye disc, the images were taken locating the morphogenetic furrow at the side with an additional margin of three to five rows of cells (to include the first and second rows of clusters of photoreceptors). For the chicken images, Hamburger and Hamilton (HH) [32] stage 10 and 17 chick embryos (see **Supplementary Fig. S3**), were fixed for 1 hour in 4% paraformaldehyde. HH st. 17 embryos were subsequently transferred to methanol then rehydrated. Immunostaining was performed with a ZO1 antibody (Zymed labs) and embryos were flat mounted. Images of neural tube epithelium, at intermediate dorsoventral positions, were obtained at the level of somite 5 of HH st. 17 embryos. Embryonic ectoderm was imaged adjacent to the most recently formed somite of HH st. 10 embryos. Imaging conditions were as for imaginal discs, image orientation: anterior, left; dorsal, up.

Image processing and generation of the epithelial network. The acquired images were converted into their respective 2-pixel BMP files. First, the confocal images were imported using Adobe Photoshop CS2. Colours were inverted to obtain a dark signal over a light background. Epithelial cells were identified from the processed images by using a semi-automated framework. The images first had their illumination corrected through polynomial fitting followed by thresholding, and were then manually checked and edited to remove artefactual connections between cells. The cells and their boundaries were defined using a foreground tracking program[4]. Adjacent cells were identified by transforming their images into an epithelial network where the centroid of each cell is represented by a node and the neighborhood relations are mapped into weighted edges (*see* **Supplementary Fig. S1**). The methodology involves the following steps: (a) all cells have their borders detected[4], (b) for each pixel p of the border of each cell i, all border pixels belonging to other cells and falling within the circle of radius r centered at p are identified and counted (we use r=6) (c) the node corresponding to this cell is connected to other cells by edges whose weights, w, correspond to the total number of neighboring pixels found during step (b). We established empirically that if the weight is greater than 40% of the minimum equivalent radius of the areas of each adjacent pair of cells, those cells can be considered to be neighbours. The equivalent radius of a cell with area A is defined as corresponding to $\sqrt{A/\pi}$. We selected an area within every image in order to have all the networks presented with similar boundary conditions. The squares were drawn to obtain the maximum possible surface without including the centroids of the cells in the border of the images. The features of the cells falling within this area were calculated. The cells outside were only used in order to provide neighbors to the cells analyzed in the network.

Calculation of geometric and network features. The area was measured by counting the number of pixels inside each cell[4]. Each epithelial image was represented as a network where each node corresponds to one of the cells and the links between nodes reflect the spatial adjacency between the epithelial cells (**Fig. 1a**). Several measurements can be estimated from these networks[24] in order to provide useful characterisation and respective biological interpretations. Let the graph be represented in terms of its adjacency matrix K, such that the presence of a connection between nodes i and j, with $1 \leq i, j \leq N$, implies $K(i,j) = K(j,i) = 1$, with $K(i,j) = K(j,i) = 0$ being enforced otherwise. In this work, we employed the following topological characteristics: (a) *Degree of a node i*, which corresponds to the number of edges attached to it, *i.e.*, $k(i) = \sum_{v=1}^{N} K(v,i)$. (b) *Clustering coefficient of a node i*, which is obtained by dividing the number of edges between the neighbours of i, represented as $n(i)$, by the maximum possible number of connections between those nodes. This measurement can be calculated as $c(i) = 2\dfrac{n(i)}{k(i)(k(i)-1)}$. Therefore, the clustering coefficient varies from 0 (no interconnections between the neighbours of i) to 1 (the neighbours are fully interconnected). (c) *Average degree of the neighbours of a node*, calculated as the average of the number of edges that are attached to the neighbours of node i. This measurement can be calculated as $b(i) = \dfrac{1}{k(i)} \sum_{v=\text{neighbors of } i} k(v)$. For every feature, both the average and standard deviation values across all cells in an epithelium were estimated and were used to characterise epithelial organisation. Cell elongation was initially considered as an independent parameter but was found not to contribute further to the classification and separation. We believe that this is because variations in the elongation tend to affect the degree, clustering coefficient and average degree of neighbours, therefore becoming correlated with those measurements.

Feature vector and multivariate statistical analysis. The geometric and network analyses of epithelial images yield a large number of features or measurements (**Supplementary Table S1**). More precisely, a total of 8 features, corresponding to the mean and standard deviation of the area of cells, as well as the degree, clustering coefficient and average degree of a node are obtained. Thus, for every image of an epithelium, a feature vector of eight dimensions was obtained. We apply an unsupervised and supervised multivariate statistical method, namely Principal Component Analysis (PCA) and Discriminant Analysis (DA)[4, 24, 25] (*see* **Supplementary Methods** for explanation). Standardisation[4] of the measurements is performed in order to eliminate the effect of the magnitude of the measurements on the respective separation between categories. More specifically, the average and standard deviation for every feature across all individuals in each of the epithelial type was calculated. For every feature in a feature vector representing an individual epithelium, the calculated average value was subtracted and divided by the standard deviation. Thus, the components of the eigenvectors associated to the largest eigenvalues provide a quantification of the degree of

contribution of each original measurement in maximising the dispersion of the projection (in the case of PCA) and optimising the separation between the categories (in the case of discriminant analysis). Since both these methods output a weighted linear combination of all the features in the feature vector to obtain the final projection (the new axes), the loading values associated to each feature provide a direct measure of the contribution of that particular feature towards the projection onto the smaller-dimensional space. For our analysis, the maximum between the magnitudes of the values of the components of each feature (*i.e.*, measurement) of the first two eigenvector (*i.e.*, the loadings associated with the first two component axes) was defined as the weight of that respective measurement.

Estimation of statistical significance. The MANOVA (multivariate analysis of variance) test, which is a reference statistical test for probing the hypotheses that two or more populations, characterized in terms of two or more dependent variables, are or not distinct, was used to assess the statistical significance of the obtained separation. The tested (null) hypothesis H0 is that the two samples come from the same population, with H1 indicating different populations. The p-values are calculated in the standard way[33, 34], after PCA or DA. Both PCA and DA provide new random variables that are linear combinations of the previous ones. In the case of PCA, the new variables are completely uncorrelated. In the case of the DA, the method does take into account the known categories of the cells and therefore enhances the separation between the categories. In this case, our interest was focused on the contribution of the measurements on the separation, not on the separation itself. The p-values in this case reflect the effect of the informed cell categories and should be treated as such. The assumptions required for MANOVA were verified even though after application of PCA and DA, as the respective clusters remained largely normal and with similar variances.

References

1. Barabasi, A. L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. Nat Rev Genet 5, 101-13 (2004).
2. Albert, R. & Barabasi, A.-L. Statistical mechanics of complex networks. Reviews of Modern Physics 74, 47-97 (2002).
3. Amaral, L. A. N. & Ottino, J. M. Augmenting the framework for the study of complex systems. The European Physical Journal B 38, 147-162 (2004).
4. Costa, L. D. & Cesar, R. M. J. Shape Analysis and Classification: Theory and Practice (CRC Press, 2000, 2009).
5. Strogatz, S. H. Exploring complex networks. Nature 410, 268-76 (2001).
6. Lecuit, T. & Le Goff, L. Orchestrating size and shape during morphogenesis. Nature 450, 189-92 (2007).
7. Axelrod, J. D. Cell shape in proliferating epithelia: a multifaceted problem. Cell 126, 643-5 (2006).
8. Farhadifar, R., Roper, J. C., Aigouy, B., Eaton, S. & Julicher, F. The influence of cell mechanics, cell-cell interactions, and proliferation on epithelial packing. Curr Biol 17, 2095-104 (2007).
9. Gibson, M. C., Patel, A. B., Nagpal, R. & Perrimon, N. The emergence of geometric order in proliferating metazoan epithelia. Nature 442, 1038-41 (2006).
10. Hayashi, T. & Carthew, R. W. Surface mechanics mediate pattern formation in the developing retina. Nature 431, 647-52 (2004).
11. Rolland-Lagan, A. G., Bangham, J. A. & Coen, E. Growth dynamics underlying petal shape and asymmetry. Nature 422, 161-3 (2003).
12. Gho, M., Bellaiche, Y. & Schweisguth, F. Revisiting the Drosophila microchaete lineage: a novel intrinsically asymmetric cell division generates a glial cell. Development 126, 3573-84 (1999).
13. Classen, A. K., Anderson, K. I., Marois, E. & Eaton, S. Hexagonal packing of Drosophila wing epithelial cells by the planar cell polarity pathway. Dev Cell 9, 805-17 (2005).
14. Martinez-Arias, M. & Stewart, A. Molecular Principles of Animal Development (OUP Oxford, 2002).
15. Patel, A. B., Gibson, W. T., Gibson, M. C. & Nagpal, R. Modeling and inferring cleavage patterns in proliferating epithelia. PLoS Comput Biol 5, e1000412 (2009).
16. Cavey, M., Rauzi, M., Lenne, P. F. & Lecuit, T. A two-tiered mechanism for stabilization and immobilization of E-cadherin. Nature 453, 751-6 (2008).
17. Lecuit, T. & Lenne, P. F. Cell surface mechanics and the control of cell shape, tissue patterns and morphogenesis. Nat Rev Mol Cell Biol 8, 633-44 (2007).
18. Rauzi, M., Verant, P., Lecuit, T. & Lenne, P. F. Nature and anisotropy of cortical forces orienting Drosophila tissue morphogenesis. Nat Cell Biol 10, 1401-10 (2008).
19. Martin, A. C., Gelbart, M., Fernandez-Gonzalez, R., Kaschube, M. & Wieschaus, E. F. Integration of contractile forces during tissue invagination. J Cell Biol 188, 735-49 (2010).
20. Martin, A. C., Kaschube, M. & Wieschaus, E. F. Pulsed contractions of an actin-myosin network drive apical constriction. Nature 457, 495-9 (2009).
21. Ma, D. et al. Cell packing influences planar cell polarity signaling. Proc Natl Acad Sci U S A 105, 18800-5 (2008).
22. Mombach, J. C., de Almeida, R. M. & Iglesias, J. R. Two-cell correlations in biological tissues. Phys Rev E Stat Phys

Plasmas Fluids Relat Interdiscip Topics 47, 3712-3716 (1993).

23.   Chiu, S. N. Aboav-Weaire's and Lewis' laws—A review. Materials Characterization 34, 149-165 (1995).

24.   Costa, L. D., Rodrigues, F. A., Travieso, G. & Boas, P. R. V. Characterization of complex networks: a survey of measurements. Advances in Physics 56, 167-242 (2007).

25.   Pincus, Z. & Theriot, J. A. Comparison of quantitative methods for cell-shape analysis. J Microsc 227, 140-56 (2007).

26.   Corrigall, D., Walther, R. F., Rodriguez, L., Fichelson, P. & Pichaud, F. Hedgehog signaling is a principal inducer of Myosin-II-driven cell ingression in Drosophila epithelia. Dev Cell 13, 730-42 (2007).

27.   Escudero, L. M., Bischoff, M. & Freeman, M. Myosin II regulates complex cellular arrangement and epithelial architecture in Drosophila. Dev Cell 13, 717-29 (2007).

28.   Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development 118, 401-15 (1993).

29.   Sharp, P. A. RNA interference--2001. Genes Dev 15, 485-90 (2001).

30.   Bertet, C., Sulak, L. & Lecuit, T. Myosin-dependent junction remodelling controls planar cell intercalation and axis elongation. Nature 429, 667-71 (2004).

31.   Pilot, F. & Lecuit, T. Compartmentalized morphogenesis in epithelia: from cell to tissue shape. Dev Dyn 232, 685-94 (2005).

32.   Hamburger, V. & Hamilton., H. L. A series of normal stages in the development of the chick embryo. J. Morphol 88, 49-92 (1951).

33.   Izenman, A. J. Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning (Springer, 2008).

34.   Jolliffe, I. T. Principal Component Analysis (Springer, 2010).

Acknowledgements

Author contributions

MMB and LME designed the study with help from LdFC and MF. LME and AK obtained the fly and chicken images, respectively. LME processed the images. LdFC wrote the software, generated the epithelial network and performed all the comparisons and statistical analysis. All authors participated in the interpretation of results, discussions and the development of the project. LME and MB wrote the manuscript with input from all authors.

Additional Information

Supplementary Information: The entire pipeline of the steps involved is shown in the **Supplementary Fig. S10**. Computer programs and source code of the implementation of the algorithms in Scilab are enclosed in the **Supplementary Software S1-S5** files. The entire dataset used in our analysis is enclosed as **Supplementary Datasets S1-S7**.

Competing financial interests: The authors declare no competing financial interests.
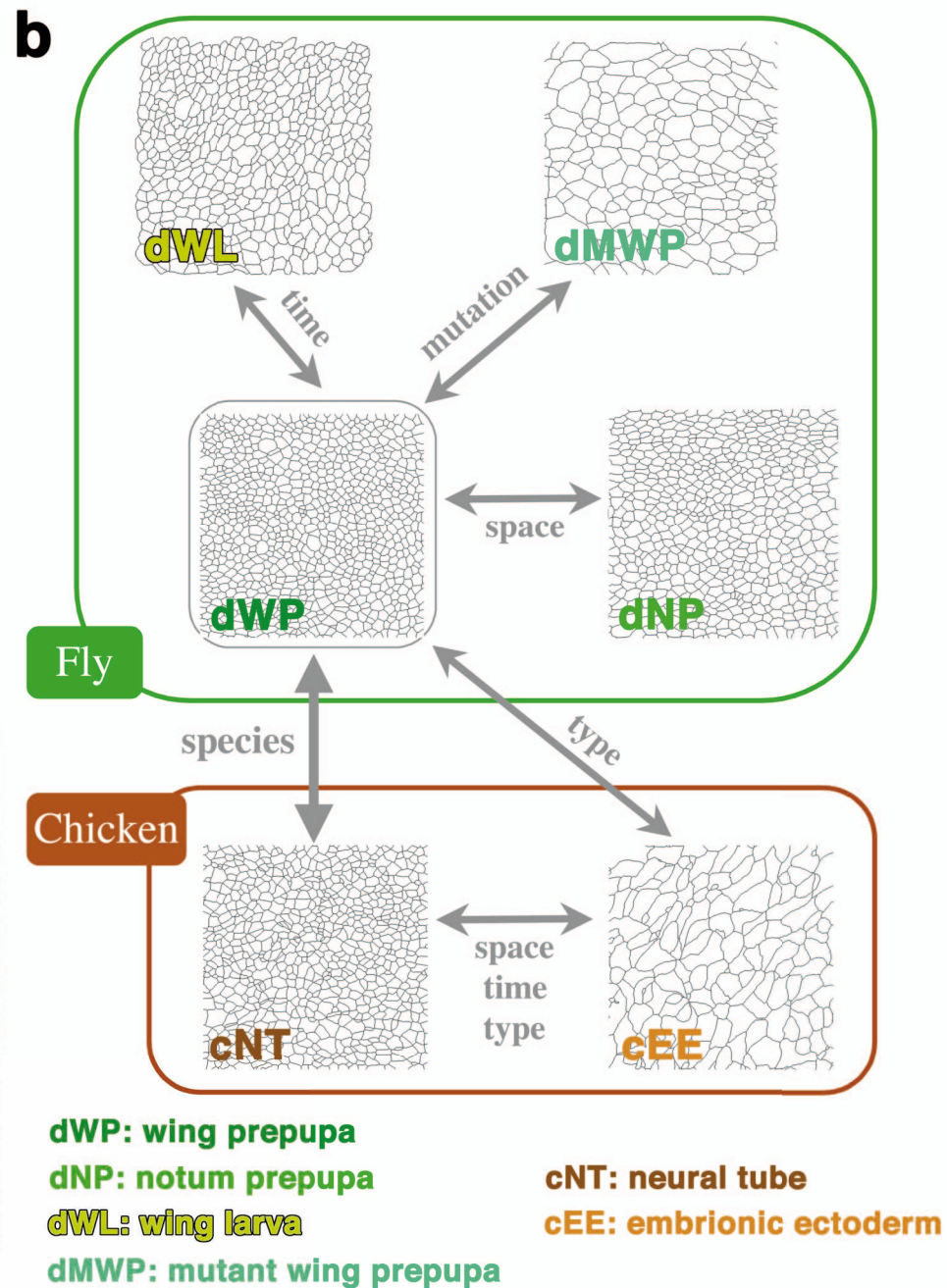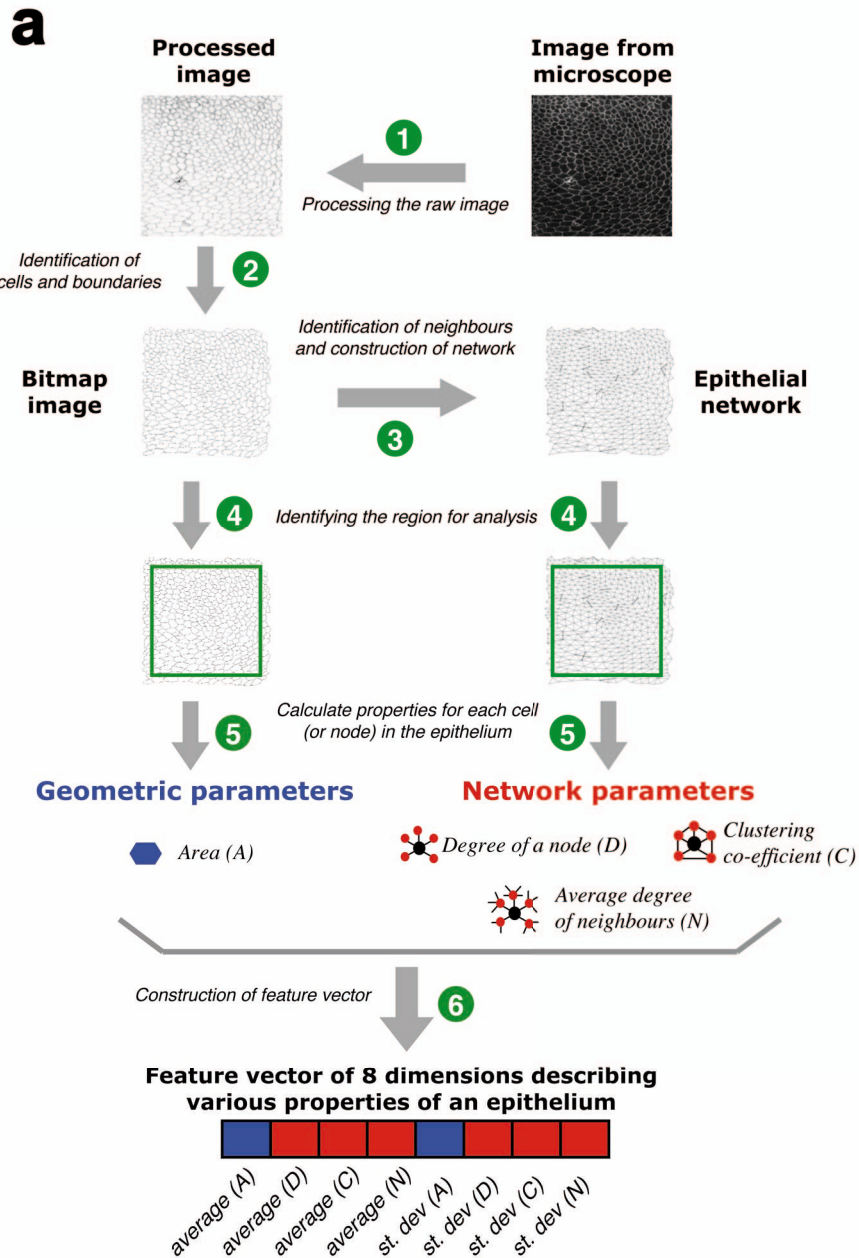
Figure Legends

Figure 1. GNEO approach and epithelial comparisons performed. (**a**) <u>G</u>eometric and <u>N</u>etwork representation of <u>E</u>pithelial <u>O</u>rganisation (GNEO) approach to characterise epithelial organisation. (1) Images from the confocal microscope are processed to get a light background with dark cell contours. (2) The processed image is the source for defining individual cells, as well as for determining the number of neighbours for every cell. (3) This information is used to produce an epithelial network where each cell is represented as a node and two nodes are connected if the two cells are neighbours in the epithelium. (4) A region of interest (ROI; shown in a green box) is selected for further analysis and the cells that border the ROI are excluded. (5) The average and standard deviation of area and three network features over all cells in an epithelium are calculated. (6) This information is represented as a feature vector, which is an 8-dimensional vector that characterises each epithelium. Each of the four features considered in this work is abbreviated by the symbols shown in this figure. (**b**) Schematic representation of the comparisons of epithelia from different sources performed in this study. Images of representative epithelial samples (2-pixel wide cell contour) from *Drosophila* (different tones of green labels, fly) and chicken (different tones of brown labels) are shown. The reference prepupal wing pouch epithelium is shown within a gray box. The text in gray denotes the relationship between epithelia from the different sources that were compared. *space*: spatially separated epithelia from the same organism; *time*: temporally separated epithelia from different stages of development; *type*: different type of epithelia (*e.g.*, squamous and columnar); *species*: epithelia from different organisms (vertebrate (chick) and invertebrate (*Drosophila*)) and *mutation*: mutant epithelia.

Figure 2. Description of features and statistical approaches. Qualitative definition of the geometric and network features and the statistical approaches used (*see* **Methods** for more details).

Figure 3. Discriminant Analysis of the different epithelia. (**a**) Colour coded matrix representation of the weights of the features contributing to the observed projection in the Discriminant Analysis of the different comparisons. A higher value (darker color) represents a relatively higher contribution of the feature to the separation. (**b-g**) Discriminant Analysis graphs of the comparisons of epithelia from different sources. (**b**) dWP-cNT. (**c**) cNT-cEE. (**d**) dWP-cNT-cEE. (**e**) dWP-dNP. (**f**) dWP-dWL. (**g**) dWP-dNP-cNT-cEE.

Table 1. Coefficient of Variation for the different features. Coefficient of Variation values describing the individual-to-individual variation for the different features in the feature vector. dWP: Wing prepupa, dNP: Notum prepupa, dMWP: mutant Wing prepupa, dWL: Wing larva, cNT: chicken Neural Tube, cEE: chicken Embryonic Ectoderm.

Figure 4. PCA graphs of the wildtype and mutant epithelia. (**a**) dWP-dMWP. (**b**) dWP-dNP-dMWP. (**c**) dWP-dMWP-cEE. (**d**) dWP-dMWP-cEE-cNT. dWP: Wing prepupa, cNT: chicken Neural Tube, cEE: chicken Embryonic Ectoderm, dMWP: mutant Wing prepupa.

a

Processed image ← **1** Image from microscope
Processing the raw image

Identification of cells and boundaries **2**

Bitmap image → **3** → Epithelial network
Identification of neighbours and construction of network

Identifying the region for analysis **4** ... **4**

Calculate properties for each cell (or node) in the epithelium **5** ... **5**

Geometric parameters   **Network parameters**

Area (A)   *Degree of a node (D)*   *Clustering co-efficient (C)*

Average degree of neighbours (N)

Construction of feature vector **6**

Feature vector of 8 dimensions describing various properties of an epithelium

average (A) | average (D) | average (C) | average (N) | st. dev (A) | st. dev (D) | st. dev (C) | st. dev (N)

b

dWL —*time*— dWP —*mutation*— dMWP

dWP —*space*— dNP

Fly

Chicken

species   *type*

cNT —*space time type*— cEE

dWP: wing prepupa
dNP: notum prepupa
dWL: wing larva
dMWP: mutant wing prepupa

cNT: neural tube
cEE: embrionic ectoderm

Geometric parameters

 Cell area (A)   The area of a cell measures the size of the apical surface, which is related to the 3D shape of the cell.

Network parameters

 Degree of a node (D)   The degree of a node (cell) is the number of neighbours of each cell and its average value for an epithelium provides a good indication of the overall connectivity between cells.

 Clustering coefficient (C)   The clustering coefficient measures the inter-connectedness between the neighbors of a cell, and is a measure of the local packing.

 Average degree of neighbours (N)   The average degree of the neighbors of a cell reflects the connectivity of its neighbors, and provides information about the wider-context of connectivity around the reference cell.

Multivariate statistical analysis

Principal Component Analysis (PCA)   PCA transforms the correlated data points of the feature vector into a small number of uncorrelated variables called principal components. The projection maximises the dispersion of the individual data points without prior knowledge of whether these are expected to form groups. This allows the unbiased identification of naturally separated sets of data points that may form groups because they are similar.

Discriminant analysis (DA)   DA projects the data points of the feature vector onto a smaller-dimensional space so as to optimise the separation of the defined groups, thereby providing information about which features contribute most to the separation of the groups.

a Weights from Discriminant Analysis

0.01 — 0.92

	average				standard deviation			
	A	D	C	N	A	D	C	N

Rows:
- dWP/cNT
- dWP/dNP/cNT
- dWP/cEE
- dWP/dNP/cEE
- cNT/cEE
- dWP/cNT/cEE
- dWP/dNP/cNT/cEE
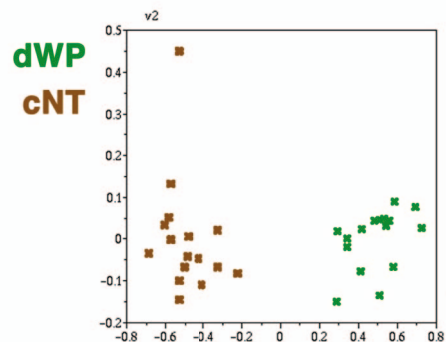- dWP/dNP
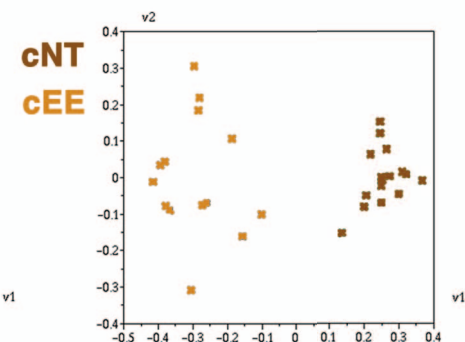- dWP/dWL
- dWP/dNP/dWL
- dWP/dMWP
- dWP/dNP/dMWP
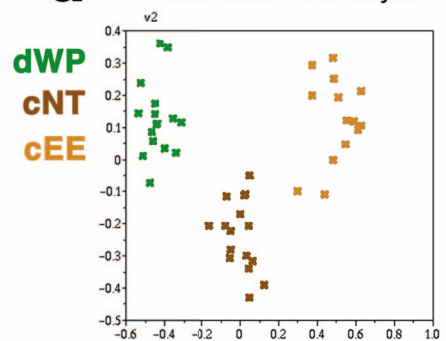- dWP/dMWP/cEE
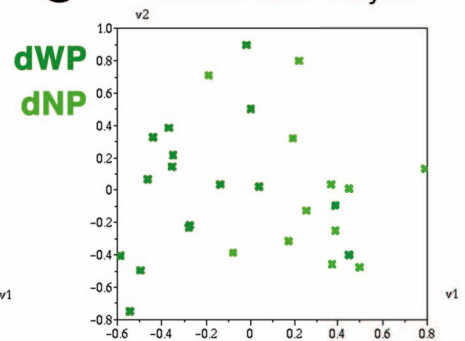- dWP/dMWP/cNT/cEE

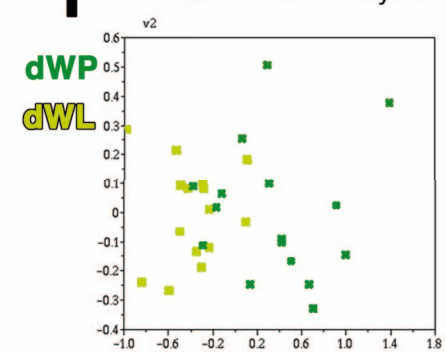**b** Discriminant Analysis
dWP
cNT

c Discriminant Analysis
cNT
cEE

d Discriminant Analysis
dWP
cNT
cEE

e Discriminant Analysis
dWP
dNP

f Discriminant Analysis
dWP
dWL

g Discriminant Analysis
dWP
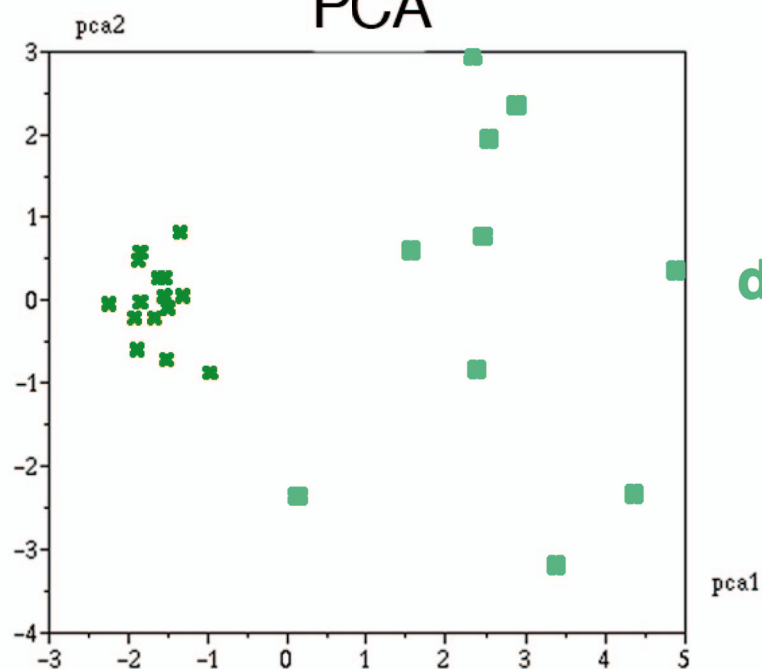dNP
cNT
cEE
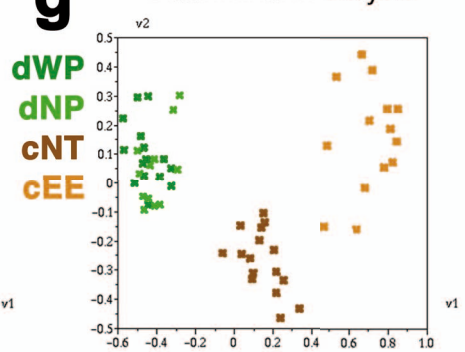
dWP: wing prepupa
dNP: notum prepupa
dWL: wing larva
dMWP: mutant wing prepupa

cNT: neural tube
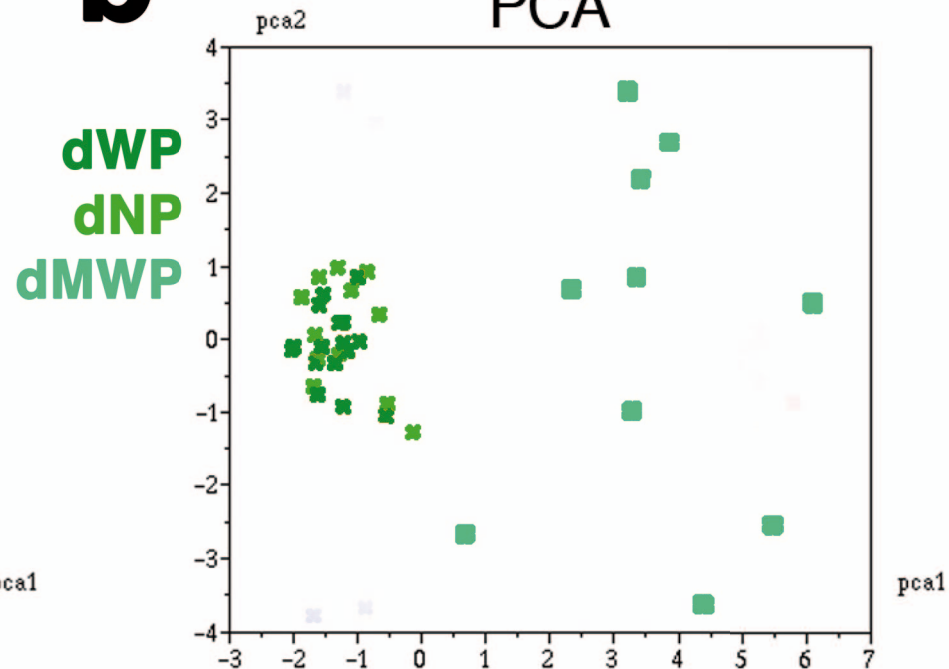cEE: embrionic ectoderm

a

PCA

dWP
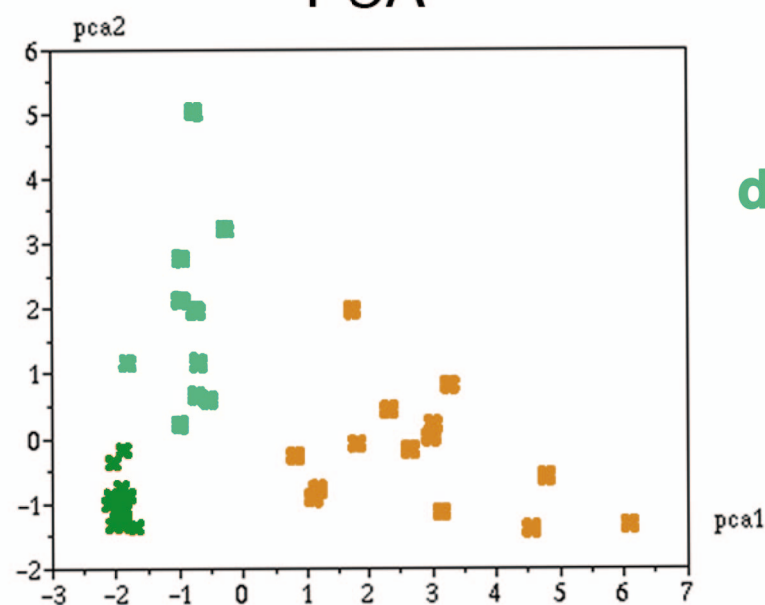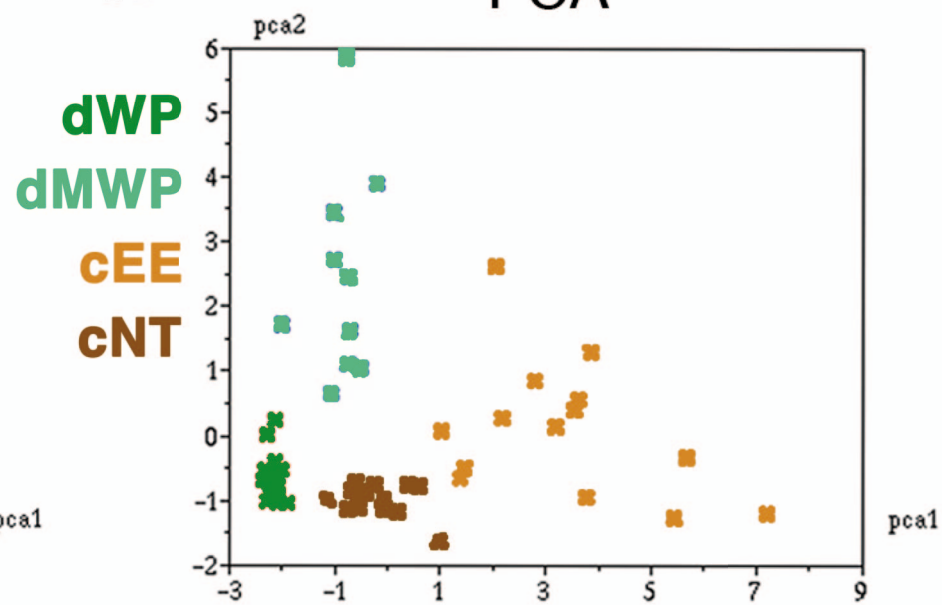dMWP

b

PCA

dWP
dNP
dMWP

c

PCA

dWP
dMWP
cEE

d

PCA

dWP
dMWP
cEE
cNT

Coefficient of Variation x100	average				standard deviation			
	A	D	C	N	A	D	C	N
dWP n=16	36.21	0.55	0.42	0.64	37.59	6.49	7.28	4.90
dNP n=12	40.25	0.47	0.86	0.96	34.75	4.19	6.86	7.01
dWL n=15	27.83	0.34	0.50	0.41	33.62	4.15	5.20	3.60
dMWP n=10	30.81	1.94	2.84	2.33	51.69	11.66	15.75	10.78
cNT n=16	18.59	0.64	2.77	1.53	24.30	8.93	12.70	10.74
cEE n=14	32.57	3.11	8.99	6.20	31.42	16.41	14.69	17.88

Area (A)   Degree of a node (D)   Clustering coefficient (C)   Average degree of neighbours (N)

```
Coefficient of Variation x100


            av_A  av_D  av_C  av_N        st_A  st_D  st_C  st_N

dWP   n=16  36.21 0.55  0.42  0.64        37.59 6.49  7.28  4.90
dNP   n=12  40.25 0.47  0.86  0.96        34.75 4.19  6.86  7.01
dWL   n=15  27.83 0.34  0.50  0.41        33.62 4.15  5.20  3.60
dMWP  n=10  30.81 1.94  2.84  2.33        51.69 11.66 15.75 10.78
cNT   n=16  18.59 0.64  2.77  1.53        24.30 8.93  12.70 10.74
cEE   n=14  32.57 3.11  8.99  6.20        31.42 16.41 14.69 17.88
```