Manuscript for

Topological properties accurately predict cell division events and organization of *Arabidopsis thaliana's* shoot apical meristem

Timon W. Matz[1,2], Yang Wang[3], Ritika Kulshreshtha[3], Arun Sampathkumar[3], Zoran Nikoloski[1,2,*]

[1]Bioinformatics, Institute of Biochemistry and Biology, University of Potsdam,14476 Potsdam, Germany; [2]Systems Biology and Mathematical Modelling, Max Planck Institute of Molecular Plant Physiology, 14476 Potsdam, Germany; [3]Plant Cell Biology and Microscopy, Max Planck Institute of Molecular Plant Physiology, 14476 Potsdam, Germany

*** Correspondence:** nikoloski@mpimp-golm.mpg.de

This PDF file includes:

Main Text
Figures 1 to 5
Supplementary Figures 1 to 9
Supplementary Table 1 to 4

Abstract

Cell division and the resulting changes to the cell organization affect the shape and functionality of all tissues. Thus, understanding the determinants of the tissue-wide changes imposed by cell division is a key question in developmental biology. Here, we use a network representation of live cell imaging data from shoot apical meristems (SAMs) in *Arabidopsis thaliana* to predict cell division events and their consequences at a tissue level. We show that a classifier based on the SAM network properties is predictive of cell division events, with validation accuracy of 82%, on par with that based on cell size alone. Further, we demonstrate that the combination of topological and biological properties, including: cell size, perimeter, distance, and shared cell wall between cells, can further boost the prediction accuracy of resulting changes in topology triggered by cell division. Using our classifiers, we demonstrate the importance of microtubule mediated cell-to-cell growth coordination in influencing tissue-level topology. Altogether, the results from our network-based analysis demonstrates a feedback mechanism between tissue topology and cell division in *A. thaliana*'s SAMs.

Summary statement

we use a network representation of live cell imaging data from SAMs in Arabidopsis thaliana to predict cell division events and their consequences at a tissue level.

Introduction

The adjacency of cells, specifying the tissue topology, defines the organization of cells and affects function of organs in multicellular organisms. Therefore, deciphering the organizational principles of cellular connectivity networks are fundamental to improve our understanding of the development of multicellular organisms. The shoot apical meristem (SAM) of plants is a highly organized structure composed of continuously proliferating cells that differentiate and give rise to all aerial organs and is under the control of an intricate signaling network influencing plant growth and response to different stimuli. The SAM epidermis in plants serves as an excellent system to identify organizational principles of cellular connectivity networks (Varner and Lin, 1989).

Since the cells in the SAM are glued to each other by a rigid cell wall, changes in the topology of SAMs are only brought about by cell division events. Cell division in plants is a cell-size-dependent, cell autonomous process (Jones et al., 2017), and crossing multiple checkpoints allows the final transition towards cell division (Veylder et al., 2007; Qi and Zhang, 2019). Willis et al. (2016) recently showed that initial cell size at birth influences the increase in size (sizer model), even though there seems to also be a component of constant size increase (adder model) in the shoot apical meristem (SAM) of *Arabidopsis thaliana*. This study has hinted at the possibility that a combination of both models may best describe cell division (see D'Ario and

Sablowski (2019) for a comparison of models). Although size-dependent cell division seems to be independent from position and cell to cell contact (Willis et al., 2016), recent study of Jackson et al. (2019) has pointed out that dividing cells display higher centralities in the network representation of the *A. thaliana*'s SAM; however, this observation was not sufficient to accurately predict cell division from network properties alone.

Since biochemical and physical signals are transmitted across tissues and affect cell division, growth, and morphology in a spatio-temporal fashion, the question arises of how tissue topology could influence such processes to help the plant respond to a variety of stimuli. In the context of physical signals, the ability of plant cells to respond to growth driven mechanical signals requires the activity of microtubule severing protein KATANIN (Uyttewaal et al., 2012). It has been shown that the lack of mechanical feedback, as in the *katanin1-2* mutant, results in changes to the topological features as a consequence of modified cell shape (Jackson et al., 2019). Therefore, this mutant can be employed to test if topological features are indeed relevant for cell division and related processes.

This question can be readily addressed due to the availability of plant lines expressing stable fluorescence reporters that allow for monitoring cellular outlines in combination with confocal imaging techniques (Reddy et al., 2004). In addition, the combination of user-friendly tools for accurate segmentation, like MorphoGraphX (Barbier de Reuille et al., 2015), with different machine learning (Bhavsar and Panchal, 2012; Pisner and Schnyer, 2020) and deep learning techniques (Camacho et al., 2018) has led to massive advances in the analysis of high-throughput imaging data. Further, the analysis of imaging phenotypes has been facilitated by adopting the network paradigm (Breuer et al., 2017; Nowak et al., 2021). To this end, topological features have been employed in cell wall placement models for dividing cells, by using the degree (i.e. number of neighbors) in combination with a spring based model (Gibson et al., 2011) or other individual topological features (Jackson et al., 2019). It has been shown that some of these individual topological features can better predict the placement of certain cell walls compared with more traditional approaches (Jackson et al., 2019), such as: dividing cells using the shortest wall placement, generalized Errera's rule (Besson and Dumais, 2011) or by minimization of tensile stress in other models (Louveaux et al., 2016). Although these models present an important step to solve the problem of cell wall placement, each model underperforms on some cells in the central region of the SAM (Shapiro et al., 2015; Jackson et al., 2019).

Although there are attempts of combining network properties with imaging data from SAMs, minor progress has been achieved in predicting individual cells divisions in this plant tissue. Here, we provide a network-based perspective to model cell division and cell wall placement in the SAM of *Arabidopsis thaliana*, a well-established system for studying cell division. To this end, we combine network-based analysis of live cell imaging data with classifiers that allow us to simulate

tissue-wise topological changes of the *A. thaliana* SAMs and test these classifiers independently on the *katanin1-2* mutant.

Results

Topology and surface area accurately predict cell division events

The question of whether division of a cell embedded in a tissue is driven by the topology of the neighboring cells, the area of the cell, or combination of the two is still open. To address this question, we imaged SAMs of five *A. thaliana* expressing a plasma membrane reporter (pUBQ10:acyl-YFP) every 24 hours over five days using confocal microscopy (Figure 1A). First, we manually determined the number of dividing and non-dividing cells between two consecutive time points in the central zone of the SAM. We defined the central zone of a SAM as the area covered by a circle of 30 µm radius around the highest point in the analyzed SAM (Figure 1B), and found that 24.3%±3.5% of cells divided per tissue between two successive time points, with a total number of 329 dividing cells and 896 non-dividing cells (Figure 1B).

Next, we represented the topology of the central zone as a network, in which every node corresponds to a cell and two nodes are connected by an edge if the cells share cell wall. For each cell we calculated 16 properties, referred to as topological features (Supplementary Table 1), in an unweighted network, in which every edge is of weight 1. We also applied different edge weights based on the mean surface area, shared cell wall, and distance of the cell centroid between two nodes representing those cells (Figure 1C). In addition, we considered the surface area of each cell in the central zone as a biological feature.

Previous studies have shown that there exists a critical cell size threshold for cell division in the SAM of *A. thaliana* (Jones et al., 2017). To show that topological features capture information distinct from that provided by the cell surface area, we calculated its Pearson correlation with the topological features (Supplementary Figure 2, Supplementary Table 2). Using the network with edges weighted based on the cell surface area, we found that betweenness centrality, a measure for the relative number of shortest paths passing through a node, exhibited the highest correlation of 0.71 to the surface area. Nevertheless, the absolute value of the correlation with surface area was smaller than 0.5 for 63% of the features (97% of features showing correlation smaller than 0.7). Therefore, topological features in the considered network scenarios carry information that is different from that obtained by the cell surface area alone. To further show the predictive power of the classifiers trained on the topological features, we considered two reduced feature sets that only included features with absolute value of the Pearson correlation coefficient (r) smaller than 0.5 and 0.7, respectively (Supplementary Figure 2). In such a way, we aimed to remove bias due

to consideration of features which may, to a certain extent, include information about surface area.

As a result of these considerations, we trained six classifiers based on non-linear support vector machines (SVMs) with Gaussian kernel ((Bhavsar and Panchal, 2012)) to predict cell division based on: all topological features (topo), surface area alone, the combination of topological features and surface area (topo + area), topological features with low absolute value of correlation with surface area ($r < 0.7$ and $r < 0.5$), and on unweighted topological features. To this end, we selected an equal number of dividing and non-dividing cells from four SAM for training the SVMs, to ensure balancing of cell labels. We kept the data from the remaining, fifth SAM, as a testing set (Methods). Further, we partitioned the 502 selected cells into training and validation sets composed of equal numbers of dividing and non-dividing cells, and used five-fold cross validation to train the classifiers (see Methods).

While the training accuracy of the SVM using only the surface area was 79.4%, the training accuracy solely based on topological features was significantly higher, at 88.7% (11.0% higher; p-value=0.0011, one-way ANOVA); this was also the case when the combined set of topological features and surface area was used, with training accuracy of 86.3% (8.3% higher; p-value=0.0101, one-way ANOVA). However, we observed no difference in the validation accuracies for the three types of SVMs (~81%). For the test SAM, the classifier based on the combination of topological features and surface area exhibited the best performance, with an accuracy of 78.9%, followed by the SVM that considered the topological features (76.9%) and the surface area (72.4%) alone (see Figure 2, Supplementary Table 3). The area under the curve (AUC) of the receiver operating characteristic (ROC) - curve, used as another measure of performance, showed similar trends (Supplementary Figure 3A, Supplementary Table 3).

The removal of topological features that were highly correlated with area does not significantly change the validation accuracy (Figure 2, Supplementary Table 3). Moreover, using only the topological features from the unweighted network scenario (Figure 1C), resulted in 16.7% smaller accuracy of the classifier on the validation set in comparison to that based on all topological features (p-value < 0.001, one-way ANOVA). However, the classifier based on the features from the unweighted topology only performs slightly worse (relative difference in accuracy of 5.5%) compared with the classifier based on the surface area (Figure 2, Supplementary Table 3). Inspection of the learning curves showed that the classifiers did not suffer from high bias and variance and that the training set was sufficiently large (Supplementary Figure 4). Therefore, we concluded that topological features led to marginally improved performance in predicting cell division compared to the surface area alone, while the performance could be further increased by the combination of both feature types.

To further corroborate the biological relevance of these findings, we randomly permuted the labels and retrained the classifiers, repeating this procedure 1000 times for each feature set (see Methods). All but the classifier based on surface area were able to partially predict training data. However, no classifier was able to generalize on the validation or test set, exhibiting accuracies expected by chance (Supplementary Figure 5A). Therefore, the classifiers trained on the randomized labels demonstrated that the used features capture information important for classification of dividing and non-dividing cells in the 1-day long intervals.

Independent testing of the trained SVMs with data from the *katanin1-2* (*ktn*) mutant to predict cell division events showed reduced accuracy in comparison with data from the wild type (WT) for all classifiers. The relative difference between the test accuracies for WT and *ktn* was smaller for the classifier trained on surface area (2.0%) compared with all topological (6.0%) as well as the topological and biological features combined (7.2%). Removing the features with Pearson correlation to surface area higher than 0.7 led to relative decrease of 6.0% between the test accuracies for WT and *ktn* (Figure 2). While these findings demonstrate the importance of surface area as a determinant of cell division, they also support the claim that topology plays an important role in predicting cell division events.

Combination of topological and biological features enables recreation of the local topology after cell division

To examine whether properties derived from the tissue connectivity network as well as biological properties (i.e. cell size and perimeter, distance, and shared cell wall between cells) are predictive in the time-dependent connectivity of daughter cells, we trained classifiers based on SVMs (with Gaussian kernel) to predict which of the cells adjacent to a dividing (parent) cell are neighbors of the divided (daughter) cells. We distinguished neighbors that were only adjacent to one daughter cell: adjacent to the daughter cell closer to the SAM center are labelled as class 0, while those adjacent to the daughter farther from the center were classified as class 1; neighboring cells adjacent to both daughter cells were considered to be of class 2 (Supplementary Figure 1C).

To predict changes of cell divisions based on local topology from the data collected at 1-day time interval, we first determined all neighbor-parent pairs and then predicted the adjacency of the neighbor to the daughter cells. To this end, we considered the topological features as well as biological properties of the parent and the neighbor cells in the preceding time point. We used the topological features of the neighbor cell to distinguish neighbor-parent pairs in which a neighbor is adjacent to two dividing cells. We also included the difference between the topological features of the neighbor and the parent, thus generating 100 unique topological-based properties for each neighbor-parent-pair (Supplementary Figure 1B). For the biological feature set, we extracted the

surface area and perimeter of both the parent cell and neighbors as well as their shared cell wall and distance between the centers of mass. For the combined features, we concatenated both topological and biological features. Given a parent cell, we determined the class of its neighbors in the next time point by aligning the tissues between 1-day time intervals manually, and determining their adjacency of the neighbor with respect to the daughter cells (see Methods and Supplementary Figure 1B).

We excluded neighbor-parent-pairs for which the neighbor also divided in the considered 1-day interval to avoid bias due to guessing which cell divides first (Figure 3A). Following this procedure, we created 1638 neighbor-parent pairs with 546 representatives (balanced classes) in each of the three classes (0, 1, and 2) from five different SAMs, tracked every 24 hours over 5 days. The data was split into three parts: training, validation, and test data, such that the SAM of one plant was kept as test data, while the rest of the plants were used in a nested five-fold cross validation for training the SVM.

The training and validation accuracy was best for the SVM based on the topological features combined with biological features, at 70.4% and 64.0%, respectively. The topological and biological features alone showed 12.7% or 17.0% reduction in validation accuracy compared to the combined classifier, and similar reduction in training accuracies. Regarding the performance on the test set, the combined classifier performed best, with an accuracy of 65.7%, followed by the classifier based on topological features alone with 60.6%, and that using biological features alone, with the worst accuracy (equivalent to guessing) of 50.3% (Figure 3B, Supplementary Table 4).

Investigating the area under the ROC-curve (AUC) measure for individual classes highlighted the differences between the two classifiers trained on topological or biological features alone: The SVM based on the topological features showed better performance for the neighbors only adjacent to one cell (class 0 and 1) in comparison with the classifier based on the biological features (i.e. relative increase of 10.7% and 23.9% for class 0 and 1, respectively). In contrast, the SVM based on the biological features performed 22.6% better for neighbors adjacent to both daughter cells in comparison with the classifier based on the topological features. Combining both feature sets improved the average AUC on the validation data of the classifier by 9.4% and 13.0% (relative increase compared to topological and biological features alone, respectively) while retaining high performance for all classes (Figure 3C, Supplementary Figure 6). Investigating the reduced topological feature set (i.e. removing features with Pearson correlation coefficients larger than 0.5 or 0.7 with any biological feature) as well as only considering unweighted features resulted in almost identical training, validation, and test accuracies compared with all topological feature trained classifiers. The training accuracy of the unweighted set showed a slight relative reduction of 7.0% (p-value: 0.013, one-way ANOVA) (Figure 3A,

Supplementary Table 4). These findings indicated the importance of both topological as well as biological properties to predict local topology after a division event.

To further corroborate the biological relevance of these results, we randomly permuted the labels and retrained the classifiers, repeating this procedure 1000 times for each feature set. While the resulting classifiers showed performance better than expected at random on the training data with the three sets of features, they did not generalize well and exhibited accuracy on the validation set similar to expected by chance (Supplementary Figure 5B). Further, we investigated the more difficult scenario of including neighbor-parent-pairs whose neighbors also divide and repeat the topology prediction procedure. Here, we found similar performance to that on the training, validation, and test sets for all combinations of feature sets (Supplementary Table 4). Therefore, our findings demonstrated that the used features capture information important to classify changes in local topology predictions surrounding dividing cells in 24-hour intervals.

We tested how well the trained classifiers based on the wild-type data performed on the *ktn* mutant. With the data from the *ktn* mutant, we found a reduction in accuracy for all classifiers for local topology prediction, trained on the wild type data, except for those using the biological feature set (50.3% WT testing vs 51.6% *ktn* testing). All classifiers trained on topology-related feature sets showed strong relative reduction in test accuracy between WT and *ktn* (topo: 31.6%; topo + bio: 17.0%; $r < 0.7$: 33.3%; $r < 0.5$: 28.9%; unweighted: 25.5%; orange vs purple bars, Figure 3B). The reduced performance on *ktn* data of classifiers trained on topological features (feature sets: topo, topoAndBio) can be mainly attributed to the worse prediction of class 0 and 1 (cells being predicted to be adjacency to only one of the divided daughter cells, Supplementary Figure 1C). In contrast, the classifiers trained on biological features performed similarly on *ktn* test as those based on the WT test data (Figure 3B). These results highlight the importance of both topological and biological information in local topology rearrangement after cell division.

Combined application of division event and local topology prediction enables to predict tissue topologies

To apply the classifiers and compare the resulting topologies, we used the data from the test plant and successively predicted division events and changes in local topology using classifiers trained on the combined biological and topological features (see following procedure in Figure 4A). We compared the predicted and observed topologies by investigating the unweighted topological features (Figure 1C, Supplementary Table 1) of non-dividing cells in the next time points of both scenarios. We selected non-dividing cells of both scenarios, i.e. predicted and observed, to pairwisely compare their unweighted topological features. We did not consider other network scenarios (see Figure 1C) since we would need to estimate the weights for the topology, adding a layer of uncertainty. Here, 126 of the possible 155 non-dividing cells in the observed

topology were also non-dividing in the predicted topology. For these cells, we calculated the Pearson correlation coefficient (r) of all unweighted features between observed and predicted topologies, with the harmonic centrality showing the largest value of r = 0.80 and eight of 16 features having Pearson correlation larger than 0.5 (Figure 4D, blue bars). We compared the predicted and observed values of harmonic centrality of the non-dividing cells of the next time step, and found strong correlation (Figure 4E, Supplementary Figure 8).

For comparison, given the same test plants observed topologies, we selected all predicted non-dividing cells to divide and randomly connected the neighbors with the divided cells reorienting their local neighborhood; we then repeated the correlation analysis of the resulting topology with the observed one. This “random propagation” scenario allowed us to construct and investigate the most opposite example to our predictions (Figure 4D, orange bars). Comparing the predicted and random propagations correlations shows that only five out of 16 topological features showed higher correlation in the random propagation. The random propagation showed the lowest correlating features and the trained classifiers showed the highest correlations with a total of eight being higher (Figure 4D, blue vs orange bars).

To further investigate the performance of the local topology prediction on the test plant, we calculated the percentage of correctly predicted neighbors for each cell dividing in the predicted and observed tissue (example in Figure 4B). The distribution of correctly labelled neighbors per dividing cell was significantly shifted towards higher accuracy when comparing the predicted and random topology predictions (Figure 4C).

Discussion

The biochemical pathway of cell division control has been extensively studied (Dewitte and Murray, 2003), but only recently external cues have also been considered to understand the effect of cell divisions in a tissue context (Hartig and Beck, 2006; Shimotohno et al., 2021). It has been known that the outer epidermal cell wall resists most forces (Beauzamy et al., 2015) and, thus, division in the SAM outer-layer needs to serve both meristematic functions. This raises the question if cell division and their subsequent local topology rearrangement are affected by the tissue topology and if tissue topology contains sufficient information for their accurate prediction.

Based on our extensive network-based modelling, we showed that both surface area, as an approximation of cell size, as well as the characteristics of topology allow for prediction of cell division events in the central epidermal region of *A. thaliana* SAM, in contrast to earlier reports (Jackson et al., 2019). The cyclin-dependent kinase (CDK) G1 is known to bind DNA and serves as a ruler after cell division, allowing for size dependent division in *C. reinhardtii* (Li et al., 2016), while KIP-related protein 4 has a similar function in the *A. thaliana* SAM niche (D'Ario et al., 2021). Modelling cell division in the SAM of *A. thaliana* also revealed the importance of CDKs in

G1-S and G2-M phase transition (Jones et al., 2017). Further, the work of Willis et al. (2016) showed that cell division events in SAMs of *Arabidopsis* treated with naphthylphthalamic acid, an inhibitor of auxin transport that generates naked meristem, are influenced by both cell size increase and a cell size threshold affecting cell division. Both models explain the importance of surface area in predicting cell division events, but they do not explain the importance of topological features. Here, the mechanical feedback loop, envisioned by the cells ability to react to changes in turgor pressure with MT and cell wall rearrangement affecting cell divisions (for detail of the feedback loop, see Sampathkumar (2020)) may serve as an explanation linking topology with the summed turgor and supracellular mechanical stress. Alternatively, the predictive ability of topological properties may result from long distance communication by different phytohormones (Shimotohno et al., 2021), or due to cell-to-cell communication by plasmodesmata (Kitagawa and Jackson, 2017).

However, not only the cell division, but also the cell wall positioning affects the tissue organization; a prime example is the effect of division patterning in lateral root initiation (Wangenheim et al., 2016). Our study relies on the adjacency of cells in the tissue topology, in contrast to other cell wall models, such as: the generalized Errera's rule (Besson and Dumais, 2011), the spring-based model (Gibson et al., 2011), and the mechanical stress related model (Louveaux et al., 2016), that predict the placement of the cell wall based on the individual cell geometry. Our classifier employs the biological feature set composed of six cellular features, having limited information about the dividing and neighbor cell geometry, and allows for reliable prediction of the changes in the local topology. These local changes in the topology mirror the effect of the cell wall placement on the tissue. In addition, we showed that topological features alone sufficed to accurately predict local topological changes. While single topological properties were already used to estimate cell wall placement (Jackson et al., 2019), the percentage of dividing epidermal cells in this study was only 12% (total n=7/57 dividing and non-dividing cells) per tissue every 22h. In contrast, our results rely on experiments in which cells divided more regularly, with an average of 24% of dividing cells per tissue every 24h (total n=329/896 dividing and non-dividing cells), allowing us to train robust classifiers. We showed that the combination of both feature sets boosted performance of local topology reorientation prediction (Figure 3), indicating that the inclusion of multiple viewpoints of information available to cells needed to be involved to solve the problem of cell wall placement in the SAM. This raises the question how information of the topology is biologically transferred to cells, either via mechanical stress, hormones, or cell to cell communication with plasmodesmata.

To demonstrate the generalizability of the classifiers, we showed that they can be used to make accurate predictions for *ktn* mutants that are defective in mechanical feedback regulation. Our results indicated similar performance for the classifiers with biological feature sets from WT

and *ktn*. In contrast, the classifiers trained on topological features showed reductions in performance in *ktn* compared to WT. This difference in performance is not due to differences in topological features, since the normalized features showed similar distributions (Supplementary Figure 9). These results suggest a potential role of KATANIN in linking sub- and supracellular mechanical stress, known to affect leaf epidermal cells (Eng et al., 2021) and KATANINs role in positioning of the preprophase band, spindle, and phragmoplast (Komis et al., 2017). In addition, the cell geometry of the *ktn* mutant differs compared to the WT and might also influence the topology. Therefore, the combination of network-based modelling with machine learning provides a method to screen SAMs under different conditions and mutants. More specifically, reduction in test performance of either the classifiers trained on surface area or on topological features compared to the wild type may hint to effects only disturbing function related to the cell cycle or to a topological effect (in the case of classifier trained on area or topological features being lower, respectively).

When combining division prediction and the resulting changes to the tissue, previous studies mostly focus on single cell division or propagating tissues based on division likelihoods using the number of neighbors (Gibson et al., 2011) or just using area as a fixed threshold (Sahlin and Jönsson, 2010; Alim et al., 2012), while our classifier incorporates more diverse tissue-level information. Here, we combined our best classifiers to predict future tissue topology using the combined topological and biological features. Although the results of this propagation of classifiers is promising, the careful inspection of the finding, particularly with respect to planarity and topological properties of the reconstructed topologies point out that further research should consider simultaneous modelling of cell neighborhoods of higher order to improve the reconstruction.

Furthermore, as information is not only be passed along the epidermis (L1-layer), the assessment of cell division events and their changes on the topology could be expanded beyond the epidermis of the SAM as we know that the L2- and L3-layer play a vital role in supporting the meristematic function through the feedback of CLAVATA 1, 2, and 3 and WUSCHEL (Schoof et al., 2000). Transferring the classifiers to other plant species, such as maize (that has only two distinct layers forming the SAM), may provide insights into how meristematic function can be conserved with fewer cells. As other tissues and organs are also experiencing mechanical stresses, hormone gradients and other transport related feedbacks, e.g. growth resulting stress (Sampathkumar et al., 2014), auxin gradients (due to PIN; Shi et al., (2018)), soil thickness in roots, and bending through wind in the stem, there are bound to be feedback loops of cells and tissues to sense and react to those cues on a topological level to integrate this information into the plants development.

Material and Methods

Plant materials and growth condition

We grew *Arabidopsis thaliana* wild-type (WT; Wassilewskija ecotype) plants with the membrane reporter pUBQ10::acyl-YFP (previously described in (Willis et al., 2016)) and katanin1-2 mutant in Columbia-0 background with the membrane reporter Lti6b-GFP (Eng et al., 2021) in short day (8 h/16 h day/night), 20 °C/16 °C conditions for 3 weeks and then transferred to long day (16 h/8 h day/night), 20 °C/16 °C conditions till shoot apical meristem sampling. We cultured sampled shoot apical meristems (SAMs) in transparent imaging boxes containing apex culture media under long day, 22 °C conditions as previously described (Wang and Sampathkumar, 2020).

Time-lapse data acquisition and pre-processing

We acquired confocal Z-stacks (3D images) at an excitation wavelength of 514 nm and 488 nm for imaging YFP and GFP respectively with a 40X/0.8 water immersion objective every 24 hours for 5 days (WT) or 3 days (*ktn*). Next, we used MorphoGraphX (MGX) (Barbier de Reuille et al., 2015) to obtain 2 ½ D surface mesh of the meristem L1 layer from the 3D images and from there we extract the cellular connectivity network (topology). In addition, we measured the shared cell wall of the neighboring cells (MGX function: Mesh/Export/Save Cell Neighborhood 2D), the surface area, and cell positions (MGX: Mesh/Heat Map/Heat Map Classic). The cellular connectivity network is composed of nodes, representing the centroids of the extracted cells. Edges connect two nodes if the corresponding cells are adjacent to each other. We lineage-tracked all cells between 1-day time steps manually in MGX (Figure 1A). We refer to dividing cells, at time t (days), as parent cells and their descendants, at time t + 1 (days), as daughter cells. To select the cells for the downstream analysis, we first manually determined the cells closest to the center of the SAM surface, given by the highest curvature. To this end, we compared the positions of cells to the average position over all cells.

Prediction of dividing cells

To predict cell division events of central and non-peripheral cells, we selected all cells in a radius of 30 µm around the center (Figure 1B). In such a way, we only analyzed central and exclude peripheral cells. We considered a cell as peripheral with respect to a connectivity network in case the graph induced by the adjacent nodes does not form a cycle. We then labelled each of the selected cells as dividing (label: 1) or non-dividing (label: -1) cell within 24 hours (one day). In addition, we determined six sets of features (see below; for unreduced sets see Supplementary Figure 1) for each cell.

Five of six feature sets are based on the entire tissue (i.e. including peripheral cells as well as cells outside of the central region from the cellular connectivity network) and consist of topological features for all central cells; the sixth set includes only the surface area of the central cell. While calculating the topological properties, we considered different scenarios for weighting the edges. In the case of the unweighted topology, we weighted all edges equally (edge weight = 1). For the area-induced topology, we used the inverse of the mean surface areas of the two adjacent cells as edge weight. For the wall-induced topology, we defined the edge weights as the inverse of the shared cell wall area between two cells. For the distance-induced topology, we determined the inverse of the distance between the centroid positions of two adjacent cells as the edge weight (Figure 1C). We calculated ten topological properties for each central cell and network scenario (see Supplementary Table 1). Furthermore, we considered topological properties based on the induced subgraph of the first neighborhood (see Supplementary Table 1). We estimated all properties in python 3.8.1 using the networkx 2.4 package.

To train the classifiers for prediction of division events between two successive time points, we split the WT data from the five plants into two data sets, a training-validation and a testing set with four and one plant, respectively, while keeping three *ktn* plants as a separate test set. As there are fewer dividing cells, their class is the minority class. As a result, we down-sampled the majority class of non-dividing cells to balance the two classes. We applied a support vector machine (SVM) with a Gaussian kernel to predict the occurrence of cell division events within 1 day. To this end, we used the six different feature sets, namely: the unweighted topological features (unweighted topology), all topological features combined (topo), the surface area (surface area), topological features and area (topo + area), as well as two reduced feature sets including only topological features with Pearson correlation coefficients with surface area smaller than 0.5 or 0.7 (denoted by $r < 0.5$ and $r < 0.7$) (Supplementary Figure 1A). We trained each classifier with the topological properties as features of the training-validation set using five-fold cross-validation.

To this end, we z-normalised ((X-mean)/std) the topological properties with the corresponding mean and standard deviation (std) for the train-validation and the WT test data sets, respectively. The *ktn* data is z-normalised using its mean and standard deviations. We estimated the hyperparameters on the training set using another five-fold cross-validation using grid search (sklearn 0.22.1, GridSearchCV) regularly spacing 50 hyper-parameters for each power of 10. We further tested the classifiers by retraining the SVMs on all training-validation data with newly selected parameters and applied them on the unseen test data. We quantified the performance of the classifiers by calculating five measures of performance, including: the accuracy, F1-score, true positive rate, false positive rate, and area under the curve (AUC) of the receiver-operator

characteristic (ROC). For comparative analysis between two performance measures p1 and p2, we used the relative difference ($100 \cdot \frac{|p_1 - p_2|}{\mathrm{mean}(p_1, p_2)}$).

To further inspect the training of the classifiers, we generated the learning curves by retraining each classifier on a different number of training data (keeping the hyper-parameters from above). We further determined the feature sets information content by shuffling the labels 1000 times, retraining the classifier using the default RBF SVM parameters (sklearn 0.22.1, svm.SVC) on each set of shuffled labels, and calculating the performance of the resulting classifiers.

Recreating of local topology after cell division

For the prediction of the changes in local topology of dividing cells, we selected all non-peripheral neighbor-parent-pairs of dividing cells. Next, we categorized the adjacency of these neighbors with respect to the newly divided (daughter) cells. To this end, we inspected if the neighbor of a neighbor-parent pair is adjacent to only one or both of the daughter cells.

To automate the procedure, we distinguished the divided daughter cells into the daughter closer to the center of the SAM which we termed cell “A” and the second daughter cell we named cell “B”. We labelled each neighbor cell in a neighbor-parent-pair with class 0, 1, or 2 according to whether it is connected only to cell “A”, cell “B”, or both. We then predicted the local topology excluding and including dividing neighbors using six feature sets, similar to the analysis above.

To distinguish neighbor-parent-pairs which are adjacent to two dividing cells, we considered the difference of topological features between neighbor and dividing parental cells in addition to the parental topological properties as features (Supplementary Figure 1B). As a result, we obtained the following feature sets: unweighted topology, topological features from all weightings (topo), biological features (bio, consisting of surface area and perimeter from neighbor and parent, as well as the shared cell wall and distance between the two), the combination of all topological and biological features (topo + bio), and two reduced feature sets ($r < 0.5$ and $r < 0.7$) including only topological features with Pearson correlation coefficients of smaller than 0.5 or 0.7 with all biological features. We performed training, validation, and testing as well as inspected the learning curves and estimated the information content of the used features as specified in the analysis above for wild type data. Additionally, we tested the classifiers on the *ktn* data.

Application of the classifiers for division event and local topology

To combine the predictions of division events and local topology changes, we used the previously developed classifiers and applied them to predict how the topology of the test plants would change. To this end, we selected the classifiers including both topological and biological features (based on validation performance) and applied them one after another on to the test tissues to generate the topology of the next time points. Here, the predictions were only made for one time

step (24 hours), since longer periods required us to estimate changes in the biological features as cells predicted to divide would not necessarily divide in the observed tissue one step later.

To arrive at the predicted cellular connectivity network, we determined the cells predicted to divide and divided cells' future adjacency with their neighbors. Next, we repeated the following four steps for all cells predicted to divide at time t, starting with a random cell: (1) We removed the dividing cell along with the edges connecting the neighbors that is dividing. (2) We added the daughter cells representing the cell closer (cell A) and farther (cell B) away from the SAM center. (3) We connected the daughter cells with their neighbors based on the prediction from the local topology classifier (Figure 4A).

To evaluate the performance of the combined application of division and topology prediction, we calculated all unweighted topology features for the cells which are neither dividing in the predicted nor in the observed topology. Next, we plotted the non-dividing cells observed against the predicted features, determined best linear fit, and the Pearson correlation coefficient of all unweighted topological properties. In addition, we divided all predicted non-dividing cells, randomly assigned labels to the neighbors of dividing cells how they will be connected to the divided cells based on the training-validation set representation. Then, we repeated the correlation analysis from above 1000 times (differently reconnecting topologies), and compared the correlations between predicted and random topology propagation.

To also evaluate the local recreation of the topology around dividing cells, we compare the first neighborhoods of cells dividing in the predicted and observed tissue of the test plant by calculating the percentage of correctly labelled neighbors. The distributions of predicted accuracies are compared with an estimated random labelling of the neighbors using Kolmogorov-Smirnov-Test (scipy 1.4.1, ks_2samp).

Code availability

The entire code to reproduce the findings is available at
https://github.com/matz2532/SAM_division_prediction

Acknowledgments

Z.N., T.M., A.S., Y.W., and R.K. acknowledge the support by the project SHAPENET, 031L0177A (to Z.N.) and 031L0177B (to A.S.), of the German Federal Ministry of Education and Research.

Author Contributions: Z.N. and A.S. designed the research. T.W.M. implemented the computational approaches and performed the computational experiments. Y.W. and T.W.M. performed segmentation. Y.W. and R.K. performed the image generation. T.W.M. and Z.N. prepared the manuscript. All co-authors contributed to the final version of the manuscript.

Competing Interest Statement: The authors declare no competing interests.

Figures

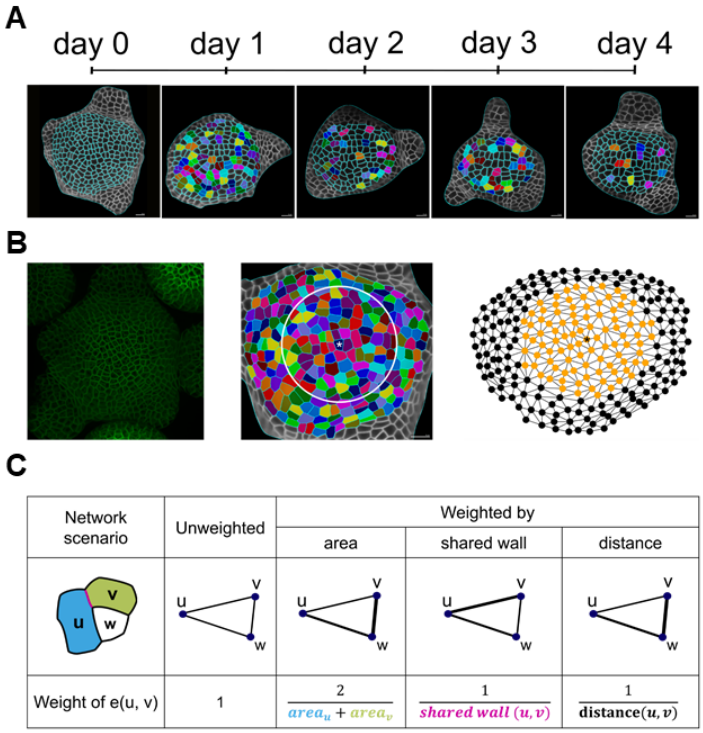


Figure 1. Feature generation from three-dimensional (3D) images of the shoot apical meristem (SAM). (A) The surface of A. thaliana SAM is imaged every 24 hours over four days. Pairs of dividing cells, depicted with the same colors, are determined manually (see Methods). (B) The 3D images of SAMs are converted to 2½D surfaces by employing MGX (de Reuille et al., 2015) (left panel). The surface is abstracted by its topology, capturing the connectivity of neighboring cells in a radius of 30 µm (white circle) around the central cell, marked with * (center panel). The topology of the analyzed cells inside the circle is colored in orange (right panel). Two nodes are connected by an edge if the cells they represent share cell wall. (C) Four different network scenarios are considered: (i) unweighted edges and edges weighted by (ii) area, (iii) shared cell wall, and (iv) distance, illustrated for the case of three cells u (blue), v (green), and w (white). In the unweighted network scenario, all edge weights have a value of one. The edge weight for the network weighted on area, shared wall, and distance is the inverse of the mean cell areas of u and v, of the shared cell wall area (magenta), and of the inverse distance of the center of mass for the graph weighted on the distance (black). The weights of the edge e(u, v) in the four scenarios are illustrated with different line widths.

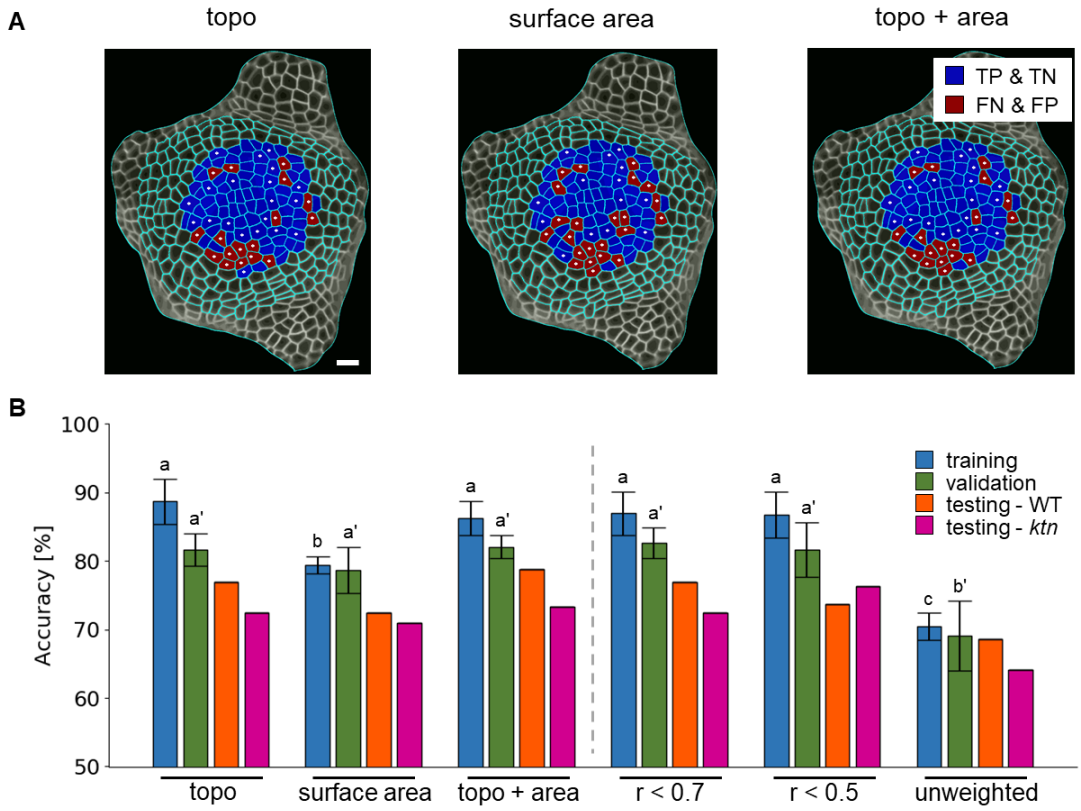


Figure 2. Surface area and topology-based features generate different predictions for cell division in the SAM. (A) Comparison of predicted and observed labels on one test plant tissue from Figure 1A day 0 highlighting the difference between predicted and observed division events. The predictions are from classifiers trained on different feature sets: combined topological features (topo), surface area, combined topological features and surface area (topo + area; from left to right) with the coloring scheme of correct predictions in blue and wrong prediction in red. The combined topological features include 16 centrality measures (see Methods) calculated based on the four network scenarios (see Figure 1C). Dividing cells are marked with a white star. Scale bar is 10 µm. (B) Accuracy of the support vector machine classifier to predict cell division for the training (blue), validation (green), and testing of wild type (orange) and *ktn* mutant (purple) data sets using: topo, surface area, and topo + area, reduced set of topological features that show an absolute Pearson correlation coefficient with surface area smaller than 0.7 or 0.5 ($r < 0.7$ and $r < 0.5$), as well as only the topological features derived from the unweighted network scenario (unweighted). The performance on the training and validation set is determined from five-fold cross-validation with mean and the standard deviation shown as error bars. Different letters indicate significance between groups using one-way ANOVA with Tukey's pairwise comparison: p-value < 0.05. Statistical testing for differences of classifier performance for the

556 training and validation sets was conducted separately (letter without and with apostrophe,
557 respectively). N_{WT} = 5 plants, 4 time steps (4 plants for training-validation and 1 plant for testing);
558 N_{ktn} = 3 plants, 3 time steps; n_{WT} = 502 and 156, train-validation and test cells respectively; n_{ktn} =
559 334 (balanced data).
560
561

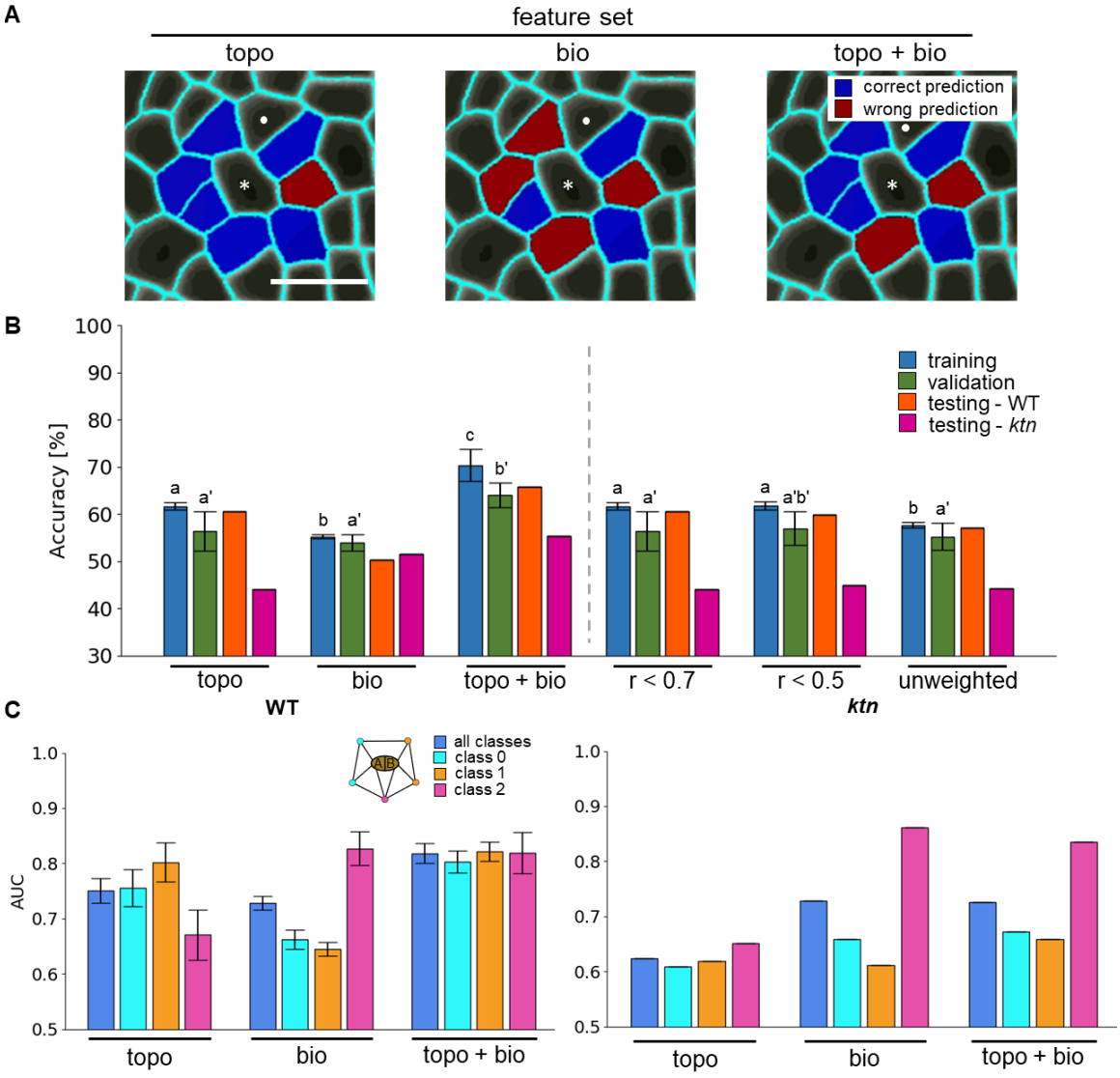


Figure 3. Topological and biological features are required for accurate prediction of the local neighborhood after cell division in the SAM. (A) Comparison of predicted and observed local neighborhoods on one local topology from Figure 1A, day 0. The predictions are made with classifiers trained on different feature sets: combined topological (topo, including features of the four network scenarios, see Figure 1C, Supplementary Figure 1), biological (bio, including area, perimeter, shared cell wall, and distance), and combined topological and biological features (topo + bio; from left to right) with the coloring scheme of correct predictions in blue and wrong prediction in red. The combined topological features include 16 centrality measures (see Methods) calculated based on the four network scenarios (see Figure 1C). The central and neighboring dividing cells are marked with a white star and circle, respectively. No color is displayed on the dividing neighbor cell as all dividing neighbors were removed. Scale bar is 10 µm. (B) Accuracy of classification on the training (blue), validation (green), and testing of wild

type (orange) and *ktn* mutant (purple) and (C) Area under the curve (AUC) of the ROC based on the wild-type validation (left) and *ktn* test data (right) for the classification of all classes (blue), for the class of neighbors adjacent to the daughter cell (cell A, see legend in B) closer to the SAM center (denoted as class 0; cyan), adjacent to the daughter (cell B, see legend in C) farther from the center (denoted as class 1; orange), or adjacent to both cells (denoted as class 2; magenta). The classifiers are based on topo, bio, topo + bio, reduced set of topological features that show an absolute Pearson correlation coefficient with all biological features smaller than 0.7 or 0.5 ($r < 0.7$ and $r < 0.5$), as well as topological features derived from the unweighted network scenario (unweighted). The performance on the training and validation set is determined from five-fold cross-validation with mean and the standard deviation shown as error bars. Different letters indicate significance between groups using one-way ANOVA with Tukey's pairwise comparison (p-value < 0.05). Statistical testing for differences of classifier performance for the training and validation sets was conducted separately (letter without and with apostrophe, respectively). N_{WT} = 5 plants, 4 time steps (4 plants for training-validation and 1 plant for testing); N_{ktn} = 3 plants, 3 time steps; n_{WT} = 1317 and 312, train-validation and test cells respectively; n_{ktn} = 888 (balanced data).

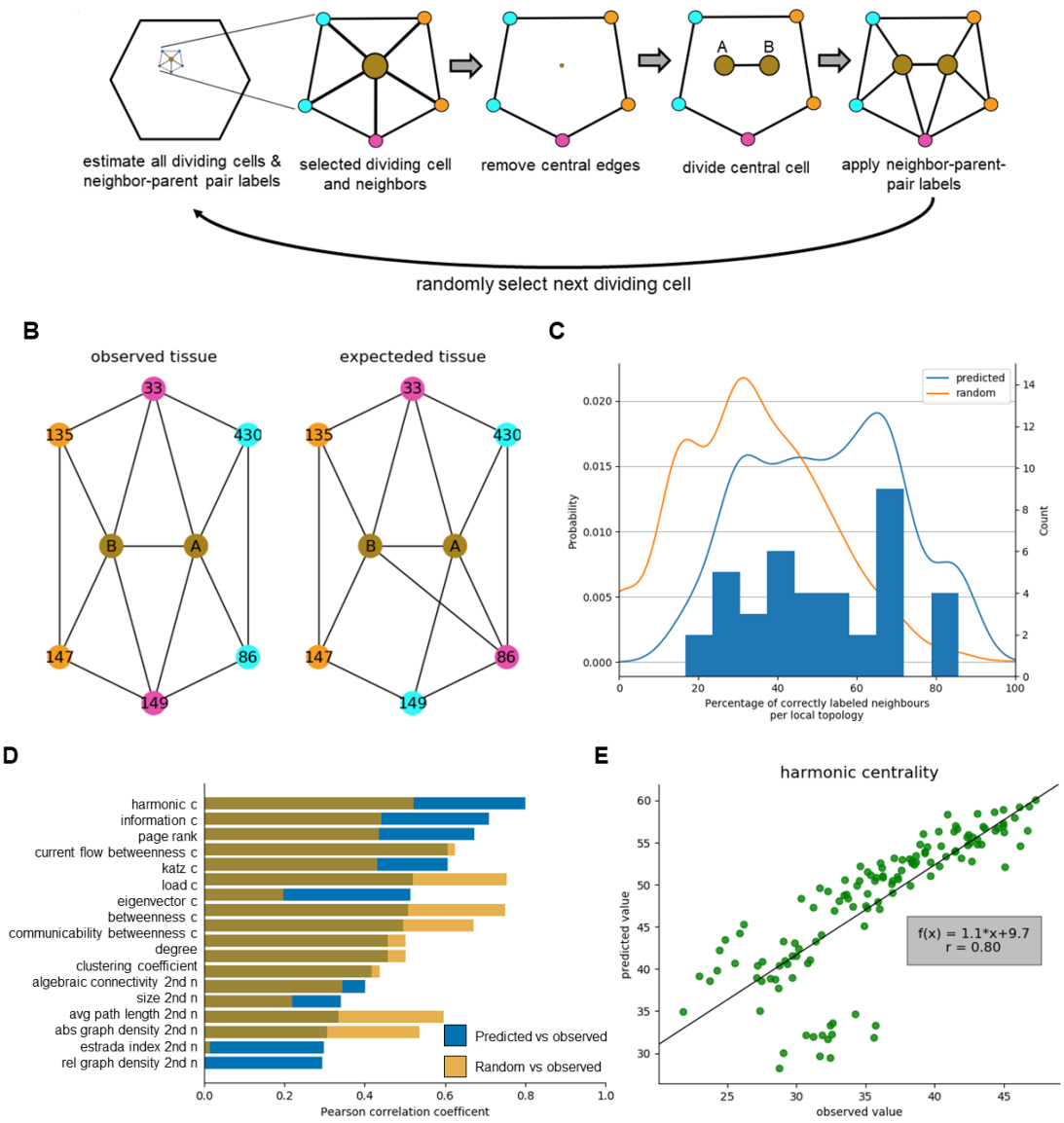
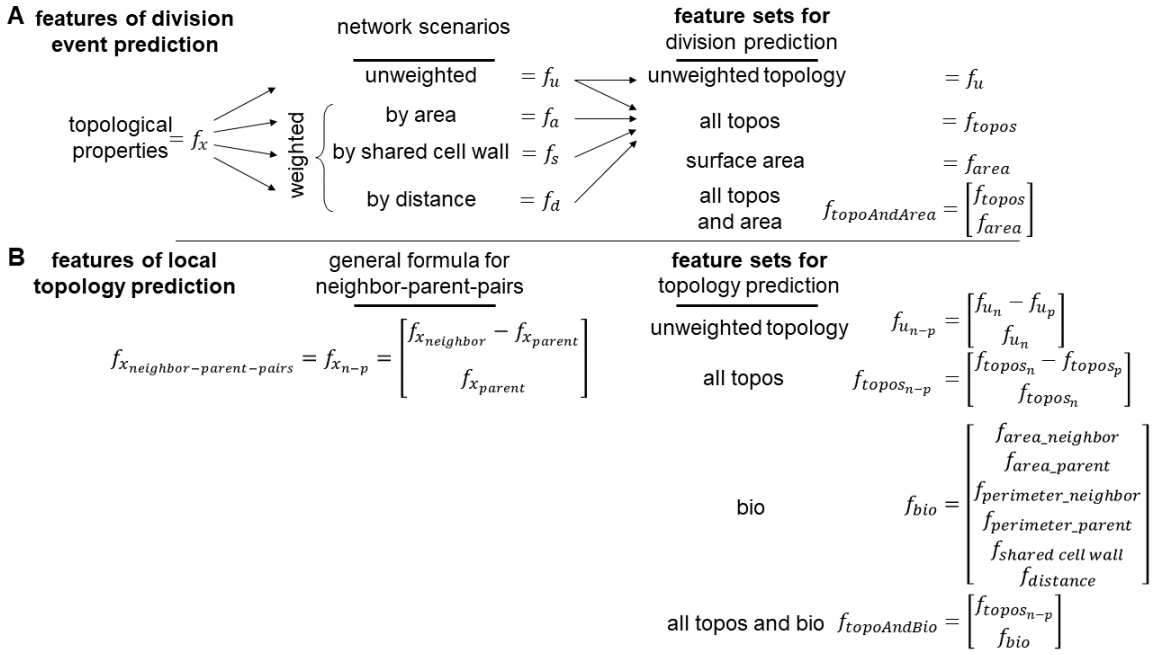


Figure 4. Concordance between observed and predicted topologies. The cell connectivity network of the test plant was predicted by applying classifiers for division event and topology prediction. (A) Illustration of the procedure applying the division and local topology classifiers to generate the topology of the next time point (24-hour time interval). After division and topology prediction, a dividing cell (dot in brown) is selected along with its neighbors (blue circles) and its adjacency relationship (edges, black lines) (left). The selected cell (predicted to divide) along with the edges incident to the corresponding node are removed and replaced by the divided daughter cells (A, B: representing the cell closer and farther away from the SAM center) that are adjacent to each other (three in the middle). The daughter cells are connected with their neighbors based on the prediction from the local topology classifier (right). The next dividing cell is randomly selected and the previous steps are repeated until all dividing cells are selected. (B) One example

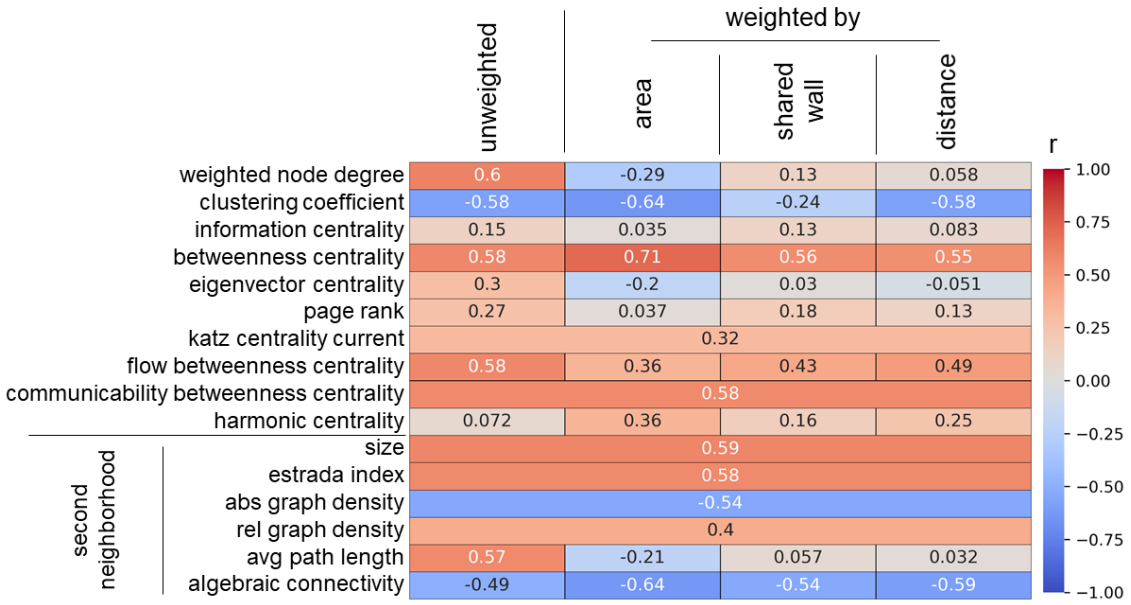
of the predicted local topology with an overall accuracy of 66.6% for the full local topology is compared with the observed local topology. The divided daughter cells “A” and “B” (brown) are adjacent with the predicted or observed cells (numbers display the same cells) coloring their respective parent-neighbor-class (cyan/orange: cell connected only with daughter A or B, respectively; magenta: cell adjacent with both daughter cells) (C) Histogram and density plot of the percentage of correctly estimated neighbors per local topology of cells predicted as dividing in the test plant (blue) are compared with the density plot of randomly assigning parent-neighbor-classes (orange). Difference between distributions is tested using Kolmogorov-Smirnov-Test, p-value < 0.01. N = 1 plant, 4 time steps, n = 39 local topologies. (D) The concordance between the observed and predicted topologies was quantified (blue) for non-dividing cells in both topologies by calculating and ranking the Pearson correlation coefficient based on 16 topological features from the unweighted networks (see Figure 1C; Supplementary Table 1). The procedure was repeated dividing all cells predicted to be non-dividing, randomly assigning classes to the neighbors, and calculating the correlation as described before (yellow). (E) The observed harmonic centrality is plotted against the predicted harmonic centrality for all non-dividing cells and the best linear fit (solid line) with its function f(x) and the respective Pearson correlation coefficient r is overlaid. N_{WT} = 1 plant, 4 time steps, n_{WT} = 126 cells not dividing in both observed and predicted tissue.

Supplementary Information

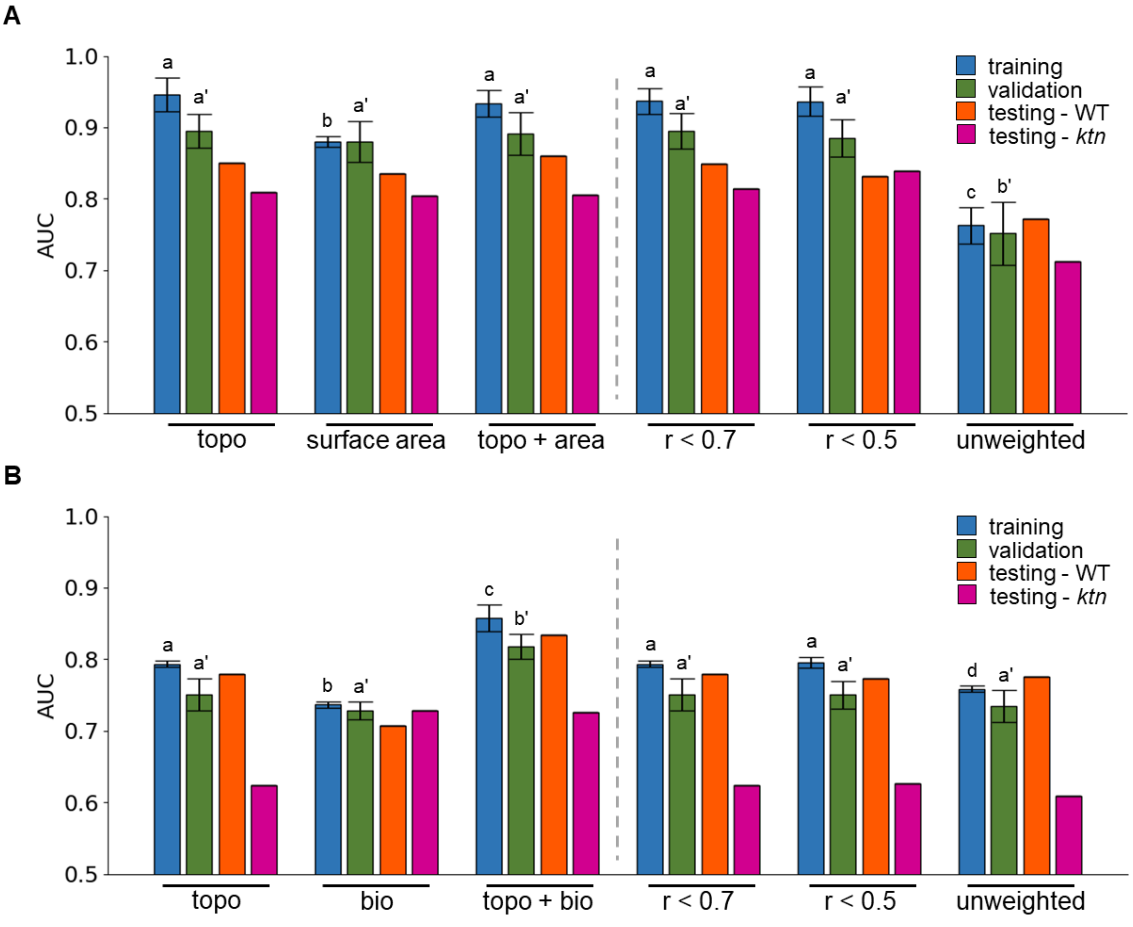


Supplementary Figures 1. Overview about feature sets for division event and local topology prediction as well as an example for local topology class assignment. (A) For division event prediction, we consider 16 topological features (Supplementary Table 1) calculated from four network scenarios (unweighted, weighted by area, shared cell wall, and distance), creating four different feature sets: unweighted topology, all topologies (all topos), surface area, as well as topological features and surface area combined (all topos and area) for all central cells. (B) For local topology prediction, we calculate features for each neighbor-parent-pair of dividing parent cells using the difference in topological features of the neighbor and parent features as well as the features of the neighbor. Using this general formula, we generated four feature sets: unweighted topology, all topos, biological features (bio, including surface area, perimeter, shared cell wall, and distance), as well as topological and biological features combined (all topos and bio). (C) Parent cell (brown circle) divides into two daughter cells (A, B: representing the cell closer and farther away from the SAM center) changing the local topology in the process. The

colors of the neighbors after division of the central cell represents the adjacency of the neighbor with the daughter cells: class 0 (cyan) neighbor is adjacent to cell A, class 1 (orange) pair neighbor is adjacent cell B, and class 2 (magenta) neighbors are adjacent to both cells. The classes are then used to be predicted from the earlier time point.

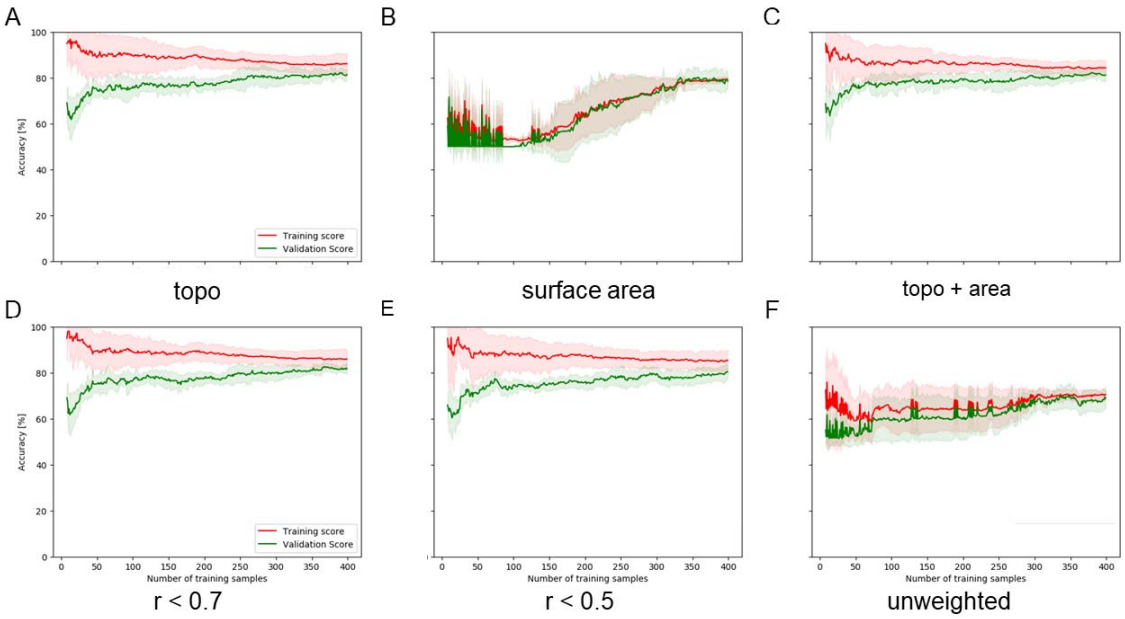


Supplementary Figures 2. Heat map of Pearson correlation coefficients between topological features and surface area. We consider the 16 topological features calculated from the four network scenarios (see Figure 1C): unweighted edges and edges weighted by area, shared wall, and distance. Majority of topological features exhibit small Pearson correlation coefficients (r, legend range from -1 (blue) to 1 (red)). N_{WT} = 5 plants, 4 time steps, n_{WT} = 1225 cells.

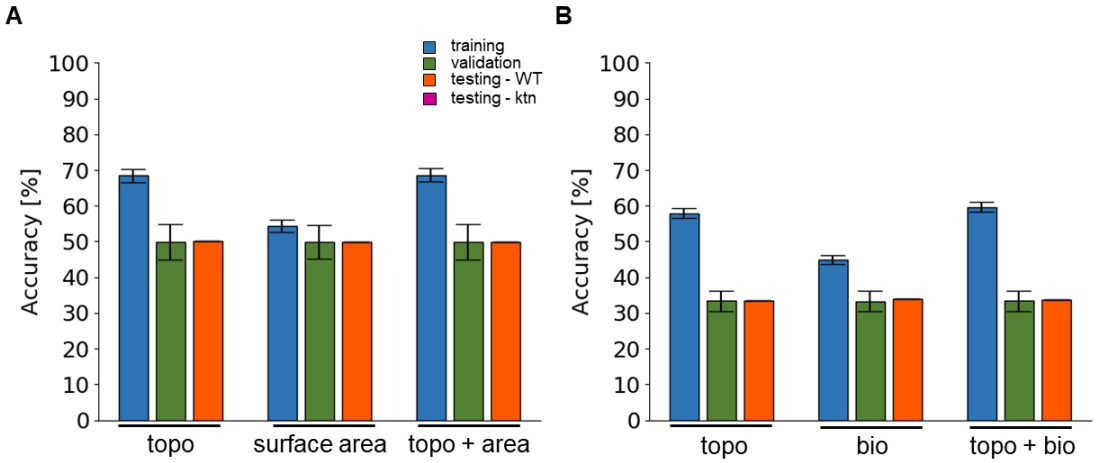


Supplementary Figure 3. Comparative analysis of predictions based on reduced sets of features. Area under the curve (AUC) of the ROC of the support vector machine (SVM) classifier on the training (blue), validation (green), and testing of wild type (orange) and ktn mutant (purple) set of (A) division event and (B) local topology prediction. SVMs are trained on the combined topological features (topo), (A) surface area or (B) (bio, including surface area, perimeter, shared cell wall, and distance), topological features with (A) surface area (topo + area) or (B) bio (topo + bio), reduced set of topological features that show an absolute Pearson correlation coefficient with (A) surface area or (B) bio smaller than 0.7 or 0.5 ($r < 0.7$ and $r < 0.5$), as well as only the topological features derived from the unweighted network scenario (unweighted). The performance on the training and validation set is determined from five-fold cross-validation with mean and the standard deviation shown as error bars. Different letters indicate significance between groups using one-way ANOVA with Tukey's pairwise comparison (p-value < 0.05). Statistical testing for differences of classifier performance for the training and validation sets was conducted separately (letter without and with apostrophe, respectively). N_{WT} = 5 plants, 4 time steps (4 plants for training-validation and 1 plant for testing); N_{ktn} = 3 plants, 3 time steps; (A) n_{WT}

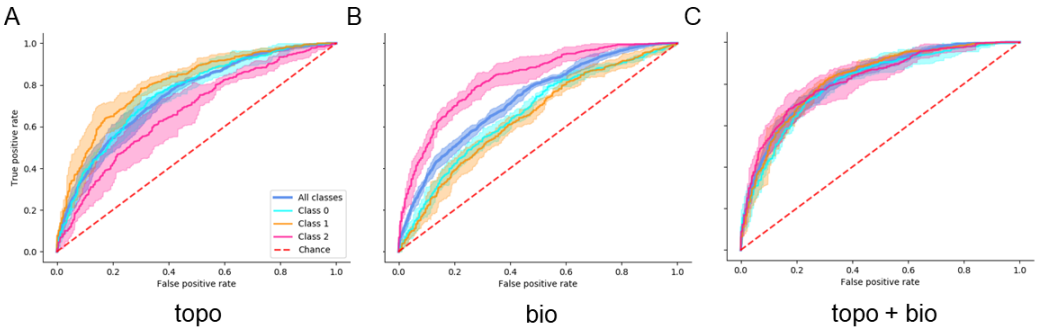
676 = 502 and 156, train-validation and test cells, n_{ktn} = 334 (balanced data) and (B) n_{WT} = 1317 and
677 321, train-validation and test cells, n_{ktn} = 888 (balanced data).
678



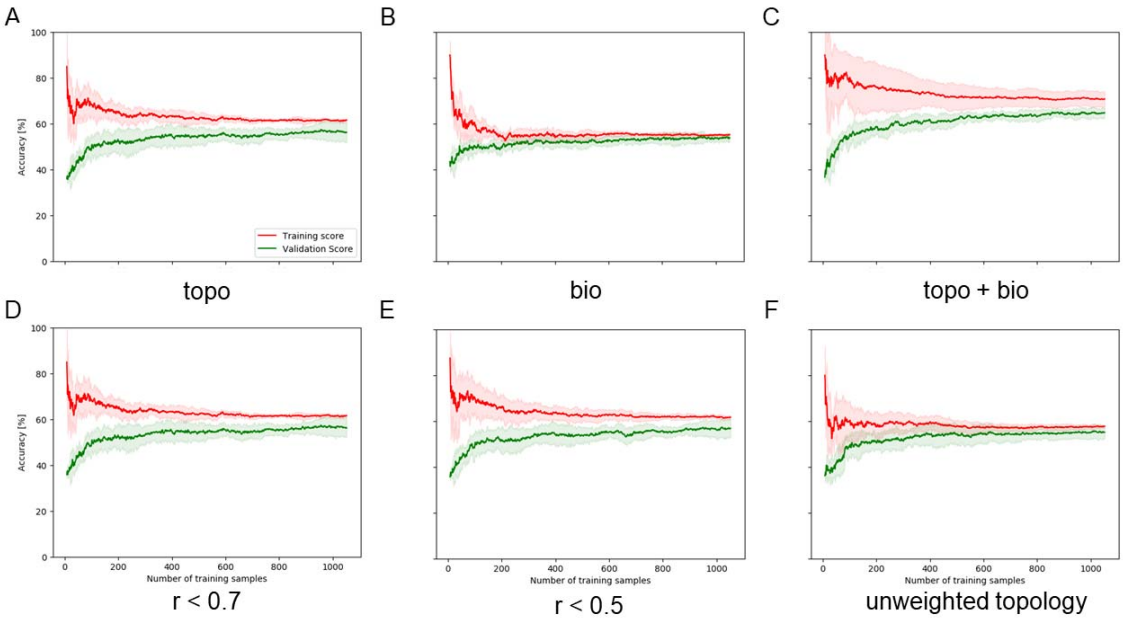
Supplementary Figure 4. Learning curves for the classifiers that predict division events. Learning curves of SVMs predicting cell division events, based on four different feature sets, showing the accuracy in the validation (green) and training (red) set (line: mean, area: ± 1 standard deviation). Feature sets: (A) combined topological features (topo; including features calculated from the four network scenarios, see Figure 1C), (B) surface area as a single feature, (C) topo with surface area (topo + area), (D, E) topological features which have an absolute Pearson correlation coefficient (r) with surface area smaller than 0.7 and 0.5, respectively, and (F) unweighted topological features (unweighted topology). N_{WT} = 4 plants, 4 time steps; n_{WT} = 502 train-validation cells (balanced data), one-way ANOVA.



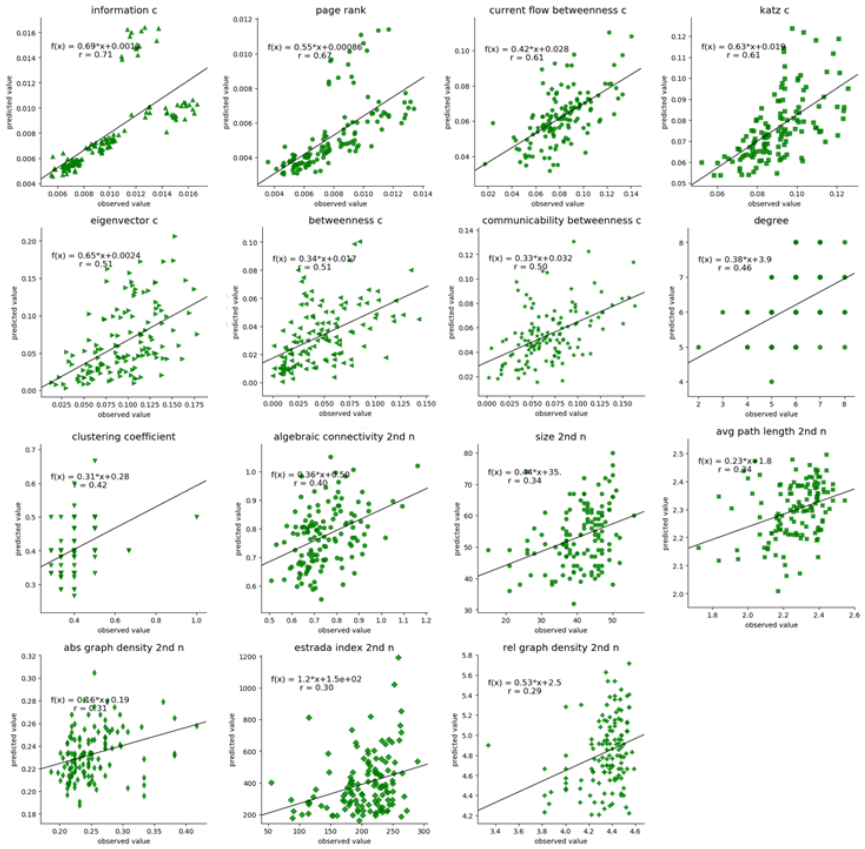
Supplementary Figure 5. Performance of classifiers trained on randomized labels. Support vector machines (SVMs) trained on surface area/biological features and unweighted topological features perform similarly with respect to the prediction of a division event and local topology of shoot apical meristems (SAM). Accuracy of the SVM classifier on the training (blue), validation (green), and testing (orange) set for a division event (A) and local topology prediction (B) based on unweighted topological features (unweighted topology), surface area, or biological features (bio). Shown are the mean and standard deviation on the training and validation sets from five-fold cross-validation. The performance on the training and validation set is determined from five-fold cross-validation with mean and the standard deviation shown as error bars. Different letters indicate significance between groups using one-way ANOVA with Tukey's pairwise comparison: p-value < 0.05. Statistical testing for differences of classifier performance for the training and validation sets was conducted separately (letter without and with apostrophe, respectively). N_{WT} = 5 plants, 4 time steps (4 plants for training-validation and 1 plant for testing); (A) n_{WT} = 502 and 156, and (B) n_{WT} = 1317 and 321 train-validation and test cells respectively (balanced data) shuffling labels 1000 times.



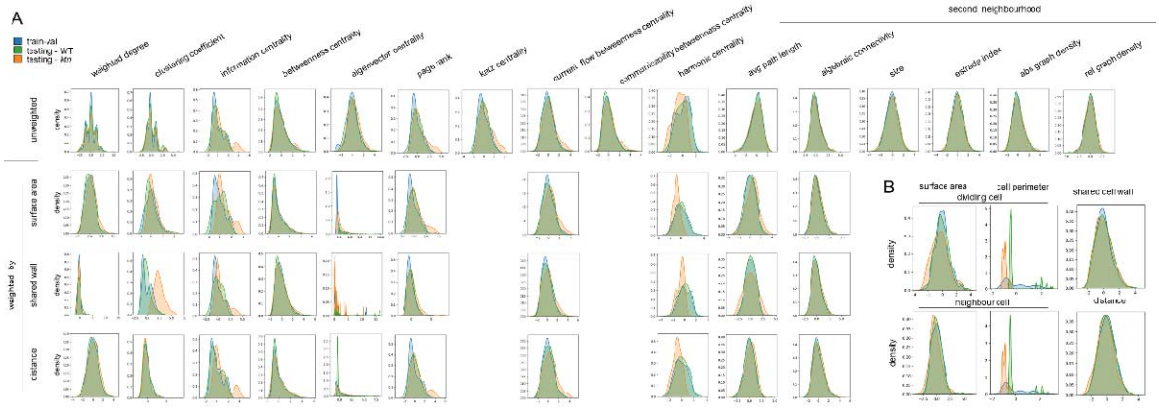
Supplementary Figure 6. Difference between classifiers trained on topological and biological features to predict local topology. Classifiers based on topological and biological features predict the three different classes of cells that are neighbors of cells that have divided in comparison to the previous time point. Receiver operating characteristic-curve predicting cell division events on five-fold cross-validation of (A) combined topological features (topo), including features calculated from the four network scenarios (see Figure 1C), (B) biological features (bio), including surface area, perimeter, shared cell wall, and distance, and (C) topological and biological features combined (topo + bio). The mean performance is shown as a straight line, together with the area of ± 1 standard deviation obtained from the five-fold cross validation of the average ROC-curve combining all classes (blue), of class 0 (cyan), class 1 (orange), and class 2 (magenta). The performance expected by change is marked with a red dashed line. N_{WT} = 4 plants, 4 time steps; n_{WT} = 1317 train-validation cells (balanced data).



Supplementary Figure 7. Learning curves for the classifiers predicting local topological changes. Learning curves of support vector machines (SVMs) that predict local topology based on six different feature sets: (A) combined topological features (topo, including features calculated from the four network scenarios, see Figure 1C), (B) biological features (bio), including surface area, perimeter, shared cell wall, and distance, (C) topological and biological features combined (topo+bio), (D, E) topological features which have an absolute value of Pearson correlation coefficients with all biological features (cor) smaller than 0.7 or 0.5, and (F) only unweighted features showing validation (green) and training (red) accuracy (line: mean, area: ± 1 standard deviation). N_{WT} = 4 plants, 4 time steps; n_{WT} = 1317 train-validation cells (balanced data).



Supplementary Figure 8. Observed and predicted features from non-dividing cells plotted against each other. The observed feature values are plotted against the predicted features values for each cell non-dividing in the observed and predicted tissue. A line fitting the linear regression line is plotted to the data including its function f(x) and Pearson correlation coefficient r for the unweighted topological features: information c (centrality), page rank, current flow betweenness c, katz c, communicability betweenness c, load c, betweenness c, eigenvector c, clustering coefficient, degree, avg. path length on 2 neighborhood (2nd n), abs graph density 2nd n, size 2nd n, algebraic connectivity 2nd n, rel. graph density 2nd n, estrada index 2nd n. The predicted tissue is estimated using the division and topology prediction classifiers trained on topological and biological features of the training-validation data (see Figure 4E). N_{WT} = 1 plant (test), 4 time steps; n_{WT} = 126.



Supplementary Figure 9. Density distributions of topological and biological features of WT and ktn data. The density distributions of the normalized (A) topological and (B) biological features are plotted to differentiate between pooled samples used during training and validation from wild-type (WT; blue), and testing from WT (green), and ktn (orange). The topological features are grouped into unweighted and weighted by area, shared cell wall, and distance (see Figure 1C), while topological features independent of the network scenario are only displayed in the unweighted row. The columns are further grouped into features calculated on the second neighborhood and features from the dividing cell or its neighbor. (A) n_{WT} = 502, 156, train-validation and test cells respectively; n_{ktn} = 334 (balanced data) (B) n_{WT} = 1317 and 312, train-validation and test cells respectively; n_{ktn} = 888 (balanced data).

763 **Supplementary Table 1. Definition of network centralities applied on**
764 **differently weighted network scenarios to be used as feature sets for cell**
765 **division and local topology prediction.** Each network centrality was applied on
766 the unweighted, weighted by area, shared cell wall, and distance network
767 scenario. The network centralities were concatenated together for each cell and
768 used to train and predict different classifiers.
769

Network centrality	Definition	Reference
weighted node degree	$C_D(v) = \frac{\deg(v)}{\lvert V\rvert - 1}$	Freeman, 1979
clustering coefficient	$C_u = \frac{1}{deg(u)(deg(u)-1))} \sum_{vw} \hat{w}_{uv}\hat{w}_{uw}\hat{w}_{vw}$	Saramäki, 2007
information centrality	$C_I(v) = \frac{\lvert V\rvert - 1}{\sum_{u \in V} p_{uv}(v) \quad p_{uv}(u)}$	Sabidussi, 1966
betweenness centrality	$C_B(v) = \frac{2}{(\lvert V\rvert-1)(\lvert V\rvert-2)} \sum_{s,t \in V} \frac{\sigma_v(s,t)}{\sigma(s,t)}$	Freeman, 1977
eigenvector centrality	$C_E(v) = \frac{1}{\lambda} \sum_{u \in V} a_{v,u} C_E(u)$	Bonacich, 1987
PageRank	$R'(u) = c \sum_{v \in B_u} \frac{R'(v)}{N_v} + cE(u)$	Page, 1999
Katz centrality	$x_i = \alpha \sum_j A_{ij} + \beta$	Katz, 1953
current flow betweenness centrality	$c_{CB}(v) = \frac{1}{(n-1)(n-2)} \sum_{s,t \in V} \tau_{st}(v)$	Newman, 2005
communicability betweenness centrality	$C_{CFB}(v) = \frac{2}{(\lvert V\rvert-1)(\lvert V\rvert-2)} \sum_{s,t \in V} \tau_{st}(v)$	Newman, 2005
harmonic centrality	$C_H(v) = \frac{\sum_{u \in V} \frac{1}{d(u,v)}}{\lvert V\rvert - 1}$	Marchiori & Latora, 2000
size	number of edges in a graph	
Estrada index	$EEG(G) = \sum_{j=1}^{n} e^{\lambda_j}$	Estrada, 200
absolute graph density	$C = \frac{2m}{n(n-q)}$	Wilkis & Meara, 2002
relative graph density	$C = \frac{2m}{n}$	
average path length	$a = \sum_{s,t \in V} \frac{d(s,t)}{n(n-1)}$	Mao & Zhang, 2013
algebraic connectivity	second smallest eigenvalue of Laplacian matrix	Fiedler, 1973

773 **Supplementary Table 2. Features ordered by Pearson correlation**
774 **coefficient with area.** Each network centrality was applied on the unweighted,
775 weighted by area, shared cell wall, and distance network scenario. The Pearson
776 correlation coefficient with area was calculated and were ordered from highest to
777 lowest.

Feature name	Pearson correlation coefficient	p-value
betweenness centrality weighted by area	0.711959318	4.858135677288066e-211
degree	0.603854449	4.04599560747419e-136
size on 2 neighborhood	0.592735441	5.437779671764918e-130
betweenness centrality	0.583915039	2.6889326200117693e-125
current flow betweenness centrality	0.580737704	1.217610917275356e-123
load centrality	0.579975133	3.0215367111277033e-123
estrada index on 2 neighborhood	0.577117774	8.912704904812637e-122
estrada index on 2 neighborhood weighted by area	0.577117774	8.912704904812637e-122
communicability betweenness centrality	0.57662665	1.5891934730974062e-121
communicability betweenness centrality weighted by area	0.57662665	1.5891934731075239e-121
avg path length on 2 neighborhood	0.571341971	7.534891764175505e-119
betweenness centrality weighted by shared wall	0.562314483	2.17398436729873e-114
betweenness centrality weighted by distance	0.551932822	1.9924782594347459e-109
current flow betweenness centrality weighted by distance	0.49002158	3.548421338397669e-83
current flow betweenness centrality weighted by shared wall	0.425038752	7.271052967178761e-61
rel graph density on 2 neighborhood	0.402000467	4.72658491702921e-54
current flow betweenness centrality weighted by area	0.357475281	2.5263809251463147e-42
harmonic centrality weighted by area	0.356480643	4.4059338421761555e-42
katz centrality weighted by area	0.318100276	2.1117164600213396e-33
katz centrality	0.318100276	2.1117164600213396e-33
eigenvector centrality	0.296303239	5.28165100633828e-29
page rank	0.268477018	6.447099545102352e-24
harmonic centrality weighted by distance	0.247926272	1.5906836425501315e-20
page rank weighted by shared wall	0.181567285	1.4700578049314804e-11
harmonic centrality weighted by shared wall	0.164231644	1.0815009609397551e-09
information centrality	0.14740273	4.633731043029875e-08
information centrality weighted by shared wall	0.13173255	1.0692961604944774e-06
weighted node degree weighted by shared wall	0.128973752	1.7935045888103348e-06
page rank weighted by distance	0.127130861	2.5188037924397987e-06
information centrality weighted by distance	0.082746096	0.002241568
harmonic centrality	0.071745324	0.008078663

778

779

weighted node degree weighted by distance	0.057524217	0.033774282
avg path length on 2 neighborhood weighted by shared wall	0.056763189	0.03620415
page rank weighted by area	0.037413208	0.167598817
information centrality weighted by area	0.035267233	0.19334169
avg path length on 2 neighborhood weighted by distance	0.0323718	0.232514552
eigenvector centrality weighted by shared wall	0.03031164	0.263614219
eigenvector centrality weighted by distance	-0.051181347	0.05897689
eigenvector centrality weighted by area	-0.203077703	3.8271343446663024e-14
avg path length on 2 neighborhood weighted by area	-0.214150561	1.361632834653018e-15
clustering coefficient weighted by shared wall	-0.236129215	1.0319617716515097e-18
weighted node degree weighted by area	-0.289769264	9.320458571568519e-28
algebraic connectivity on 2 neighborhood	-0.493010144	2.5317849899442773e-84
algebraic connectivity on 2 neighborhood weighted by shared wall	-0.535130612	9.326326866241923e-102
abs graph density on 2 neighborhood	-0.54315277	2.3001423114310698e-105
clustering coefficient	-0.580538544	1.544179245441756e-123
clustering coefficient weighted by distance	-0.584348465	1.5932034339322458e-125
algebraic connectivity on 2 neighborhood weighted by distance	-0.594101302	9.896429835985306e-131
algebraic connectivity on 2 neighborhood weighted by area	-0.639772476	1.1317137430247356e-157
clustering coefficient weighted by area	-0.642874421	1.1369324719240177e-159

780
781

782

783

Supplementary Table 3. Performance measures of division event prediction from SVMs trained on features from four differently weighted topologies and/or area. F1-score (F1), accuracy (Acc), true positive rate (TPR), false positive rate (FPR), and area under the ROC-curve (Auc) of training (train) and unseen data (val) on the splits of the five-fold cross-validations, as well as retraining on train+val data testing on test data (testing), their mean and standard deviation (std) as well as the performance of testing on a never seen plant training on the full training-validation data set. Feature sets: combined topological features (topo; including features of unweighted, weighted by area-, shared cell wall-, and distance topologies), area as a single feature, topo with area, topological features which have an absolute Pearson correlation coefficient with area smaller than 0.7 or 0.5 (cor < 0.7 and cor < 0.5, respectively) and unweighted topological features (unweighted topology).

topo

	train F1	train Acc	train TPR	train FPR	train Auc	val F1	val Acc	val TPR	val FPR	val Auc
split 0	91.36	91.25	92.50	10.00	0.9697	82.69	82.35	84.31	19.61	0.9250
split 1	85.31	84.58	89.55	20.40	0.9133	80.00	80.00	80.00	20.00	0.8708
split 2	93.63	93.53	95.02	7.96	0.9702	78.85	78.00	82.00	26.00	0.8724
split 3	86.46	85.82	90.55	18.91	0.9244	84.62	84.00	88.00	20.00	0.9220
split 4	88.45	88.31	89.55	12.94	0.9511	84.62	84.00	88.00	20.00	0.8836
mean	89.04	88.70	91.44	14.04	0.9457	82.15	81.67	84.46	21.12	0.8948
std	3.08	3.33	2.09	4.87	0.0233	2.37	2.35	3.19	2.44	0.0239
testing	82.17	81.67	84.46	21.12	0.8922	76.62	76.92	75.64	21.79	0.8498
ktn testing						69.93	72.46	64.07	19.16	0.8087

area

	train F1	train Acc	train TPR	train FPR	train Auc	val F1	val Acc	val TPR	val FPR	val Auc
split 0	76.60	78.00	72.00	16.00	0.8674	80.00	81.37	74.51	11.76	0.9285
split 1	78.43	78.11	79.60	23.38	0.8773	82.83	83.00	82.00	16.00	0.8940
split 2	80.60	80.85	79.60	17.91	0.8873	74.51	74.00	76.00	28.00	0.8536
split 3	78.97	79.60	76.62	17.41	0.8836	78.35	79.00	76.00	18.00	0.8664
split 4	80.40	80.60	79.60	18.41	0.8864	75.00	76.00	72.00	20.00	0.8556
mean	79.00	79.43	77.48	18.62	0.8804	78.14	78.67	76.10	18.75	0.8796
std	1.46	1.20	2.98	2.51	0.0074	3.12	3.32	3.29	5.37	0.0284
testing	79.36	79.48	78.88	19.92	0.8804	69.06	72.44	61.54	16.67	0.8358
ktn testing						67.34	70.96	59.88	17.96	0.8046

topo with area

	train F1	train Acc	train TPR	train FPR	train Auc	val F1	val Acc	val TPR	val FPR	val Auc
split 0	89.43	89.25	91.00	12.50	0.9594	83.81	83.33	86.27	19.61	0.9319
split 1	85.37	84.83	88.56	18.91	0.9172	82.00	82.00	82.00	18.00	0.8804
split 2	86.34	86.07	88.06	15.92	0.9307	80.00	79.00	84.00	26.00	0.8424
split 3	83.25	82.59	86.57	21.39	0.9093	84.31	84.00	86.00	18.00	0.9072
split 4	88.89	88.81	89.55	11.94	0.9498	82.35	82.00	84.00	20.00	0.8924
mean	86.66	86.31	88.75	16.13	0.9333	82.50	82.07	84.45	20.32	0.8909
std	2.28	2.49	1.48	3.64	0.0189	1.52	1.72	1.56	2.95	0.0297
testing	80.94	80.68	82.07	20.72	0.8935	77.85	78.85	74.36	16.67	0.8601
ktn testing						71.01	73.35	65.27	18.56	0.8056

cor < 0.7

	train F1	train Acc	train TPR	train FPR	train Auc	val F1	val Acc	val TPR	val FPR	val Auc
split 0	85.57	85.25	87.50	17.00	0.9343	84.91	84.31	88.24	19.61	0.9239
split 1	85.58	84.83	90.05	20.40	0.9182	81.19	81.00	82.00	20.00	0.8696
split 2	93.37	93.28	94.53	7.96	0.9706	80.00	79.00	84.00	26.00	0.8688
split 3	86.26	85.57	90.55	19.40	0.9253	85.44	85.00	88.00	18.00	0.9236
split 4	86.47	86.07	89.05	16.92	0.9358	84.91	84.00	90.00	22.00	0.8884
mean	87.45	87.00	90.34	16.34	0.9369	83.29	82.66	86.45	21.12	0.8949
std	2.98	3.17	2.34	4.40	0.0181	2.24	2.29	2.97	2.75	0.0246
testing	81.78	81.27	84.06	21.51	0.8960	77.22	76.92	78.21	24.36	0.8488
ktn testing						69.13	72.46	61.68	16.77	0.8145

cor < 0.5

	train F1	train Acc	train TPR	train FPR	train Auc	val F1	val Acc	val TPR	val FPR	val Auc
split 0	68.50	68.50	68.50	31.50	0.7510	72.00	72.55	70.59	25.49	0.7559
split 1	75.41	73.88	80.10	32.34	0.8109	62.96	60.00	68.00	48.00	0.6796
split 2	69.52	69.90	68.66	28.86	0.7602	69.31	69.00	70.00	32.00	0.7288
split 3	69.44	68.91	70.65	32.84	0.7396	76.19	75.00	80.00	30.00	0.8012
split 4	70.26	71.14	68.16	25.87	0.7509	65.93	69.00	60.00	22.00	0.7920
mean	70.63	70.47	71.21	30.28	0.7625	69.28	69.11	69.72	31.50	0.7515
std	2.46	1.94	4.53	2.60	0.0251	4.61	5.09	6.39	8.96	0.0443
testing	68.94	69.12	68.53	30.28	0.7512	70.66	68.59	75.64	38.46	0.7720
ktn testing						64.71	64.07	65.87	37.72	0.7115

Supplementary Table 4. Performance measures of local topology prediction from SVMs trained from four different topologies and/or biological features. F1-score (F1), accuracy (Acc), true positive rate (TPR), false positive rate (FPR), and area under the ROC-curve (Auc) of training (train) and unseen data (val) on the splits of the five-fold cross-validations, as well as retraining on train+val data testing on test data (testing) (A) excluding or (B) including dividing neighbours, their mean and standard deviation (dev) as well as the performance of testing on a never seen plant training on the full training-validation data set. Feature sets: combined topological features (topo; including features of unweighted, weighted by area-, shared cell wall-, and distance topologies), biological features (bio; including area, perimeter, shared cell wall, and distance), topo with bio, topological features which have an absolute value of the Pearson correlation coefficients with all biological features (cor) smaller than 0.7 and 0.5, respectively, and unweighted topological features (unweighted topology).

exclude dividing neighbors

		train F1	train c0 F1	train c1 F1	train c2 F1	train Acc	train TPR	train FPR	train Auc	val F1	val c0 F1	val c1 F1	val c2 F1	val Acc	val TPR	val FPR	val Auc
topos	split 0	61.91	60.89	68.74	56.09	61.92	69.23	19.04	0.7979	51.58	53.93	60.12	40.68	51.52	59.09	24.24	0.7215
	split 1	61.96	62.57	68.34	54.98	62.11	69.80	18.95	0.7880	52.80	57.78	56.04	44.58	53.03	57.95	23.48	0.7388
	split 2	62.29	63.28	69.57	54.01	62.39	70.66	18.80	0.7992	60.70	63.16	65.54	53.42	60.98	65.91	19.51	0.7790
	split 3	61.67	63.10	68.90	53.02	61.82	70.37	19.09	0.7909	54.43	48.48	64.80	50.00	54.55	65.91	22.73	0.7402
	split 4	60.25	61.34	65.54	53.86	60.23	66.19	19.89	0.7897	61.70	55.00	71.76	58.33	61.69	70.11	19.16	0.7729
	mean	61.61	62.24	68.22	54.39	61.69	69.25	19.15	0.7931	56.24	55.67	63.65	49.40	56.35	63.80	21.82	0.7505
	std	0.71	0.95	1.39	1.05	0.76	1.60	0.38	0.0045	4.16	4.81	5.31	6.25	4.19	4.59	2.09	0.0219
	testing	62.92	64.57	69.34	54.84	62.95	69.02	18.53	0.7969	60.62	60.58	66.99	54.29	60.58	66.35	19.71	0.7789
	ktn testing									43.96	41.85	42.62	47.42	44.03	39.53	27.98	0.6243

		train F1	train c0 F1	train c1 F1	train c2 F1	train Acc	train TPR	train FPR	train Auc	val F1	val c0 F1	val c1 F1	val c2 F1	val Acc	val TPR	val FPR	val Auc
bio	split 0	54.59	48.47	46.92	68.39	55.27	45.58	22.36	0.7406	51.80	46.34	46.24	62.83	52.27	45.45	23.86	0.7084
	split 1	54.99	51.61	46.43	66.93	55.56	43.59	22.22	0.7365	53.45	49.16	42.77	68.42	54.17	38.64	22.92	0.7331
	split 2	54.76	49.63	46.70	67.96	55.46	43.30	22.27	0.7412	52.51	48.24	44.30	65.00	53.41	39.77	23.30	0.7210
	split 3	53.75	48.16	45.75	67.36	54.42	44.44	22.79	0.7295	56.52	51.76	49.08	68.72	57.20	45.45	21.40	0.7466
	split 4	54.96	49.48	49.17	66.23	55.49	46.31	22.25	0.7342	52.26	43.31	43.58	69.89	52.87	44.83	23.56	0.7317
	mean	54.61	49.47	46.99	67.37	55.24	44.65	22.38	0.7364	53.31	47.76	45.19	66.97	53.98	42.83	23.01	0.7281
	std	0.45	1.21	1.16	0.76	0.42	1.15	0.21	0.0043	1.69	2.83	2.26	2.64	1.72	2.99	0.86	0.0128
	testing	56.48	52.42	49.64	67.37	56.95	47.15	21.53	0.7521	49.18	44.24	35.75	67.54	50.32	30.77	24.84	0.7067
	ktn testing									50.39	41.61	38.87	70.69	51.58	37.16	24.21	0.7281

		train F1	train c0 F1	train c1 F1	train c2 F1	train Acc	train TPR	train FPR	train Auc	val F1	val c0 F1	val c1 F1	val c2 F1	val Acc	val TPR	val FPR	val Auc
topo And Bio	split 0	67.70	65.89	68.44	68.76	67.71	69.80	16.14	0.8426	60.25	61.54	60.44	58.76	60.23	62.50	19.89	0.7886
	split 1	70.28	68.75	70.35	71.74	70.28	70.66	14.86	0.8539	62.86	62.64	59.65	66.29	62.88	57.95	18.56	0.8187
	split 2	77.11	76.86	77.47	77.01	77.11	76.92	11.44	0.8943	63.64	62.15	64.33	64.44	63.64	62.50	18.18	0.8165
	split 3	68.38	68.25	67.62	69.29	68.38	67.52	15.81	0.8457	65.36	61.82	62.92	71.35	65.53	63.64	17.23	0.8249
	split 4	68.47	68.19	67.97	69.25	68.47	69.32	15.77	0.8532	67.83	63.16	68.13	72.19	67.82	71.26	16.09	0.8430
	mean	70.39	69.59	70.37	71.21	70.39	70.84	14.81	0.8579	63.99	62.26	63.09	66.61	64.02	63.57	17.99	0.8183
	std	3.47	3.77	3.68	3.08	3.47	3.21	1.73	0.0187	2.53	0.58	3.03	4.90	2.55	4.31	1.28	0.0175
	testing	71.99	71.06	72.50	72.41	71.98	72.67	14.01	0.8676	65.70	66.35	65.70	65.05	65.71	65.38	17.15	0.8344
	ktn testing									54.78	51.01	45.84	67.50	55.41	41.89	22.30	0.7255

		train F1	train c0 F1	train c1 F1	train c2 F1	train Acc	train TPR	train FPR	train Auc	val F1	val c0 F1	val c1 F1	val c2 F1	val Acc	val TPR	val FPR	val Auc
cor 0.7	split 0	61.91	60.89	68.74	56.09	61.92	69.23	19.04	0.7979	51.58	53.93	60.12	40.68	51.52	59.09	24.24	0.7215
	split 1	61.96	62.57	68.34	54.98	62.11	69.80	18.95	0.7880	52.80	57.78	56.04	44.58	53.03	57.95	23.48	0.7388
	split 2	62.29	63.28	69.57	54.01	62.39	70.66	18.80	0.7992	60.70	63.16	65.54	53.42	60.98	65.91	19.51	0.7790
	split 3	61.67	63.10	68.90	53.02	61.82	70.37	19.09	0.7909	54.43	48.48	64.80	50.00	54.55	65.91	22.73	0.7402
	split 4	60.25	61.34	65.54	53.86	60.23	66.19	19.89	0.7897	61.70	55.00	71.76	58.33	61.69	70.11	19.16	0.7729
	mean	61.61	62.24	68.22	54.39	61.69	69.25	19.15	0.7931	56.24	55.67	63.65	49.40	56.35	63.80	21.82	0.7505
	std	0.71	0.95	1.39	1.05	0.76	1.60	0.38	0.0045	4.16	4.81	5.31	6.25	4.19	4.59	2.09	0.0219
	testing	62.92	64.57	69.34	54.84	62.95	69.02	18.53	0.7969	60.62	60.58	66.99	54.29	60.58	66.35	19.71	0.7789
	ktn testing									43.96	41.85	42.62	47.42	44.03	39.53	27.98	0.6243

		train F1	train c0 F1	train c1 F1	train c2 F1	train Acc	train TPR	train FPR	train Auc	val F1	val c0 F1	val c1 F1	val c2 F1	val Acc	val TPR	val FPR	val Auc
cor 0.5	split 0	62.38	61.10	69.02	57.02	62.39	69.52	18.80	0.7985	53.36	56.04	62.43	41.62	53.41	61.36	23.30	0.7258
	split 1	61.73	62.01	67.89	55.29	61.82	68.66	19.09	0.7886	53.98	58.56	56.67	46.71	54.17	57.95	22.92	0.7418
	split 2	62.78	64.15	69.21	54.97	62.87	69.80	18.57	0.8080	60.81	62.43	66.67	53.33	60.98	65.91	19.51	0.7730
	split 3	61.63	62.91	68.35	53.64	61.73	69.23	19.14	0.7903	54.48	49.10	65.17	49.18	54.55	65.91	22.73	0.7387
	split 4	60.35	61.67	65.72	53.65	60.32	65.91	19.84	0.7907	61.70	55.00	71.76	58.33	61.69	70.11	19.16	0.7722
	mean	61.77	62.37	68.04	54.92	61.83	68.62	19.09	0.7952	56.87	56.23	64.54	49.83	56.96	64.25	21.52	0.7503
	std	0.83	1.07	1.25	1.25	0.86	1.41	0.43	0.0072	3.61	4.39	4.97	5.70	3.60	4.19	1.80	0.0190
	testing	62.35	63.67	68.63	54.76	62.41	69.02	18.79	0.7926	60.07	58.54	67.98	53.70	59.94	66.35	20.03	0.7734
	ktn testing									44.77	43.30	42.99	48.03	44.82	39.86	27.59	0.6265

		train F1	train c0 F1	train c1 F1	train c2 F1	train Acc	train TPR	train FPR	train Auc	val F1	val c0 F1	val c1 F1	val c2 F1	val Acc	val TPR	val FPR	val Auc
unweight topology	split 0	58.07	61.08	64.43	48.71	58.40	68.38	20.80	0.7647	54.61	62.64	56.35	44.85	54.92	57.95	22.54	0.7241
	split 1	57.92	61.37	64.02	48.37	58.31	66.67	20.85	0.7617	50.66	58.51	55.32	38.16	51.52	59.09	24.24	0.7237
	split 2	56.83	59.31	62.72	48.46	56.98	64.96	21.51	0.7538	59.05	62.86	64.92	49.38	59.47	70.45	20.27	0.7604
	split 3	57.03	61.71	61.52	47.87	57.17	62.39	21.42	0.7564	52.88	51.19	59.46	48.00	53.03	62.50	23.48	0.7033
	split 4	56.90	61.67	62.93	46.11	57.39	67.05	21.31	0.7552	57.11	56.00	68.21	47.13	57.09	67.82	21.46	0.7597
	mean	57.35	61.03	63.13	47.90	57.65	65.89	21.17	0.7584	54.86	58.24	60.85	45.50	55.21	63.56	22.40	0.7342
	std	0.53	0.89	1.03	0.94	0.59	2.06	0.30	0.0042	2.97	4.37	4.97	3.96	2.83	4.86	1.42	0.0224
	testing	58.24	62.28	65.07	47.37	58.62	68.11	20.69	0.7684	56.75	60.10	65.16	45.00	57.05	69.23	21.47	0.7757
	ktn testing									44.14	44.63	43.78	44.01	44.14	44.59	27.93	0.6093

including dividing neighbors

	train F1	train c0 F1	train c1 F1	train c2 F1	train Acc	train TPR	train FPR	train Auc	val F1	val c0 F1	val c1 F1	val c2 F1	val Acc	val TPR	val FPR	val Auc
split 0	59.86	64.54	65.50	49.54	60.22	67.12	19.89	0.7862	59.11	59.21	69.86	48.25	59.18	69.39	20.41	0.7756
split 1	61.88	64.41	68.06	53.16	62.11	69.40	18.95	0.7959	58.58	64.36	63.12	48.25	58.68	65.07	20.66	0.7741
split 2	63.08	66.94	69.51	52.79	63.48	71.11	18.26	0.8037	56.29	60.55	62.76	45.56	56.85	62.33	21.58	0.7657
split 3	65.81	67.95	70.25	59.22	65.93	71.45	17.04	0.8199	53.38	56.87	62.95	40.31	54.11	65.75	22.95	0.7375
split 4	63.82	66.34	70.21	54.93	64.10	72.31	17.95	0.8100	55.02	59.66	61.54	43.87	55.48	65.75	22.26	0.7472
mean	62.89	66.03	68.71	53.93	63.17	70.28	18.42	0.8031	56.47	60.13	64.05	45.25	56.86	65.66	21.57	0.7600
std	1.98	1.38	1.79	3.17	1.92	1.84	0.96	0.0116	2.15	2.44	2.96	2.98	1.91	2.25	0.95	0.0151
testing	68.82	71.97	73.21	61.28	69.04	74.01	15.48	0.8402	49.86	46.88	60.48	42.24	50.68	71.43	24.66	0.7093

topos

	train F1	train c0 F1	train c1 F1	train c2 F1	train Acc	train TPR	train FPR	train Auc	val F1	val c0 F1	val c1 F1	val c2 F1	val Acc	val TPR	val FPR	val Auc
split 0	54.15	50.94	42.95	68.56	54.62	40.41	22.69	0.7445	54.18	49.84	39.69	73.02	55.10	35.37	22.45	0.7507
split 1	55.37	52.84	43.89	69.38	55.95	39.32	22.02	0.7515	50.77	47.14	35.96	69.23	51.60	32.88	24.20	0.7247
split 2	53.74	48.70	43.94	68.58	54.30	41.54	22.85	0.7436	54.04	48.08	44.06	69.97	54.34	43.15	22.83	0.7491
split 3	54.20	50.13	42.54	69.93	54.81	39.49	22.59	0.7498	52.70	48.30	43.57	66.23	52.97	41.78	23.52	0.7305
split 4	54.45	52.87	41.00	69.50	55.38	35.21	22.31	0.7500	52.05	50.66	36.14	69.35	53.42	30.82	23.29	0.7342
mean	54.38	51.09	42.86	69.19	55.02	39.19	22.49	0.7479	52.75	48.80	39.88	69.56	53.49	36.80	23.26	0.7379
std	0.54	1.61	1.08	0.54	0.59	2.14	0.29	0.0032	1.27	1.27	3.48	2.16	1.20	4.86	0.60	0.0103
testing	55.60	52.15	45.49	69.15	56.18	41.04	21.91	0.7549	46.26	53.18	17.50	68.11	51.19	10.71	24.40	0.7158

bio

	train F1	train c0 F1	train c1 F1	train c2 F1	train Acc	train TPR	train FPR	train Auc	val F1	val c0 F1	val c1 F1	val c2 F1	val Acc	val TPR	val FPR	val Auc
split 0	67.78	66.95	66.61	69.78	67.75	67.47	16.12	0.8424	68.93	65.33	68.07	73.40	68.93	65.99	15.53	0.8579
split 1	70.67	70.07	68.98	72.97	70.66	68.03	14.67	0.8657	67.36	67.60	66.22	68.28	67.35	67.81	16.32	0.8517
split 2	71.06	70.05	70.32	72.82	71.05	70.26	14.47	0.8636	69.68	69.28	65.74	74.02	69.63	65.07	15.18	0.8565
split 3	70.77	68.95	70.01	73.35	70.77	70.43	14.62	0.8703	64.38	63.70	64.67	64.79	64.38	66.44	17.81	0.8240
split 4	70.55	69.27	69.62	72.77	70.54	69.91	14.73	0.8663	66.62	65.07	65.71	69.08	66.67	63.01	16.67	0.8288
mean	70.17	69.06	69.11	72.34	70.15	69.22	14.92	0.8617	67.40	66.20	66.08	69.91	67.39	65.66	16.30	0.8438
std	1.21	1.14	1.33	1.30	1.21	1.22	0.61	0.0099	1.86	1.99	1.12	3.43	1.84	1.59	0.92	0.0144
testing	70.01	69.05	69.58	71.38	70.00	69.63	15.00	0.8619	67.69	68.38	67.20	67.48	67.69	64.29	16.16	0.8442

topo
And
Bio

	train F1	train c0 F1	train c1 F1	train c2 F1	train Acc	train TPR	train FPR	train Auc	val F1	val c0 F1	val c1 F1	val c2 F1	val Acc	val TPR	val FPR	val Auc
split 0	59.86	64.54	65.50	49.54	60.22	67.12	19.89	0.7862	59.11	59.21	69.86	48.25	59.18	69.39	20.41	0.7756
split 1	61.88	64.41	68.06	53.16	62.11	69.40	18.95	0.7959	58.58	64.36	63.12	48.25	58.68	65.07	20.66	0.7741
split 2	63.08	66.94	69.51	52.79	63.48	71.11	18.26	0.8037	56.29	60.55	62.76	45.56	56.85	62.33	21.58	0.7657
split 3	65.81	67.95	70.25	59.22	65.93	71.45	17.04	0.8199	53.38	56.87	62.95	40.31	54.11	65.75	22.95	0.7375
split 4	63.82	66.34	70.21	54.93	64.10	72.31	17.95	0.8100	55.02	59.66	61.54	43.87	55.48	65.75	22.26	0.7472
mean	62.89	66.03	68.71	53.93	63.17	70.28	18.42	0.8031	56.47	60.13	64.05	45.25	56.86	65.66	21.57	0.7600
std	1.98	1.38	1.79	3.17	1.92	1.84	0.96	0.0116	2.15	2.44	2.96	2.98	1.91	2.25	0.95	0.0151
testing	68.82	71.97	73.21	61.28	69.04	74.01	15.48	0.8402	49.86	46.88	60.48	42.24	50.68	71.43	24.66	0.7093

cor 0.7

	train F1	train c0 F1	train c1 F1	train c2 F1	train Acc	train TPR	train FPR	train Auc	val F1	val c0 F1	val c1 F1	val c2 F1	val Acc	val TPR	val FPR	val Auc
split 0	59.55	63.78	65.10	49.77	59.87	66.44	20.06	0.7765	59.38	58.71	69.44	50.00	59.41	68.03	20.29	0.7751
split 1	59.98	63.36	66.67	49.91	60.40	68.03	19.80	0.7850	56.75	62.94	63.26	44.04	57.08	67.81	21.46	0.7653
split 2	64.20	66.94	70.53	55.15	64.50	71.79	17.75	0.8096	56.76	60.74	65.10	44.44	57.53	66.44	21.23	0.7672
split 3	63.92	67.27	68.64	55.83	64.10	69.40	17.95	0.8091	52.19	54.78	61.33	40.46	52.74	63.01	23.63	0.7363
split 4	66.18	68.26	71.92	58.36	66.38	73.33	16.81	0.8203	54.57	59.41	61.18	43.12	55.02	63.70	22.49	0.7457
mean	62.77	65.92	68.57	53.80	63.05	69.80	18.47	0.8001	55.93	59.31	64.06	44.41	56.36	65.80	21.82	0.7579
std	2.58	1.97	2.48	3.41	2.51	2.49	1.25	0.0165	2.41	2.69	3.05	3.12	2.28	2.08	1.14	0.0145
testing	69.13	71.73	73.50	62.17	69.36	74.56	15.32	0.8439	51.04	47.53	60.35	45.23	51.70	69.90	24.15	0.7154

cor 0.5

	train F1	train c0 F1	train c1 F1	train c2 F1	train Acc	train TPR	train FPR	train Auc	val F1	val c0 F1	val c1 F1	val c2 F1	val Acc	val TPR	val FPR	val Auc
split 0	55.18	61.71	62.55	41.28	56.16	67.64	21.92	0.7459	53.36	58.57	64.13	37.40	54.65	68.71	22.68	0.7484
split 1	55.99	61.93	63.29	42.76	56.87	67.18	21.57	0.7628	56.15	62.98	61.89	43.57	56.39	65.07	21.80	0.7423
split 2	56.15	64.01	64.56	39.88	57.38	69.74	21.31	0.7687	54.60	61.03	61.94	40.85	55.94	65.75	22.03	0.7455
split 3	55.44	62.47	63.01	40.83	56.41	66.84	21.79	0.7598	52.17	58.49	62.38	35.63	53.42	66.44	23.29	0.7316
split 4	58.84	62.41	66.07	48.03	59.43	69.57	20.28	0.7754	52.18	62.34	59.94	34.26	53.42	65.07	23.29	0.7324
mean	56.32	62.51	63.90	42.55	57.25	68.19	21.38	0.7625	53.69	60.68	62.05	38.34	54.77	66.21	22.62	0.7400
std	1.31	0.80	1.28	2.89	1.17	1.22	0.58	0.0099	1.52	1.87	1.34	3.42	1.24	1.35	0.62	0.0068
testing	58.80	65.42	65.44	45.53	59.69	70.31	20.16	0.7778	56.13	61.58	63.87	42.94	56.80	69.90	21.60	0.7620

unweight
topology

831

832

833

834

References

Alim, K., Hamant, O., Boudaoud, A., 2012. Regulatory role of cell division rules on tissue growth heterogeneity. Frontiers in plant science 3, 174.

Barbier de Reuille, P., Routier-Kierzkowska, A.-L., Kierzkowski, D., Bassel, G.W., Schüpbach, T., Tauriello, G., Bajpai, N., Strauss, S., Weber, A., Kiss, A., Burian, A., Hofhuis, H., Sapala, A., Lipowczan, M., Heimlicher, M.B., Robinson, S., Bayer, E.M., Basler, K., Koumoutsakos, P., Roeder, A.H.K., Aegerter-Wilmsen, T., Nakayama, N., Tsiantis, M., Hay, A., Kwiatkowska, D., Xenarios, I., Kuhlemeier, C., Smith, R.S., 2015. MorphoGraphX: A platform for quantifying morphogenesis in 4D. eLife 4, 5864.

Beauzamy, L., Louveaux, M., Hamant, O., Boudaoud, A., 2015. Mechanically, the Shoot Apical Meristem of Arabidopsis Behaves like a Shell Inflated by a Pressure of About 1 MPa. Frontiers in plant science 6, 1038.

Besson, S., Dumais, J., 2011. Universal rule for the symmetric division of plant cells. Proceedings of the National Academy of Sciences of the United States of America 108, 6294–6299.

Bhavsar, H., Panchal, M.H., 2012. A review on support vector machine for data classification. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) 1, 185–189.

Breuer, D., Nowak, J., Ivakov, A., Somssich, M., Persson, S., Nikoloski, Z., 2017. System-wide organization of actin cytoskeleton determines organelle transport in hypocotyl plant cells. Proceedings of the National Academy of Sciences of the United States of America 114, E5741-E5749.

Camacho, D.M., Collins, K.M., Powers, R.K., Costello, J.C., Collins, J.J., 2018. Next-Generation Machine Learning for Biological Networks. Cell 173, 1581–1592.

D'Ario, M., Sablowski, R., 2019. Cell Size Control in Plants. Annual review of genetics 53, 45–65.

D'Ario, M., Tavares, R., Schiessl, K., Desvoyes, B., Gutierrez, C., Howard, M., Sablowski, R., 2021. Cell size controlled in plants using DNA content as an internal scale. Science (New York, N.Y.) 372, 1176–1181.

Dewitte, W., Murray, J.A., 2003. The Plant Cell Cycle. Annual Review of Plant Biology 54, 235–264.

865 Eng, R.C., Schneider, R., Matz, T.W., Carter, R., Ehrhardt, D.W., Jönsson, H., Nikoloski, Z.,
866 Sampathkumar, A., 2021. KATANIN and CLASP function at different spatial scales to mediate
867 microtubule response to mechanical stress in Arabidopsis cotyledons. Current biology : CB.
868 https://doi.org/10.1016/j.cub.2021.05.019.
869 Gibson, W.T., Veldhuis, J.H., Rubinstein, B., Cartwright, H.N., Perrimon, N., Brodland, G.W.,
870 Nagpal, R., Gibson, M.C., 2011. Control of the mitotic cleavage plane by local epithelial
871 topology. Cell 144, 427–438.
872 Hartig, K., Beck, E., 2006. Crosstalk between auxin, cytokinins, and sugars in the plant cell cycle.
873 Plant biology (Stuttgart, Germany) 8, 389–396.
874 Jackson, M.D.B., Duran-Nebreda, S., Kierzkowski, D., Strauss, S., Xu, H., Landrein, B., Hamant, O.,
875 Smith, R.S., Johnston, I.G., Bassel, G.W., 2019. Global Topological Order Emerges through
876 Local Mechanical Control of Cell Divisions in the Arabidopsis Shoot Apical Meristem. Cell
877 systems 8, 53-65.e3.
878 Jones, A.R., Forero-Vargas, M., Withers, S.P., Smith, R.S., Traas, J., Dewitte, W., Murray, J.A.H.,
879 2017. Cell-size dependent progression of the cell cycle creates homeostasis and flexibility of
880 plant cell size. Nature communications 8, 15060.
881 Kitagawa, M., Jackson, D., 2017. Plasmodesmata-Mediated Cell-to-Cell Communication in the
882 Shoot Apical Meristem: How Stem Cells Talk. Plants (Basel, Switzerland) 6.
883 Komis, G., Luptovčiak, I., Ovečka, M., Samakovli, D., Šamajová, O., Šamaj, J., 2017. Katanin
884 Effects on Dynamics of Cortical Microtubules and Mitotic Arrays in Arabidopsis thaliana
885 Revealed by Advanced Live-Cell Imaging. Frontiers in plant science 8, 866.
886 Li, Y., Liu, D., López-Paz, C., Olson, B.J., Umen, J.G., 2016. A new class of cyclin dependent kinase
887 in Chlamydomonas is required for coupling cell size to cell division. eLife 5, e10767.
888 Louveaux, M., Julien, J.-D., Mirabet, V., Boudaoud, A., Hamant, O., 2016. Cell division plane
889 orientation based on tensile stress in Arabidopsis thaliana. Proceedings of the National
890 Academy of Sciences of the United States of America 113, E4294-303.
891 Nowak, J., Eng, R.C., Matz, T., Waack, M., Persson, S., Sampathkumar, A., Nikoloski, Z., 2021. A
892 network-based framework for shape analysis enables accurate characterization of leaf
893 epidermal cells. Nature communications 12, 458.
894 Pisner, D.A., Schnyer, D.M., 2020. Support vector machine. In: Machine Learning. Elsevier,
895 pp. 101–121.

Qi, F., Zhang, F., 2019. Cell Cycle Regulation in the Plant Response to Stress. Frontiers in plant science 10, 1765.

Reddy, G.V., Heisler, M.G., Ehrhardt, D.W., Meyerowitz, E.M., 2004. Real-time lineage analysis reveals oriented cell divisions associated with morphogenesis at the shoot apex of Arabidopsis thaliana. Development (Cambridge, England) 131, 4225–4237.

Sahlin, P., Jönsson, H., 2010. A modeling study on how cell division affects properties of epithelial tissues under isotropic growth. PloS one 5, e11750.

Sampathkumar, A., 2020. Mechanical feedback-loop regulation of morphogenesis in plants. Development (Cambridge, England) 147.

Sampathkumar, A., Krupinski, P., Wightman, R., Milani, P., Berquand, A., Boudaoud, A., Hamant, O., Jönsson, H., Meyerowitz, E.M., 2014. Subcellular and supracellular mechanical stress prescribes cytoskeleton behavior in Arabidopsis cotyledon pavement cells. eLife 3, e01967.

Schoof, H., Lenhard, M., Haecker, A., Mayer, K.F., Jürgens, G., Laux, T., 2000. The Stem Cell Population of Arabidopsis Shoot Meristems Is Maintained by a Regulatory Loop between the CLAVATA and WUSCHEL Genes. Cell 100, 635–644.

Shapiro, B.E., Tobin, C., Mjolsness, E., Meyerowitz, E.M., 2015. Analysis of cell division patterns in the Arabidopsis shoot apical meristem. Proceedings of the National Academy of Sciences 112, 4815–4820.

Shi, B., Guo, X., Wang, Y., Xiong, Y., Wang, J., Hayashi, K.-I., Lei, J., Zhang, L., Jiao, Y., 2018. Feedback from Lateral Organs Controls Shoot Apical Meristem Growth by Modulating Auxin Transport. Developmental cell 44, 204-216.e6.

Shimotohno, A., Aki, S.S., Takahashi, N., Umeda, M., 2021. Regulation of the Plant Cell Cycle in Response to Hormones and the Environment. Annual Review of Plant Biology 72, 273–296.

Uyttewaal, M., Burian, A., Alim, K., Landrein, B., Borowska-Wykręt, D., Dedieu, A., Peaucelle, A., Ludynia, M., Traas, J., Boudaoud, A., Kwiatkowska, D., Hamant, O., 2012. Mechanical stress acts via katanin to amplify differences in growth rate between adjacent cells in Arabidopsis. Cell 149, 439–451.

Varner, J.E., Lin, L.-S., 1989. Plant cell wall architecture. Cell 56, 231–239.

Veylder, L. de, Beeckman, T., Inzé, D., 2007. The ins and outs of the plant cell cycle. Nature reviews. Molecular cell biology 8, 655–665.

926 Wang, Y., Sampathkumar, A., 2020. Live Cell Imaging of Microtubule Cytoskeleton and
927 Micromechanical Manipulation of the Arabidopsis Shoot Apical Meristem. Journal of
928 visualized experiments : JoVE.
929 Wangenheim, D. von, Fangerau, J., Schmitz, A., Smith, R.S., Leitte, H., Stelzer, E.H.K., Maizel, A.,
930 2016. Rules and Self-Organizing Properties of Post-embryonic Plant Organ Cell Division
931 Patterns. Current biology : CB 26, 439–449.
932 Willis, L., Refahi, Y., Wightman, R., Landrein, B., Teles, J., Huang, K.C., Meyerowitz, E.M., Jönsson,
933 H., 2016. Cell size and growth regulation in the Arabidopsis thaliana apical stem cell niche.
934 Proceedings of the National Academy of Sciences of the United States of America 113,
935 E8238-E8246.
936
937
938