

RESEARCH ARTICLE

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

Timon W. Matz^{1,2}, Yang Wang³, Ritika Kulshreshtha³, Arun Sampathkumar³ and Zoran Nikoloski^{1,2,*}

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 the tissue level. We show that a support vector machine classifier based on the SAM network properties is predictive of cell division events, with test accuracy of 76%, which matches that based on cell size alone. Furthermore, 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. Together, the results from our network-based analysis demonstrate a feedback mechanism between tissue topology and cell division in *A. thaliana* SAMs.

KEY WORDS: *Arabidopsis thaliana*, Cell division, Classification models, Networks, Shoot apical meristem, Topology

INTRODUCTION

The adjacency of cells, which species 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).


As the cells in the SAM are glued to each other by a rigid cell wall, changes in the SAM topology, abstracting the tissue with the respective cellular neighborhoods, 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 (De Veylder et al., 2007; Qi and Zhang, 2020). Willis et al. (2016) recently showed that initial cell size at birth influences the increase in cell size (the sizer model – smaller cells grow faster and all cells divide at a specific size threshold), even though there seems to also be a component of constant size increase (the adder model – always adding the same size increment regardless of the initial cell size at birth) in the 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 of position and cell-to-cell contact (Willis et al., 2016), a recent study by Jackson et al. (2019) indicates that dividing cells display higher centralities, measured by their participation in more shortest paths between pairs of cells in the network representation of the *A. thaliana* SAM; however, this observation is not sufficient to accurately predict cell division from topological or cell size properties, and is only based on 32 dividing cells from the three upper layers with only seven derived from the surface layer (Jackson et al., 2019).

As 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 the microtubule-severing protein KATANIN (Uyttewaald et al., 2012). It has been shown that the lack of mechanical feedback, as seen 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 whether topological features are indeed relevant for cell division and related processes.

This issue can be readily addressed due to the availability of plant lines expressing stable fluorescence reporters that allow the monitoring of cellular outlines in combination with confocal imaging techniques (Reddy et al., 2004). In addition, the combination of user-friendly tools for accurate segmentation, such as 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. Furthermore, the analysis of micrographs for subcellular and cellular scale phenotypes has been facilitated by adopting the network paradigm by abstracting a biological system as a network (Breuer et al., 2017; Nowak et al., 2021). For example, actin

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

*Author for correspondence (nikoloski@mpimp-golm.mpg.de)

 T.W.M., 0000-0001-9802-5094; Y.W., 0000-0001-5765-1443; A.S., 0000-0003-1703-0137; Z.N., 0000-0003-2671-6763

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

Handling Editor: Ykä Helariutta
Received 15 June 2022; Accepted 15 July 2022

filaments can be represented as a network of intersecting paths (Breuer et al., 2017) or the shape of cell membrane can be represented by nodes connected via edges if they do not intersect the membrane (Nowak et al., 2021).

In the context of cellular connectivity network (i.e. topology), nodes correspond to cells and edges represent adjacency of cells. Network properties (topological features) have been employed to study and devise models of cell wall placement 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 some cell walls when compared with a more traditional approach (Jackson et al., 2019), such as dividing cells using the shortest wall placement or generalized Errera’s rule (Besson and Dumais, 2011). These resemble the minimization of tensile stress in the central region of the SAM (Louveaux et al., 2016). Although these models provide an important step towards resolving 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).

Furthermore, although individual features have been used in the context of the tissue topology, the whole landscape of topological properties is underexplored. Topological properties can be divided into three groups:

(1) Local properties pertain to individual nodes or edges and can be determined by having access to the immediate neighborhoods. Examples of local topological properties include weighted degree (averaging the edge weights to all neighbors of a given node) or the clustering coefficient, which quantifies the density of connections in the immediate neighborhood of a node.

(2) Global properties pertain to the entire network (e.g. representation of tissue topology) and require knowledge of the entire network to be quantified; e.g. the diameter and the characteristic path length are two global network properties associated with the maximum and average hops for a signal to be transferred from one to another cell.

(3) Local-global properties pertain to individual nodes or edges, but require knowledge of the entire network to be quantified. These include properties related to random walks and shortest paths between two nodes (Table S1).

Although there have been attempts of combining network properties with imaging data from SAMs, there is little progress 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 unanswered. To address this question, we imaged eight SAMs of *A. thaliana* expressing a plasma membrane reporter (pUBQ10:acyl-YFP) every 24 h over 5 days using confocal microscopy (Fig. 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.

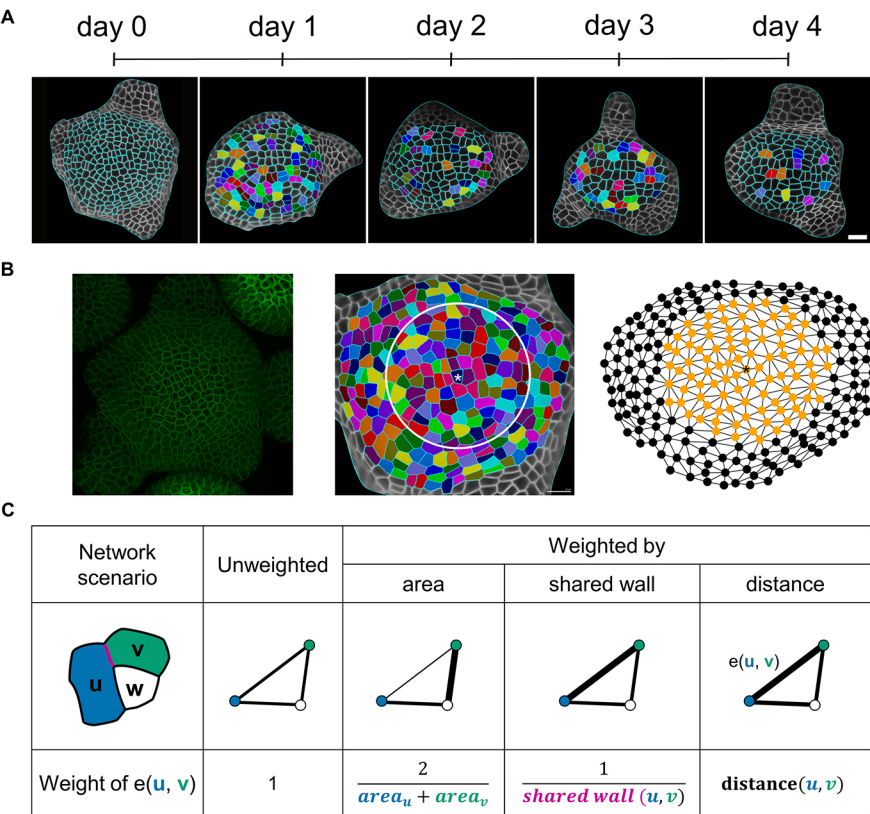


Fig. 1. Feature generation from three-dimensional (3D) images of the shoot apical meristem (SAM). (A) The surface of *A. thaliana* SAM was imaged every 24 h over 5 days. Pairs of dividing cells, depicted with the same colors, were determined manually (see Materials and Methods). (B) The 3D images of SAMs were converted to 2.5D surfaces by employing MGX (Barbier de Reuille et al., 2015) (left panel). The surface is abstracted according to its topology, capturing the connectivity of neighboring cells in a radius of 30 μ m (white circle) around the central cell(s), marked with an asterisk (center panel). The topology of the analyzed cells inside the circle is colored orange (right panel). Two nodes are connected by an edge if the cells they represent share a cell wall. (C) Four different network scenarios are considered: (1) unweighted edges, and edges weighted by (2) area, (3) a shared cell wall and (4) 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*, the inverse of the shared cell wall area (magenta) and 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 (a smaller thickness represents smaller weight values and cells being closer in terms of topological representation).

We defined the central zone of a SAM as the area covered by a circle of 30 μm radius around the most central and highest point in the analyzed SAM (Fig. 1B), and found that $28.9 \pm 17.9\%$ of cells divided per tissue between two successive tissue time points, with a total number of 605 dividing cells and 1458 non-dividing cells in 28 tissues (Fig. 1B).

Next, we represented the topology of the full tissue 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. As a result, the topology abstracts the tissue from an image into a neighborhood representation. For each cell in the central region, we calculated 17 properties, referred to as topological features (Table S1), in an unweighted network, in which every edge is of weight 1. We also applied different edge weights, which add cellular information into the topology based on the mean outer periclinal surface area, shared cell wall and distance of the cell centroids between two nodes representing those cells (Fig. 1C). In addition, we considered the surface area of each cell in the central zone as a biological feature for division prediction.

Previous studies have shown that a critical cell size threshold exists 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 coefficient (r) with the topological features (Fig. S2). We found that the degree of distance weighting was most closely correlated to the surface area ($r=0.75$). Nevertheless, the absolute value of the correlation with surface area was smaller than 0.3 for 47% of the features (98% showed correlation less than 0.7), with only three of the 50 considered features showing non-significant correlation ($P < 0.05$, Table S2). 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 included only features with an absolute value of the Pearson correlation coefficient (r) smaller than 0.3 (Fig. S2). In this way, we aimed to remove bias by considering features that may, to a certain extent, include information about surface area. As a result of these considerations, we trained five classifiers based on support vector machines (SVMs) with linear 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 the surface area ($r < 0.3$); and unweighted topological features. To this end, we selected all dividing and non-dividing cells from 20 tissue time steps (derived from six SAMs) for training the SVMs, and ensured balanced performance by weighting based on the frequency of dividing and non-dividing cells. We kept the data from the remaining eight tissue time steps (two SAMs) as a testing set (see Materials and Methods). Furthermore, we partitioned the 1445 selected cells into training and validation sets, and used six-fold cross validation to train the classifiers with each split being on plant (see Materials and Methods).

Although the training accuracy of the SVM using only the surface area was 73.8%, the training accuracy solely based on topological features was significantly higher – 81.6% ($P=0.0094$, paired t -test); this was also the case when the combined set of topological features and surface area was used, with a training accuracy of 81.2% ($P=0.0070$, paired t -test) showing that all feature sets contained information to be learned in order to predict cell division events. To ensure no overfitting, i.e. the model memorizing only training data, and to ensure transferability on unseen data, we applied the model to

validation and test data. Here, we observed a similar performance, with no difference in the validation accuracies between the feature sets for the three SVMs ($\sim 75\%$; see Fig. 2B). For the test SAM, the classifier based on the surface area alone exhibited the best performance, with an accuracy of 78.3%, closely followed by the SVM that was trained on the combination of topological features and area (76.7%), and the SVM that considered all topological features (75.9%, see Fig. 2C, Table S3). The area under the curve (AUC) of the receiver operating characteristic (ROC) curve, which is used as another measure of performance, yielded similar results (Fig. S3A, Table S3).

The removal of topological features that were highly correlated with area does not change the performance for training or for validation and test accuracy (Fig. 2B,C, Table S3). Although using only the topological features from the unweighted network scenario (Fig. 1C) decreased the training performance, the validation accuracy was only significantly reduced compared with the classifiers trained with both topological features and area, and when $r < 0.3$ ($P=0.0144$ and 0.0300 ; paired t -test). The final test accuracy was slightly reduced, but was comparable with that of the other classifiers at 69.8% (Fig. 2B,C, Table S3). 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 (Fig. S4). Therefore, we concluded that topological features and area can predict cell division independently of each other, whereas the performance could not be further increased by the combination of both feature types.

We hypothesize that feedback between topology and cell division is mediated by cell-to-cell interactions, suggesting that properties related to signal transduction and their influence in the network are important features in classification. Indeed, we found that harmonic centrality, which measures the sum of reciprocal distances of a node to all others, is of highest positive importance (Fig. 2D). Harmonic centrality implies that nodes that are closer to all nodes (Marchiori and Latora, 2000), and hence in the center of topology, are more likely to divide. Similarly, the next two positive most important features are based on Katz centrality. This centrality measure depends on the number of walks, rather than paths, of a particular length, whereby nodes that are involved in more walks to the others are considered more central (Katz, 1953), and has a similar interpretation to harmonic centrality. Interestingly, and in line with the previously discussed features, we also found that the versions of information centrality that weight the connections of cells (edges) based on shared wall, on equality (unweighted) and on distance are also of high importance in classification. If this centrality index measures the harmonic mean length of paths ending at a given node, suggesting the node is connected via many short paths to all other nodes, then it will be smaller. In addition, weighted node degree, absolute graph density in the second neighborhood and clustering coefficient denote the local density of subgraphs induced by a given node, suggesting that topologically dense regions are hotspots of cell divisions.

To further corroborate the biological relevance of these findings, we randomly permuted the labels and retrained the classifiers, repeating this procedure 100 times for each feature set (see Materials and Methods). We could not train a classifier that performed well and was able to generalize on the validation or test set, and exhibited accuracies expected by chance (Fig. S5A). Therefore, the classifiers trained on the randomized labels demonstrate that the features used capture information that is important for classification of dividing and non-dividing cells at 1-day intervals.

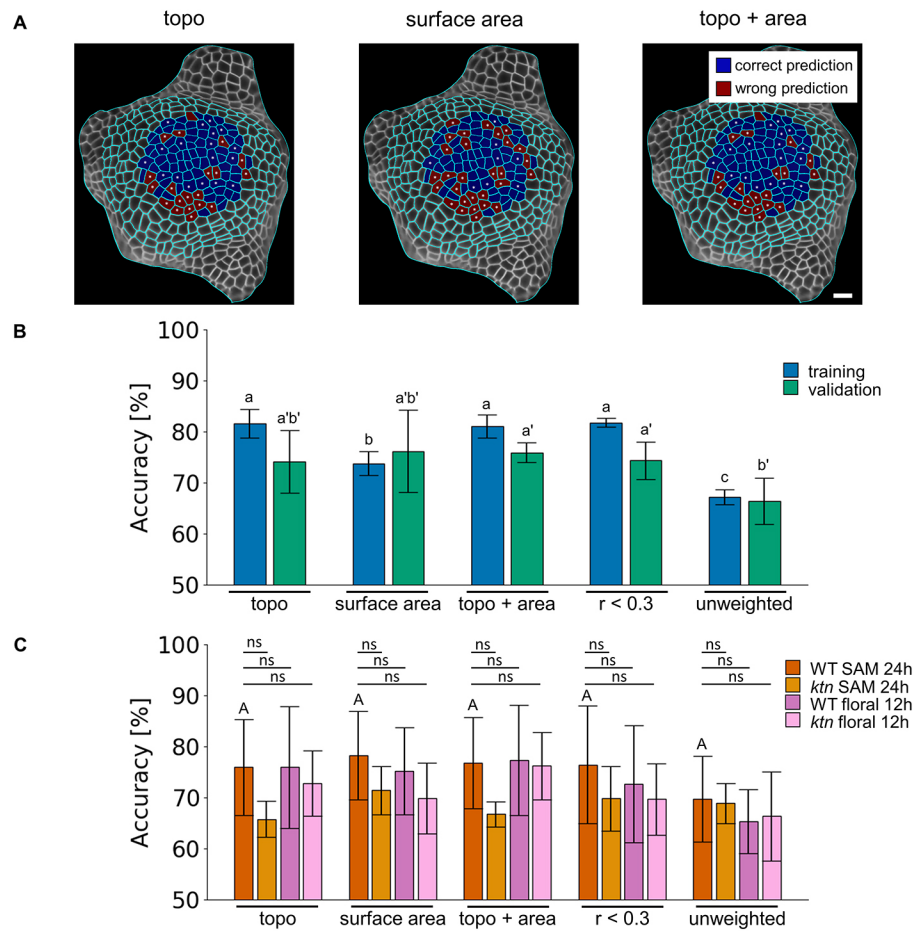


Fig. 2. Surface area and topology-based features generate similar predictions of cell division in the SAM that generalize for *ktn* mutant and floral meristems central region. (A) Comparison of predicted and observed division events on wild-type shoot apical meristem (SAM) test plant tissue from Fig. 1A day 0, highlighting correct predictions in blue and wrong predictions in red. Dividing cells are indicated with white dots. The predictions are from classifiers trained on different feature sets: combined topological features (topo), surface area and combined topological features with surface area (topo+area; from left to right). Scale bar: 10 μ m. (B,C) The accuracy of the support vector machine classifier to predict cell division of (B) wild-type SAM training (blue) and validation (green) in 24 h time steps, and (C) testing wild-type (dark orange) and *ktn* mutant (dark purple) SAMs in 24 h time steps, and wild-type (light orange) and *ktn* (light purple) floral meristem in 12 h time steps. Datasets from left to right: topo, surface area, topo+area, a reduced set of topological features that show an absolute Pearson correlation coefficient with surface area smaller than 0.3 ($r < 0.3$), and unweighted topological features (unweighted; derived from the unweighted network scenario). The performance on the training and validation set is determined from sixfold cross-validation as mean \pm s.d. The 10 most important features of the test classifier trained on the feature set $r < 0.3$ (coloring positive blue and negative coefficients orange) are shown. Different letters indicate significance between groups using Benjamini-Hochberg adjusted two-tailed paired *t*-test ($P < 0.05$). Statistical testing for differences of classifier performance for the training, validation and test sets is conducted separately (lowercase letters without prime, lowercase letters with prime and uppercase A, respectively). Lines indicate a Benjamini-Hochberg adjusted two-tailed Student's *t*-test between test wild-type SAM and other central apical meristems with non-significant (ns) pairs. $N_{WT}=28$ (20 tissue time steps for training-validation and eight for testing); $N_{ktn}=5$; $N_{WT_floral}=8$; $N_{ktn_floral}=7$; $n_{WT}=1445$ and 618 train-validation and test cells, respectively; $n_{ktn}=575$; $n_{WT_floral}=151$; $n_{ktn_floral}=295$. *N* denotes the number of tissues over the time steps considered; *n* denotes the number of train-validation cells.

To evaluate the transferability of the classifiers, we applied the SVMs trained on wild-type SAM data on the same feature sets generated from three different central meristems: SAMs of the *katanin1-2* (*ktn*) mutant, and floral meristems of wild type and *ktn* (tracked every 12 h because of their increased division speed compared with the SAM). Although the performance for predicting cell division events on the floral meristems was at the same level as the wild-type SAM test performance, the performance of the classifier slightly decreased in all feature sets from the *ktn* SAMs (non-significantly for accuracy and significantly reduced for area under the curve of all classifiers), except for the unweighted topological features (Fig. 2C, Fig. S3B). Although 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 and that similar

mechanisms exist in the central regions of both SAM and floral meristems.

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 informative in the time-dependent connectivity of daughter cells, we trained classifiers based on SVMs (with linear 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 adjacent to only one daughter cell: those adjacent to the daughter cell closer to the SAM center are labeled as class 0; those adjacent to the daughter cell further from the center are classified as class 1; neighboring cells

adjacent to both daughter cells are considered to be class 2 (Fig. S1C).

To predict changes in the local topology due to cell divisions, we first determined all neighbor-parent pairs of dividing parental cells 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. By including both parent and neighbor, we generated 96 unique topological-based properties for each neighbor-parent pair of the combined topological feature set (topo, Fig. S1B). 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 at the next time point by aligning the tissues at 1-day time intervals manually, and determining their adjacency with their neighbor with respect to the daughter cells (see Materials and Methods and Fig. S1B).

We excluded neighbor-parent pairs for which the neighbor also divided in the 1-day interval being considered to avoid bias from guessing which cell divides first (Fig. 3A). Using this procedure, we created 3015 neighbor-parent pairs with 900 representatives in each of the three classes (0, 1 and 2) from eight different SAMs, tracked every 24 h over 5 days. The data were split into three parts (training, validation and test data), such that the SAMs of two plants were kept as test data, while the rest of the plants were used in a nested fivefold 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 71.6% and 65.9%, respectively. The topological and biological features alone showed 12.8% and 23.4% reduction in validation accuracy compared with the combined classifier, and similar reduction in training accuracies ($P=0.012$, paired t -test with Benjamini-Hochberg correction). Regarding the performance on the test set, the combined classifier performed best, with an accuracy of 64.8%, followed by the classifier based on topological features alone with 53.3%. The classifier that used biological features alone had the worst accuracy (equivalent to guessing) of 50.2% (Fig. 3B, Table S4).

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 adjacent to only one cell (class 0 and 1) in comparison with the classifier based on the biological features (i.e. relative increase of 13.0% and 30.0% for class 0 and 1, respectively). In contrast, the SVM based on the biological features performed 29.4% 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 12.7% and 16.7% (relative increase compared with topological and biological features alone, respectively) while retaining high performance for all classes (Fig. 3C, Fig. S6). Investigating the reduced topological feature set (i.e. removing features with Pearson correlation coefficients larger than 0.3 with any biological feature) as well as considering only unweighted features resulted in similar validation and test accuracies compared with all topological feature trained classifiers (Fig. 3B,C, Table S4). Investigating the classifier

learning curves reveals that they did not suffer from high bias and variance, and that training was carried out on a sufficiently large number of pairs (Fig. S7). These findings indicate the importance of both topological as well as biological properties in predicting 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 100 times for each feature set. Although 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 that expected by chance (Fig. S5B). Furthermore, 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 (Table S4). Therefore, our findings demonstrated that the used features capture information important for classifying changes in local topology predictions surrounding dividing cells in 24 h intervals.

We also tested how well the trained classifiers based on the wild-type SAM data performed on the *ktn* mutant SAM and floral meristem, as well as on wild-type floral meristem. With the data from the *ktn* mutant SAM, we found a reduction in accuracy for the best-performing classifier when trained on the combined topological and biological, reduced feature set, and unweighted features (35.7%, 19.9% and 26.6% with $P<0.001$, $P<0.0463$ and $P<0.0103$ using Student's t -test, respectively, Fig. 3C). We could equally well predict floral meristem topological changes compared with the wild-type SAM test data, except for *ktn* floral meristem based on all topological features and wild-type floral meristems based on unweighted topological features (with relative reduction of 17.2% and 17.4%, and $P=0.0462$ and $P=0.0490$ using Student's t -test, respectively). The reduced performance on SAM data of classifiers trained on topological and biological features combined cannot be attributed to one specific class (Fig. S1C). 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 the prediction of 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 outlined in Fig. 4A). We compared the predicted and observed topologies by investigating the unweighted topological features (Fig. 1C, Table S1) 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 compare in a pairwise manner their unweighted topological features. We did not consider other network scenarios (see Fig. 1C) as we would need to estimate the weights for the topology, adding a layer of uncertainty. Here, 290 of the possible 437 non-dividing cells in the observed topology were also non-dividing in the predicted topology. For these cells, we calculated r of all unweighted features between observed and predicted topologies, with the information centrality showing the largest value of $r=0.71$, and 10 out of 17 features with $r>0.5$ (Fig. 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 (Fig. 4E, Fig. S8).

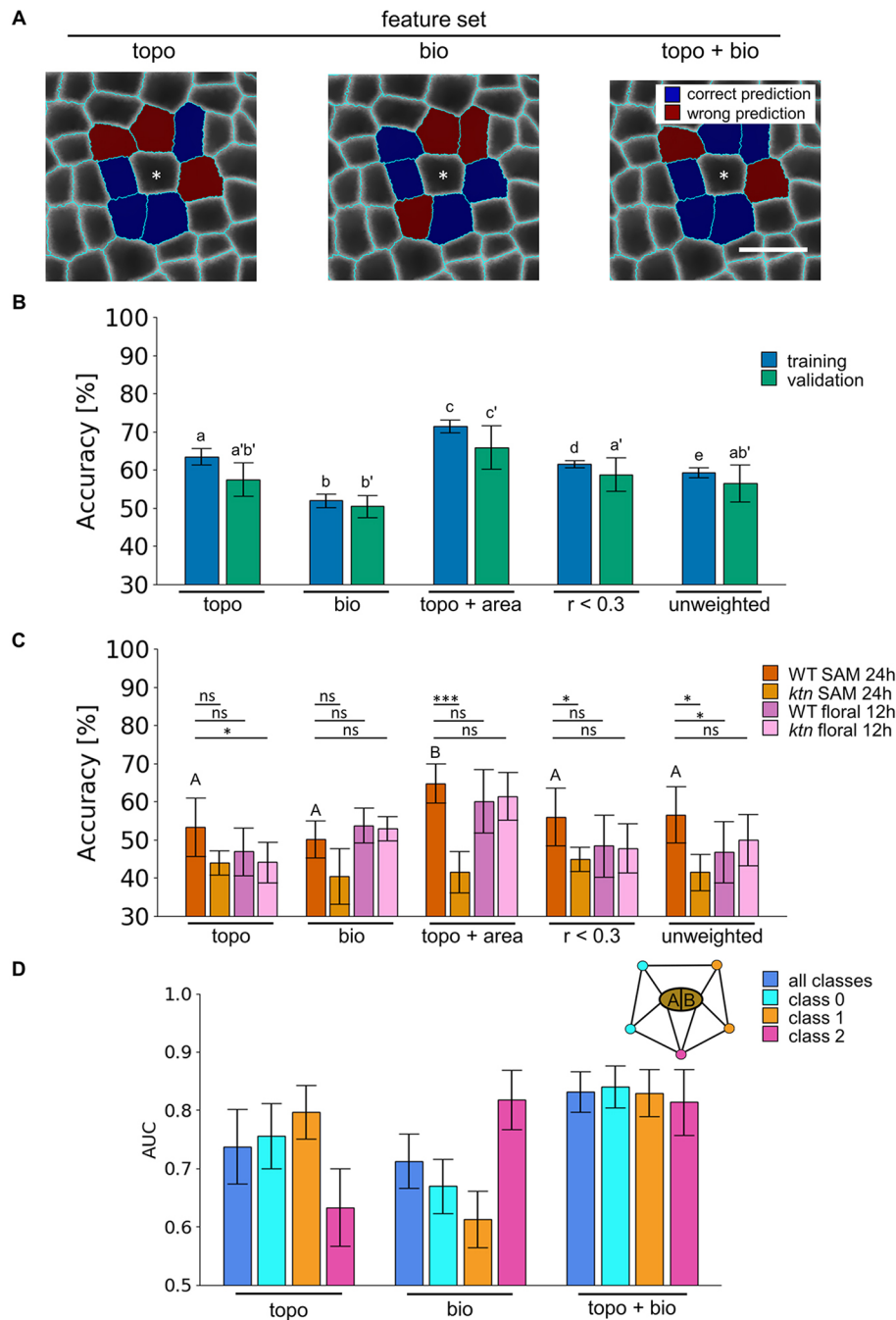


Fig. 3. Topological and biological features are required for accurate prediction of the local neighborhood after cell division in the SAM and *ktn* mutant shows reduced performance on SAM. (A) Comparison of predicted and observed local neighborhood changes around dividing cell (white asterisks) from wild-type shoot apical meristems (SAM) tissue after 24 h (as in Fig. 1A, day 0), highlighting correct (blue) and wrong (red) predictions. The predictions are made with classifiers trained on different feature sets: combined topological (topo, including features of the four network scenarios, see Fig. 1C, Fig. S1), biological (bio, including area, perimeter, shared cell wall and distance), and combined topological and biological features (topo+bio). Scale bar: 10 μ m. (B,C) Accuracy of the support vector machine classifier to predict local topological changes for dividing cell neighbor pairs of (B) wild-type SAM training (blue) and validation (green) in 24 h time steps, and (C) testing wild-type (dark orange) and *ktn* mutant (dark purple) SAMs in 24 h time steps, and wild-type (light orange) and *ktn* (light purple) floral meristem in 12 h time steps. (D) Area under the curve (AUC) of the receiver operating characteristic based on the wild-type SAM test data for the local neighborhood prediction of all classes (blue), and 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 trained on topo, bio, topo+bio, reduced set of topological features that show an absolute Pearson correlation coefficient with all biological features smaller than 0.3 ($r < 0.3$), and unweighted topological features (unweighted; derived from the unweighted network scenario). The performance on the training and validation set is determined from sixfold cross-validation as mean \pm s.d. Different letters indicate significance between groups using Benjamini-Hochberg adjusted two-tailed paired *t*-test: $P < 0.05$. Statistical testing for differences of classifier performance for the training, validation and test sets was conducted separately (lowercase letters without prime, lowercase letters with prime and uppercase letters, respectively). Lines indicate a Benjamini-Hochberg adjusted two-tailed Student's *t*-test between test wild type and other central apical meristems. ns, non-significant, $*P < 0.05$, $***P < 0.001$. $N_{WT} = 28$ (20 tissue time steps for training-validation and eight for testing); $N_{ktn} = 5$; $N_{WT_floral} = 8$; $N_{ktn_floral} = 7$; $n_{WT} = 2103$ and 912 train-validation and test cells, respectively; $n_{ktn} = 1042$; $n_{WT_floral} = 1646$; $n_{ktn_floral} = 2003$.

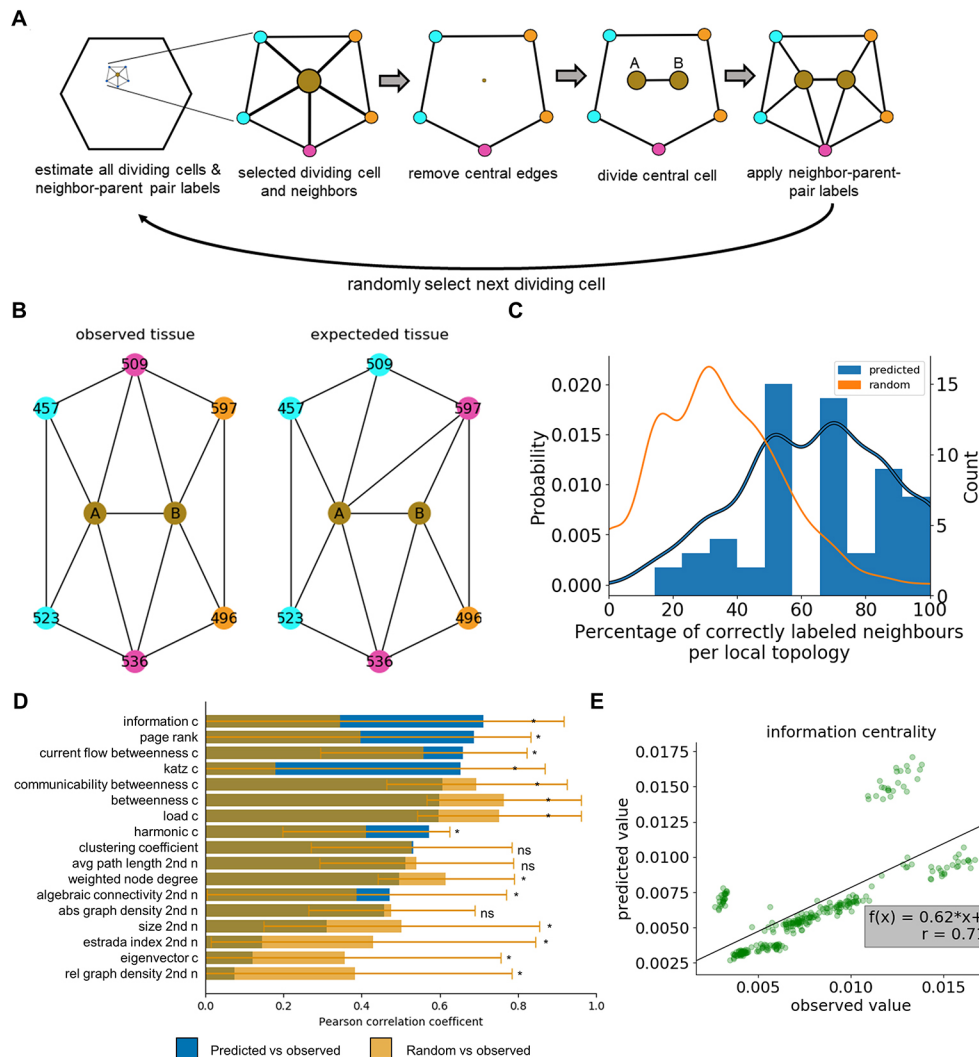


Fig. 4. Concordance between observed and predicted topologies. The cell connectivity network of the two test plants was predicted by applying classifiers for division event and topology prediction. (A) Illustration of the procedure applying the division and local topology classifiers to generate the topology of the next time point (24 h time interval). After division and topology prediction, a dividing cell (brown dot) is selected along with its neighbors (blue circles) and its adjacency relationship (edges, black lines) (left schematic). 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' and 'B'; representing the cell closer and farther away from the SAM center) that are adjacent to each other (middle three schematics). The daughter cells are connected with their neighbors based on the prediction from the local topology classifier (right schematic). 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 to the predicted or observed cells (numbers indicate the same cells) coloring their respective parent-neighbor class (cyan or orange indicate a cell connected only with daughter 'A' or 'B', respectively; magenta indicates a cell adjacent to 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 plants (blue) are compared with the density plot of randomly assigned parent-neighbor classes (orange). The difference between distributions is tested using a Kolmogorov–Smirnov test, $P < 0.001$. $N = 8$ tissue time steps, $n = 59$ local topologies. (D) The concordance between the observed and predicted topologies was quantified (mean, blue) for non-dividing cells in both topologies by calculating and ranking the Pearson correlation coefficient based on 17 topological features from the unweighted networks (see Fig. 1C; Table S1). 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 (data are mean \pm s.d., yellow). Benjamini–Hochberg-corrected two-tailed one-sample t -test: $*P < 0.05$; ns, non-significant. (E) The observed information centrality is plotted against the predicted information 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. $N_{WT} = 8$ tissue time steps, $n_{WT} = 290$ cells not dividing in both observed and predicted tissue.

For comparison, given the same test plant observed topologies, we selected all non-dividing cells predicted to divide and randomly connected the neighbors with the divided cells, which reoriented the local neighborhood; we then repeated the correlation analysis of the resulting topology with the observed topology. This 'random propagation' scenario allowed us to construct and investigate the most opposite example to our predictions (Fig. 4D, orange bars). Comparing the predicted and random propagation correlations

showed that six out of 17 topological features had a higher correlations and eight had a lower correlation (one sample t -test with Benjamini–Hochberg correction, $P < 0.05$). Although local features (calculated on the second neighborhood, i.e. the topology induced by cells in a radius of two cells around a specific cell), such as size (number of neighborhoods) and relative graph density (average number of neighbors per cell), showed small correlations, suggesting a need for a correction in the case of erroneous

topology recreations (e.g. creating non-planar topology), the overall structure in terms of flow centralities showed more and stronger correlations for the predicted propagation (Fig. 4D, blue versus 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 Fig. 4B). The distribution of correctly labeled neighbors per dividing cell was significantly shifted towards higher accuracy when comparing the predicted and random topology (Fig. 4C).

DISCUSSION

The biochemical pathway of cell division control has been extensively studied (Dewitte and Murray, 2003), but only recently have external cues also been considered in order to understand the effect of cell divisions in tissues (Hartig and Beck, 2006; Shimotohno et al., 2021). It is 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 a meristematic function and provide mechanical protection. This raises the issues of whether cell division and the subsequent local topology rearrangement are affected by tissue topology and whether tissue topology contains sufficient information for its accurate prediction.

Based on our extensive network-based modeling, 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 showing only an increase in size between dividing and non-dividing cells (Jackson et al., 2019). With the local topology rearrangement predicting model, we provide an alternative and network-based model to infer the changes applied to the network due to cell divisions. In addition, we successfully applied the division and local rearrangement classifier trained on wild-type SAM to the central region of the wild-type floral meristem, suggesting a similar mode of action. At the molecular scale 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), whereas KIP-related protein 4 has a similar function in the *A. thaliana* SAM niche (D'Ario et al., 2021). Modeling 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). Furthermore, 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. 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, manifested by the ability of cells to react to changes in turgor pressure produced by microtubules and cell wall rearrangement affecting cell divisions (for details of the feedback loop, see Sampathkumar, 2020), could explain the link between topology and the summed turgor and supracellular mechanical stress. Alternatively, the predictive ability of topological properties may result from long-distance communication of different phytohormones (Shimotohno et al., 2021) or could be due to cell-to-cell communication by plasmodesmata (Kitagawa and Jackson, 2017).

However, both cell division and the cell wall positioning affect tissue organization; a prime example is the effect of division

patterning in lateral root initiation (von 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 (Louveau 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, with limited information about the dividing and neighbor cell geometry, and allows reliable prediction of changes in the local topology. These local changes in the topology mirror the effect of cell wall placement on the tissue. In addition, we show that topological features alone suffice to accurately predict local topological changes. Although single topological properties are already used to estimate cell wall placement, the percentage of dividing epidermal cells in this study was only 5.2% (total $n=32/582$ dividing and non-dividing cells) per tissue every 11 h (Jackson et al., 2019). In contrast, our results rely on experiments in which cells divided more regularly, with an average of 28.9% of dividing cells per tissue every 24 h (total $n=605/1458$ dividing and non-dividing cells), allowing us to train robust classifiers. We show that the combination of both feature sets boosts performance of local topology reorientation prediction (Fig. 3), indicating that the inclusion of multiple viewpoints of information available to cells needs to be involved to solve the problem of cell wall placement in the SAM. This raises the question of how information of the topology is biologically transferred over the tissue, either via mechanical or via active and passive transduction of biochemical signals.

To demonstrate the generalizability of the classifiers, we show that they can be used to make accurate predictions for *ktn* mutants that are defective in mechanical feedback regulation. Our results indicate similar performance for the classifiers with biological feature sets from wild type and *ktn*. In contrast, the classifiers trained on topological features show reductions in performance in *ktn* compared with wild type. This difference in performance is not due to differences in topological features, as the normalized features showed similar distributions (Fig. S9). These results suggest a potential role for KATANIN in linking sub- and supracellular mechanical stress, which are known to affect leaf epidermal cells (Eng et al., 2021), and a role in positioning of the preprophase band, spindle and phragmoplast (Komis et al., 2017). In addition, the cell geometry of the *ktn* mutant differs from the wild type and might also influence the topology. Therefore, the combination of network-based modeling with machine learning provides a method for screening SAMs under different conditions and mutants. More specifically, reduction in test performance of the classifiers trained either on mutant surface area or on topological features compared with the wild type could suggest effects that disturb only functions related to the cell cycle or to a topological effect.

When combining division prediction and the resulting changes to the tissue, previous studies mostly focused on single cell division or on propagating tissues based on division likelihoods using the number of neighbors (Gibson et al., 2011) or using only area as a fixed threshold (Sahlin and Jönsson, 2010; Alim et al., 2012). Our classifier, on the other hand, incorporates more diverse tissue-level information. Here, we have combined our best classifiers to predict future tissue topology using the combined topological and biological features. Although the results of this propagation of classifiers are promising, the careful inspection of the findings, particularly with respect to planarity and topological properties of the reconstructed topologies indicate that further research should

consider simultaneous modeling of cell neighborhoods of higher orders to improve the reconstruction.

Furthermore, as information is not only being passed along the epidermis (L1-layer), the assessment of cell division events and their effects on the topology could be expanded beyond the epidermis of the SAM, as we know that the L2 and L3 layers play a vital role in supporting the meristematic function through the feedback of CLAVATA and WUSCHEL (Schoof et al., 2000). Transferring the classifiers to other plant species, such as maize (which 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-related mechanical stress (Sampathkumar et al., 2014), auxin gradients (due to PIN; Shi et al., (2018)), soil density 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.

MATERIALS AND METHODS

Plant materials and growth condition

We grew *Arabidopsis thaliana* wild-type (Wassilewskija ecotype) plants with the membrane reporter pUBQ10::acyl-YFP (previously described by 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 until 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 40×/0.8 water immersion objective of wild-type SAMs every 24 h for 5 days or *ktn* SAMs every 24 h for 3 days, and floral meristems of wild type and *ktn* every 12 h. Next, we used MorphoGraphX (MGX) (Barbier de Reuille et al., 2015) to obtain a 2.5 D surface mesh of the meristem L1 layer from the 3D images with at least two dividing cells; from there, we extracted 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, e.g. share a cell wall. We lineage tracked all cells for 1-day for SAMs and in 12 h time steps for floral meristems manually in MGX, resulting in tissue time steps with distinct topologies (Fig. 1A). We refer to dividing cells, at time *t* (days), as parent cells, and to 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, approximated by the highest curvature, then determined the largest distance to the boundary regions and emerging primordia. To this end, we compared the positions of cells with the average position over all cells.

Prediction of dividing cells

To predict cell division events (*n*) of central and non-peripheral cells, we employed python 3.8.1 for all further steps, e.g. to select all cells in a radius of 30 μm around the center for SAMs and 15 μm for floral meristems (Fig. 1B). Accordingly, we analyzed central cells and excluded peripheral cells. We considered a cell as peripheral with respect to a connectivity network when the topology induced by the adjacent nodes did not form a cycle. We then labeled each of the selected cells as a dividing (1) or non-dividing (−1) cell within each time step. In addition, we determined six sets of features (see below; for unreduced sets, see Fig. S1) for each cell.

Four out of five feature sets are based on entire tissue topology (i.e. including peripheral cells as well as cells outside the central region from the cellular connectivity network) and consist of topological features for all central cells; the fifth set includes only the surface area of the central cell. When 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 distance between the centroid positions of two adjacent cells as the edge weight (Fig. 1C). We calculated ten topological properties for each central cell and network scenario (see Table S1). Furthermore, we considered topological properties based on the induced subgraph of the first neighborhood (see Table S1). We estimated all properties in python using the networkx 2.4 package.

To train the classifiers for prediction of division events between two successive time points, we split the wild-type data from the eight plants into two datasets, a training-validation and a testing set, with six and two plants, respectively, while keeping three *ktn* plants, and two wild-type and three *ktn* floral meristems as a separate test set (resulting in 20 training-validation and 8 testing wild-type SAM, 5 *ktn* SAM, 8 wild-type and 7 *ktn* floral meristem tissue time steps, *N*). We split the data into training-validation and test sets to ensure no overfitting was carried out and rigorously examined our classifiers. As there are fewer dividing cells, their class is the minority class and we account for the imbalance by weighting the classifier and performance measures accordingly. We applied a support vector machine (SVM) with a linear kernel to predict the occurrence of cell division events within one time step, e.g. 1 day. To this end, we used the five different feature sets, 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 a surface area smaller than 0.3 (denoted by $r < 0.3$) (Fig. S1A). We trained each classifier with the topological properties as features of the training-validation set using six-fold cross-validation and always kept one plant as a validation set.

To this end, we z-normalized $[(X - \text{mean}) / \text{s.d.}]$ the topological properties with the corresponding mean and standard deviation (s.d.) for the training-validation and the wild-type test datasets. The *ktn* data were z-normalized using means and standard deviations. We estimated the hyperparameters on the training set by sixfold cross-validation using a grid search (sklearn 0.22.1, model_selection.GridSearchCV) with 100 regularly spaced hyperparameters for each power of 10 ($C: 10^{-4}$ – 10^1). 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, p_1 and p_2 , we used the relative difference $(100 \frac{|p_1 - p_2|}{\text{mean}(p_1, p_2)})$.

We quantified the difference between feature sets of training, validation or testing using a paired *t*-test (scipy 1.8.0, stats.ttest_rel) and the difference between testing of wild-type SAM and *ktn* SAM, and wild-type and *ktn* floral meristem using Student's *t*-test (scipy.stats.ttest_ind) with Benjamini-Hochberg testing correction (statsmodels 0.13.2, stats.multitest.multipletests).

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 100 times, retraining the classifier using the default linear 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 automatically selected all non-peripheral neighbor-parent pairs of dividing

cells (n). Next, we categorized the adjacency of these neighbors with respect to the newly divided (daughter) cells. To this end, we inspected whether 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 labeled each neighbor cell in a neighbor-parent pair as class 0, 1, or 2, according to whether it is adjacent to only cell 'A' or only cell 'B', or to both. We then predicted the local topology excluding and including dividing neighbors using five feature sets, similar to the analysis above.

To distinguish neighbor-parent pairs that 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 (Fig. S1B). 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 a reduced feature set ($r < 0.3$) including only topological features with $r < 0.3$ with all biological features. We performed training, validation and testing, and inspected the learning curves and estimated the information content of the used features, as specified in the analysis above for wild-type data using a linear kernel. Additionally, we tested the classifiers on the *ktn* SAM, and wild-type and *ktn* floral meristem 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 made for only one time step (24 h), as 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 the future adjacency of divided cells 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; and (3) we connected the daughter cells with their neighbors based on the prediction from the local topology classifier (Fig. 4A).

To evaluate the performance of the combined application of division and topology prediction, we calculated all unweighted topology features for the cells that are dividing in neither the predicted nor the observed topology. Next, we plotted the non-dividing cells observed against the predicted features, determined best linear fit and r for all unweighted topological properties. In addition, we divided all predicted non-dividing cells, randomly assigned labels to the neighbors of dividing cells depending on how they will be connected to the divided cells based on the training-validation set representation. We then repeated the correlation analysis from above 100 times (differently reconnecting topologies) and compared the correlations between predicted and random topology propagation using a one sample t -test (scipy.stats.ttest_1samp) and Benjamini-Hochberg testing correction.

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

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: A.S., Z.N.; Methodology: T.W.M., A.S., Z.N.; Software: T.W.M.; Validation: T.W.M.; Formal analysis: T.W.M.; Investigation: T.W.M., Y.W., R.K.; Resources: A.S., Z.N.; Data curation: T.W.M., Y.W., R.K.; Writing - original draft: T.W.M., Z.N.; Writing - review & editing: T.W.M., Y.W., R.K., A.S., Z.N.; Visualization: T.W.M.; Supervision: A.S., Z.N.; Project administration: A.S., Z.N.; Funding acquisition: A.S., Z.N.

Funding

Z.N., T.W.M., A.S., Y.W. and R.K. acknowledge the support from the Bundesministerium für Bildung und Forschung (project SHAPENET, 031L0177A to Z.N. and 031L0177B to A.S.). Open Access funding provided by the Max Planck Society. Deposited in PMC for immediate release.

Data availability

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

Peer review history

The peer review history is available online at <https://journals.biologists.com/dev/lookup/doi/10.1242/dev.201042.reviewer-comments.pdf>

References

- Alim, K., Hamant, O. and Boudaoud, A. (2012). Regulatory role of cell division rules on tissue growth heterogeneity. *Front. Plant Sci.* **3**, 174. doi:10.3389/fpls.2012.00174
- 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. et al. (2015). MorphoGraphX: a platform for quantifying morphogenesis in 4D. *eLife* **4**, 5864. doi:10.7554/eLife.05864
- Beauzamy, L., Louveaux, M., Hamant, O. and Boudaoud, A. (2015). Mechanically, the shoot apical meristem of Arabidopsis behaves like a shell inflated by a pressure of about 1 MPa. *Front. Plant Sci.* **6**, 1038. doi:10.3389/fpls.2015.01038
- Besson, S. and Dumais, J. (2011). Universal rule for the symmetric division of plant cells. *Proc. Natl. Acad. Sci. USA* **108**, 6294-6299. doi:10.1073/pnas.1011866108
- Bhavsar, H. and Panchal, M. H. (2012). A review on support vector machine for data classification. *Int. J. Adv. Res. Comp. Eng. Technol.* **1**, 185-189.
- Breuer, D., Nowak, J., Ivakov, A., Somssich, M., Persson, S. and Nikoloski, Z. (2017). System-wide organization of actin cytoskeleton determines organelle transport in hypocotyl plant cells. *Proc. Natl. Acad. Sci. USA* **114**, E5741-E5749.
- Camacho, D. M., Collins, K. M., Powers, R. K., Costello, J. C. and Collins, J. J. (2018). Next-generation machine learning for biological networks. *Cell* **173**, 1581-1592. doi:10.1016/j.cell.2018.05.015
- D'Ario, M. and Sablowski, R. (2019). Cell size control in plants. *Annu. Rev. Genet.* **53**, 45-65. doi:10.1146/annurev-genet-112618-043602
- D'Ario, M., Tavares, R., Schiessl, K., Desvoyes, B., Gutierrez, C., Howard, M. and Sablowski, R. (2021). Cell size controlled in plants using DNA content as an internal scale. *Science* **372**, 1176-1181. doi:10.1126/science.abb4348
- De Veylder, L., Beeckman, T. and Inzé, D. (2007). The ins and outs of the plant cell cycle. *Nat. Rev. Mol. Cell Biol.* **8**, 655-665. doi:10.1038/nrm2227
- Dewitte, W. and Murray, J. A. (2003). The plant cell cycle. *Annu. Rev. Plant Biol.* **54**, 235-264. doi:10.1146/annurev-arplant.54.031902.134836
- Eng, R. C., Schneider, R., Matz, T. W., Carter, R., Ehrhardt, D. W., Jönsson, H., Nikoloski, Z. and Sampathkumar, A. (2021). KATANIN and CLASP function at different spatial scales to mediate microtubule response to mechanical stress in Arabidopsis cotyledons. *Curr. Biol.* **31**, 3262-3274.e6. doi:10.1016/j.cub.2021.05.019
- Gibson, W. T., Veldhuis, J. H., Rubinstein, B., Cartwright, H. N., Perrimon, N., Brodland, G. W., Nagpal, R. and Gibson, M. C. (2011). Control of the mitotic cleavage plane by local epithelial topology. *Cell* **144**, 427-438. doi:10.1016/j.cell.2010.12.035
- Hartig, K. and Beck, E. (2006). Crosstalk between auxin, cytokinins, and sugars in the plant cell cycle. *Plant Biol.* **8**, 389-396. doi:10.1055/s-2006-923797
- Jackson, M. D. B., Duran-Nebreda, S., Kierzkowski, D., Strauss, S., Xu, H., Landrein, B., Hamant, O., Smith, R. S., Johnston, I. G. and Bassel, G. W. (2019). Global topological order emerges through local mechanical control of cell divisions in the Arabidopsis shoot apical meristem. *Cell Syst.* **8**, 53-65.e3. doi:10.1016/j.cels.2018.12.009
- Jones, A. R., Forero-Vargas, M., Withers, S. P., Smith, R. S., Traas, J., Dewitte, W. and Murray, J. A. H. (2017). Cell-size dependent progression of the cell cycle creates homeostasis and flexibility of plant cell size. *Nat. Commun.* **8**, 15060. doi:10.1038/ncomms15060
- Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika* **18**, 39-43. doi:10.1007/BF02289026
- Kitagawa, M. and Jackson, D. (2017). Plasmodesmata-mediated cell-to-cell communication in the shoot apical meristem: how stem cells talk. *Plants* **6**, 12. doi:10.3390/plants6010012

- Komis, G., Luptovčiak, I., Ovečka, M., Samakovli, D., Šamajová, O. and Šamaj, J. (2017). Katanin effects on dynamics of cortical microtubules and mitotic arrays in *Arabidopsis thaliana* revealed by advanced live-cell imaging. *Front. Plant Sci.* **8**, 866. doi:10.3389/fpls.2017.00866
- Li, Y., Liu, D., López-Paz, C., Olson, B. J. and Umen, J. G. (2016). A new class of cyclin dependent kinase in *Chlamydomonas* is required for coupling cell size to cell division. *Elife* **5**, e10767. doi:10.7554/eLife.10767
- Louveaux, M., Julien, J.-D., Mirabet, V., Boudaoud, A. and Hamant, O. (2016). Cell division plane orientation based on tensile stress in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **113**, E4294–E4303. doi:10.1073/pnas.1600677113
- Marchiori, M. and Latora, V. (2000). Harmony in the small-world. *Physica A* **285**, 539–546. doi:10.1016/S0378-4371(00)00311-3
- Nowak, J., Eng, R. C., Matz, T., Waack, M., Persson, S., Sampathkumar, A. and Nikoloski, Z. (2021). A network-based framework for shape analysis enables accurate characterization of leaf epidermal cells. *Nat. Commun.* **12**, 458. doi:10.1038/s41467-020-20730-y
- Pisner, D. A. and Schnyer, D. M. (2020). Support vector machine. In *Machine Learning* (ed. A. Mechelli and S. Vieira), pp. 101–121. Academic Press. doi:10.1016/B978-0-12-815739-8.00006-7
- Qi, F. and Zhang, F. (2020). Cell cycle regulation in the plant response to stress. *Front. Plant Sci.* **10**, 1765. doi:10.3389/fpls.2019.01765
- Reddy, G. V., Heisler, M. G., Ehrhardt, D. W. and Meyerowitz, E. M. (2004). Real-time lineage analysis reveals oriented cell divisions associated with morphogenesis at the shoot apex of *Arabidopsis thaliana*. *Development* **131**, 4225–4237. doi:10.1242/dev.01261
- Sahlin, P. and Jönsson, H. (2010). A modeling study on how cell division affects properties of epithelial tissues under isotropic growth. *PLoS One* **5**, e11750. doi:10.1371/journal.pone.0011750
- Sampathkumar, A. (2020). Mechanical feedback-loop regulation of morphogenesis in plants. *Development* **147**, dev177964. doi:10.1242/dev.177964
- Sampathkumar, A., Krupinski, P., Wightman, R., Milani, P., Berquand, A., Boudaoud, A., Hamant, O., Jönsson, H. and Meyerowitz, E. M. (2014). Subcellular and supracellular mechanical stress prescribes cytoskeleton behavior in *Arabidopsis* cotyledon pavement cells. *Elife* **3**, e01967. doi:10.7554/eLife.01967
- Schoof, H., Lenhard, M., Haecker, A., Mayer, K. F., Jürgens, G. and 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. doi:10.1016/S0092-8674(00)80700-X
- Shapiro, B. E., Tobin, C., Mjolsness, E. and Meyerowitz, E. M. (2015). Analysis of cell division patterns in the *Arabidopsis* shoot apical meristem. *Proc. Natl. Acad. Sci. USA* **112**, 4815–4820. doi:10.1073/pnas.1502588112
- Shi, B., Guo, X., Wang, Y., Xiong, Y., Wang, J., Hayashi, K.-I., Lei, J., Zhang, L. and Jiao, Y. (2018). Feedback from lateral organs controls shoot apical meristem growth by modulating auxin transport. *Dev. Cell* **44**, 204–216.e6. doi:10.1016/j.devcel.2017.12.021
- Shimotohno, A., Aki, S. S., Takahashi, N. and Umeda, M. (2021). Regulation of the plant cell cycle in response to hormones and the environment. *Annu. Rev. Plant Biol.* **72**, 273–296. doi:10.1146/annurev-arplant-080720-103739
- Uyttewaal, M., Burian, A., Alim, K., Landrein, B., Borowska-Wykręt, D., Dedieu, A., Peaucelle, A., Ludynia, M., Traas, J., Boudaoud, A. et al. (2012). Mechanical stress acts via katanin to amplify differences in growth rate between adjacent cells in *Arabidopsis*. *Cell* **149**, 439–451. doi:10.1016/j.cell.2012.02.048
- Varner, J. E. and Lin, L.-S. (1989). Plant cell wall architecture. *Cell* **56**, 231–239. doi:10.1016/0092-8674(89)90896-9
- von Wangenheim, D., Fangerau, J., Schmitz, A., Smith, R. S., Leitte, H., Stelzer, E. H. K. and Maizel, A. (2016). Rules and self-organizing properties of post-embryonic plant organ cell division patterns. *Curr. Biol.* **26**, 439–449. doi:10.1016/j.cub.2015.12.047
- Wang, Y. and Sampathkumar, A. (2020). Live cell imaging of microtubule cytoskeleton and micromechanical manipulation of the *Arabidopsis* shoot apical meristem. *J. Vis. Exp.* **159**, e60936. doi:10.3791/60936
- Willis, L., Refahi, Y., Wightman, R., Landrein, B., Teles, J., Huang, K. C., Meyerowitz, E. M. and Jönsson, H. (2016). Cell size and growth regulation in the *Arabidopsis thaliana* apical stem cell niche. *Proc. Natl. Acad. Sci. USA* **113**, E8238–E8246. doi: 10.1073/pnas.1616768113

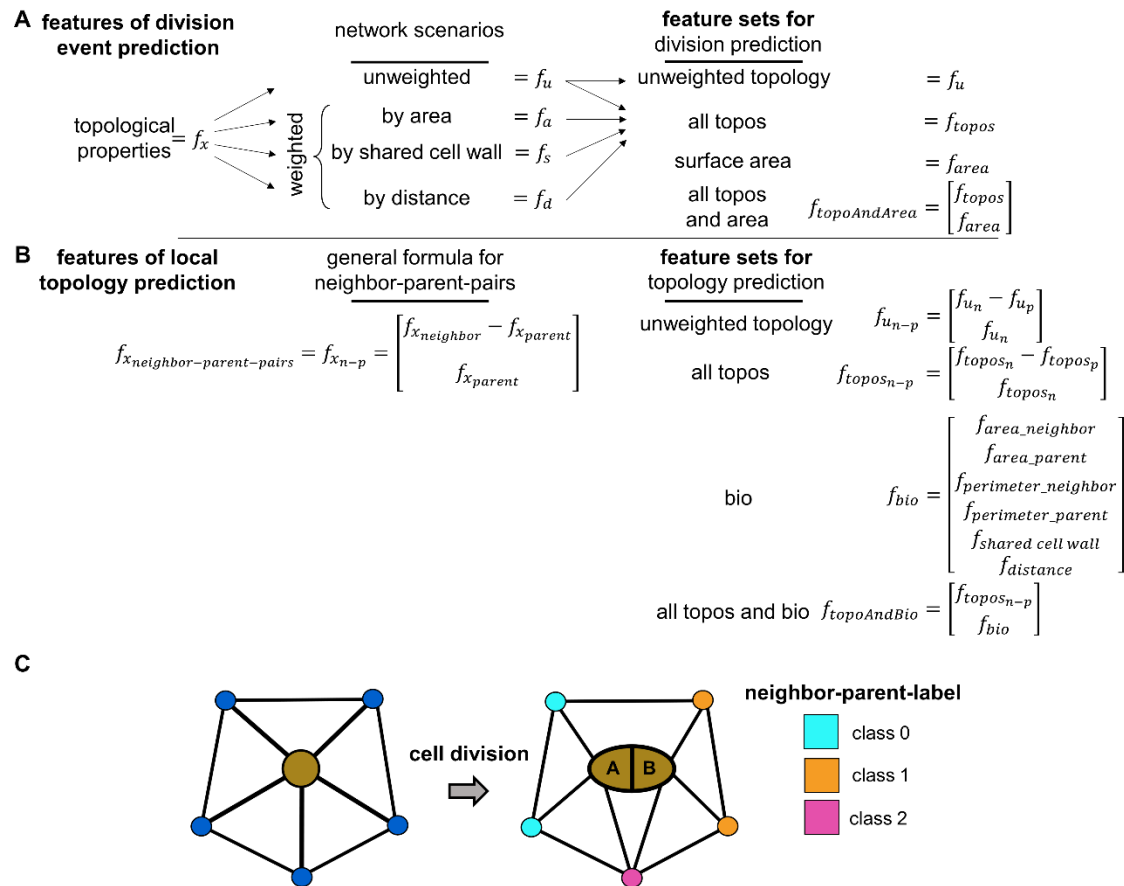


Fig. S1. Overview about feature sets for division event and local topology prediction as well as an example for local topology class assignment. (A) For division event prediction, 17 topological features (Supplementary Table 1) are calculated from four network scenarios (unweighted, weighted by area, shared cell wall, and distance, see Figure 1C), creating four different feature sets: unweighted topology (f_u), all topologies (f_{topos}), surface area (f_{area}), as well as topological features with surface area combined ($f_{toposAndArea}$) for all central cells. (B) For local topology prediction, the features are calculated for each neighbor-parent-pair of dividing parent cells using the difference in topological features of the neighbor ($f_{x_{neighbor}}$) and parent features ($f_{x_{parent}}$) as well as the features of the parent. Using this general formula, four feature sets are generated: 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 predict local topological changes from the earlier time point.

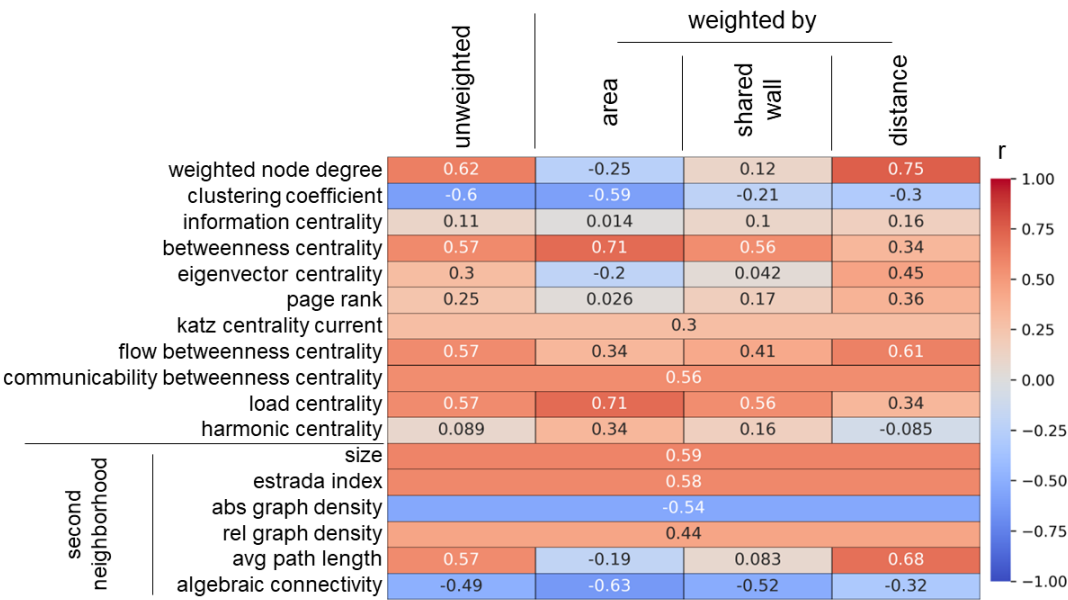


Fig. S2. Heat map of Pearson correlation coefficients between topological features and surface area. 17 topological features are calculated based on each of the four network scenarios (see Figure 1C): unweighted edges and edges weighted by area, shared wall, and distance. Majority of topological features exhibit small Pearson correlation coefficients (r , legend range from -1 (blue) to 1 (red)). $N_{WT} = 20$ tissue time steps, $n_{WT} = 1445$ cells.

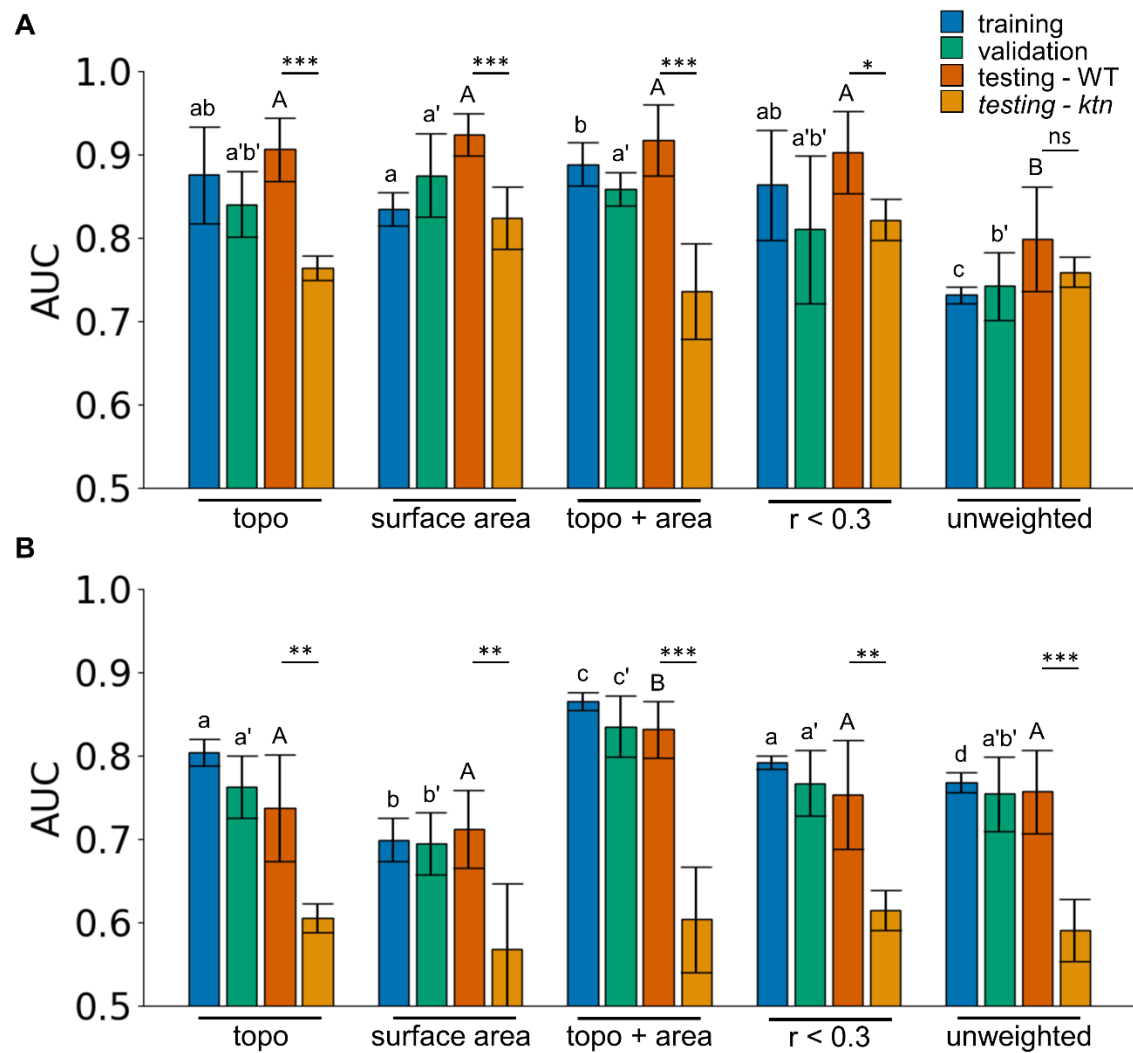


Fig. S3. Comparative analysis of predictions based using Area under the curve (AUC) of the ROC for feature sets. AUC of the ROC of the support vector machine (SVM) classifier on the training (blue), validation (green), and testing of wild type (orange) and *ktn* mutant (light orange) set of (A) division event and (B) local topology prediction. SVMs are trained on the combined topological features (topo), (A) surface area or (B) biological features (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 0.3 ($r < 0.3$), 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 six-fold cross-validation with mean and the standard deviation shown as error bars. Different letters indicate significance between groups using Benjamini-Hochberg corrected two-tailed paired t-test (p -value < 0.05).

Statistical testing for differences of classifier performance for the training, validation and test sets was conducted separately (small letter without and with apostrophe and capital letter, respectively). Lines indicate Benjamini-Hochberg corrected two-tailed Student's t-test between test WT and *ktn* SAM with non-significant (ns), p-value < 0.05 (*), p-value < 0.01 (*), p-value < 0.001 (***) pairs. $N_{WT} = 28$ (20 tissue time steps for training-validation and 8 for testing); $N_{ktn} = 5$; (A) $n_{WT} = 1445$ and 618, train-validation and test cells respectively; $n_{ktn} = 575$ and (B) $n_{WT} = 2103$ and 912, train-validation and test cells, respectively; $n_{ktn} = 1042$. bars represent mean \pm s.d.

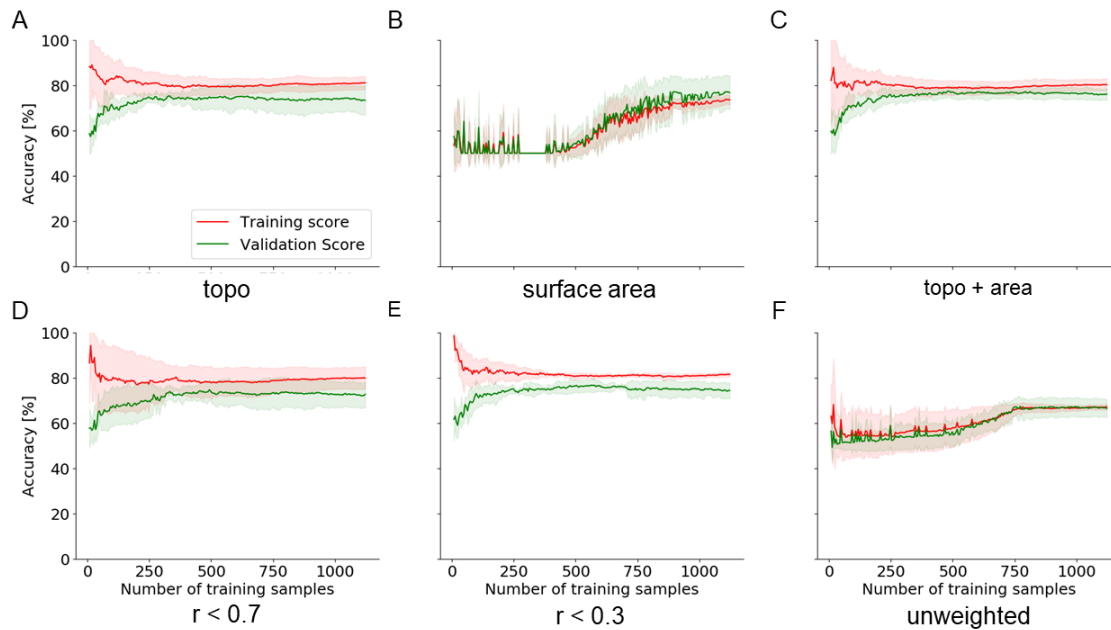


Fig. S4. Learning curves for the classifiers that predict division events. Learning curves of SVMs predicting cell division events, based on four different feature sets, showing the accuracy in the validation (green) and training (red) set (line: mean, area: ± 1 standard deviation). Feature sets: (A) 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 0.3, respectively, and (F) unweighted topological features (unweighted topology). $N_{WT} = 20$ tissue time steps; $n_{WT} = 1445$ train-validation cells.

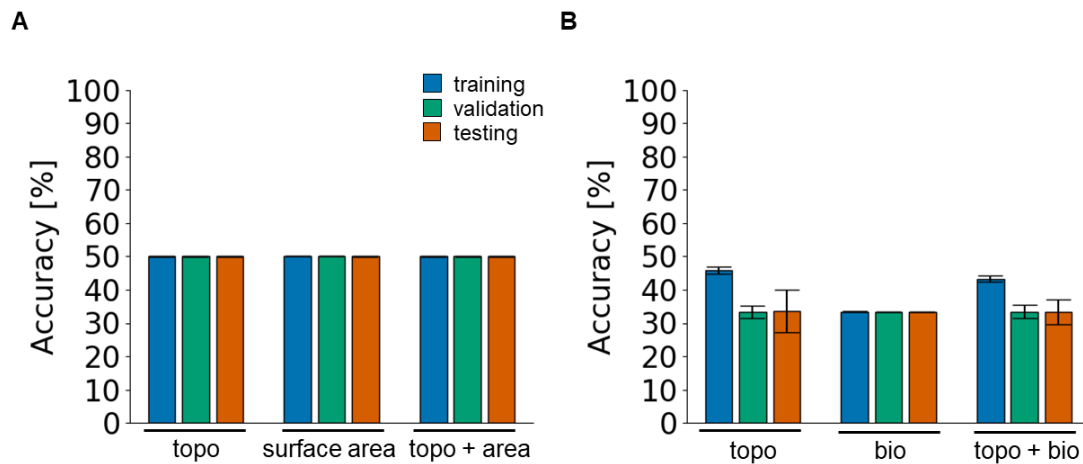


Fig. S5. Performance of classifiers trained on randomized labels. Support vector machines (SVMs) trained on surface area/biological features and unweighted topological features perform similarly with respect to the prediction of a division event and local topology of shoot apical meristems (SAM). Accuracy 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 six-fold cross-validation. The performance on the training and validation set is determined from six-fold cross-validation with mean and the standard deviation shown as error bars. Different letters indicate significance between groups using paired t-test: 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). $N_{WT} = 28$ (20 tissue time steps for training-validation and 8 for testing); $N_{ktn} = 5$; (A) $n_{WT} = 1445$ and 618, train-validation and test cells respectively; $n_{ktn} = 575$ and (B) $n_{WT} = 2103$ and 912, train-validation and test cells, respectively; $n_{ktn} = 1042$. bars represent mean \pm s.d.

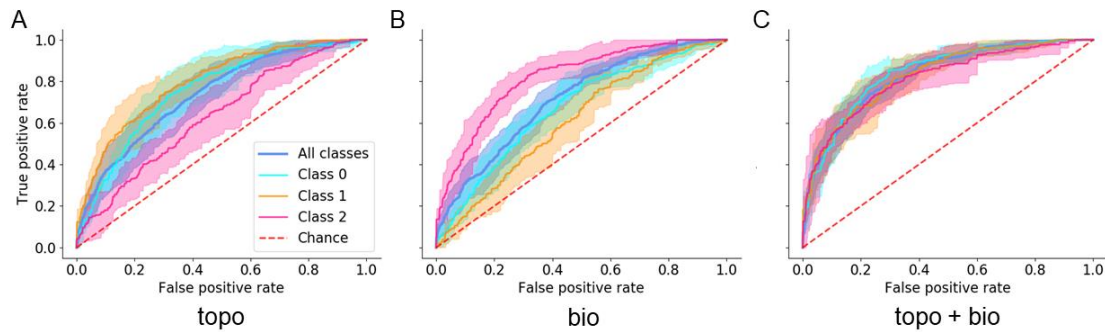


Fig. S6. Difference between classifiers trained on topological and biological features to predict local topology. Classifiers based on topological and biological features predict the three different classes of cells that are neighbors of cells that have divided in comparison to the previous time point. Receiver operating characteristic (ROC)-curve predicting cell division events on six-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 chance is marked with a red dashed line. $N_{WT} = 8$ test tissue time steps; $n_{WT} = 912$ test cells.

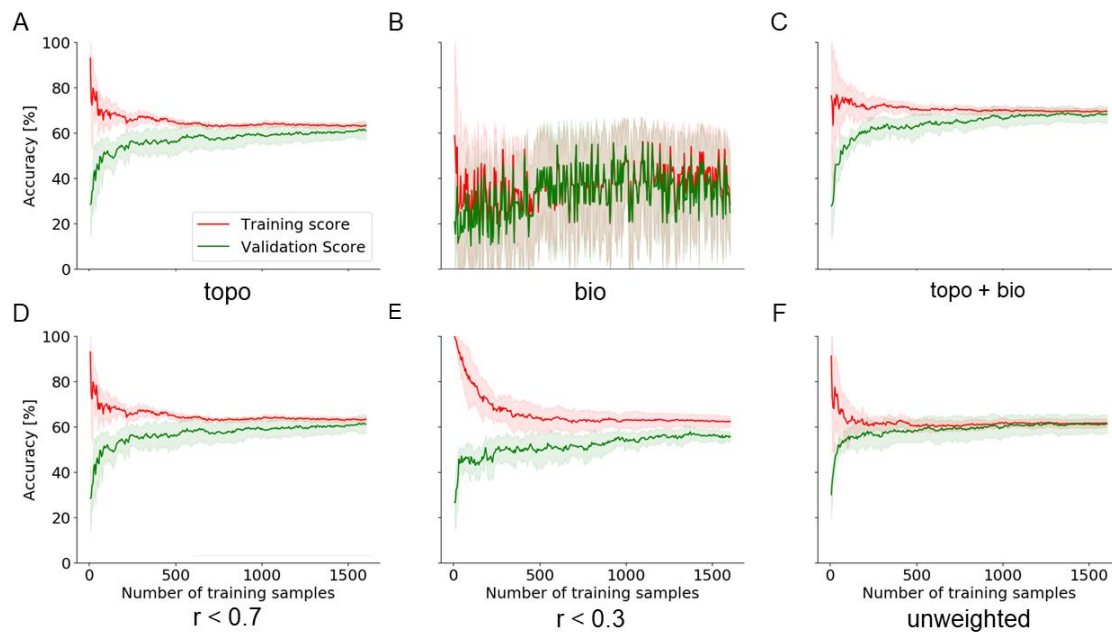


Fig. S7. Learning curves for the classifiers predicting local topological changes. Learning curves of support vector machines (SVMs) that predict local topology based on five different feature sets: (A) combined topological features (topo, including features calculated from the four network scenarios, 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 0.3, and (F) only unweighted features showing validation (green) and training (red) accuracy (line: mean, area: ± 1 standard deviation). $N_{WT} = 20$ tissue time steps; $n_{WT} = 2103$ train-validation cells.

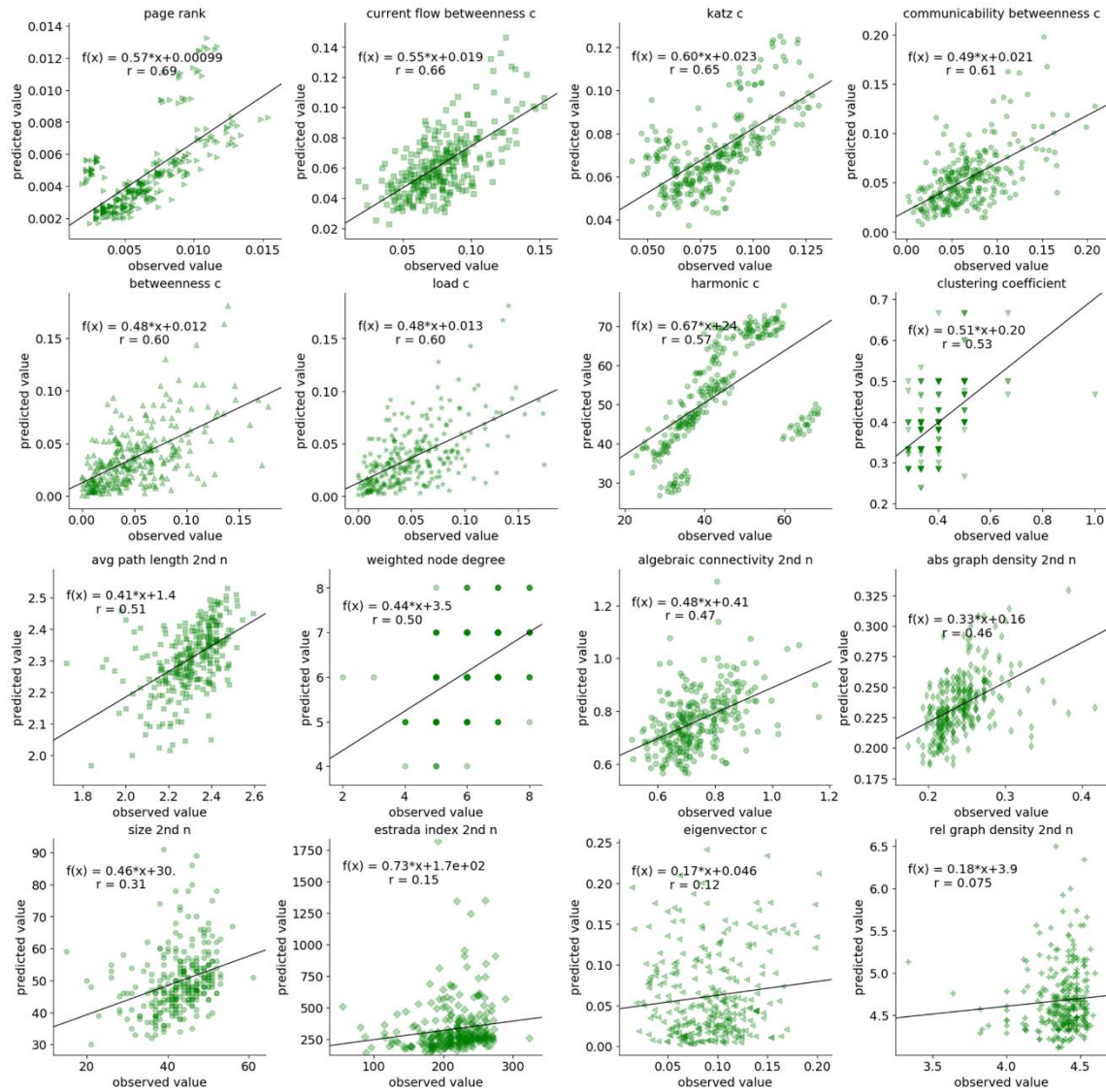


Fig. S8. Observed and predicted features from non-dividing cells plotted against each other.

The observed feature values are plotted against the predicted features values for each non-dividing cell 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: page rank, current flow betweenness c (centrality), katz c , communicability betweenness c , betweenness c , load c , harmonic c , clustering coefficient, average (avg) path length on 2 neighborhood (2nd n), weighted node degree, algebraic connectivity 2nd n , absolute (abs) graph density 2nd n , size 2nd n , estrada index 2nd n , eigenvector c , relative (rel) graph density 2nd n . The predicted tissue is estimated using the division and topology prediction classifiers trained on topological and topological with biological features of the test tissues, respectively (see Figure 4E). $N_{WT} = 8$ tissue time steps, $n_{WT} = 290$ cells.

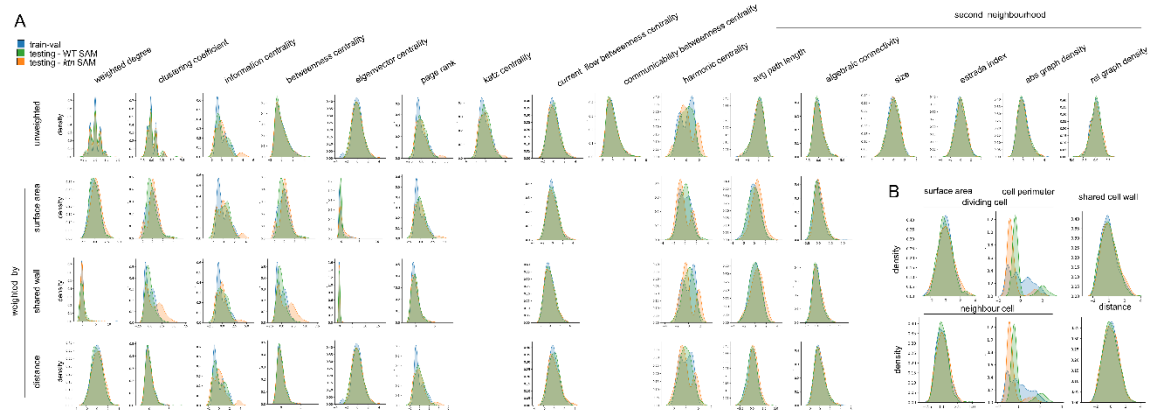


Fig. S9. Density distributions of topological and biological features of WT and ktn data. The density distributions of the normalized (A) topological and (B) biological features are plotted to differentiate between pooled samples used during training and validation from wild-type (WT; blue), and testing 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) $n_{WT} = 1445$ and 618, train-validation and test cells respectively; $n_{ktn} = 575$ (B) $n_{WT} = 2103$ and 912, train-validation and test cells respectively; $n_{ktn} = 1042$.

Table S1. Definition of network centralities applied on differently weighted network scenarios to be used as feature sets for cell division and local topology prediction. Each network centrality is applied on the unweighted, weighted by area, shared cell wall, and distance network scenario. The network centralities were concatenated together for each cell and used to train and predict different classifiers.

Network centrality	Definition	Reference
weighted node degree	$d(v) = \frac{\deg(v)}{ V - 1}$	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	$C_I(v) = \frac{ V - 1}{\sum_{u \in V} p_{uv}(v) - p_{uv}(u)}$	Sabidussi, 1966
betweenness centrality	$C_B(v) = \frac{2}{(V - 1)(V - 2)} \sum_{s,t \in V} \frac{\sigma_v(s,t)}{\sigma(s,t)}$	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	$C_{CFB}(v) = \frac{2}{(V - 1)(V - 2)} \sum_{s,t \in V} \tau_{st}(v)$	Newman, 2005
load centrality	$L = \frac{1}{(V - 1)(V - 2)} \sum_{s,t \in V} \theta_{s,t}(v)$	Goh, 2001
harmonic centrality	$C_H(v) = \frac{\sum_{u \in V} \frac{1}{d(u,v)}}{ V - 1}$	Marchiori & Latora, 2000
size	number of edges in a graph	
Estrada index	$EEG(G) = \sum_{j=1}^n e^{\lambda_j}$	Estrada, 2005
absolute graph density	$= \frac{2m}{n(n-1)}$	Wilks & Meara, 2002
relative graph density	$= \frac{2m}{n}$	
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

Table S2. Features ordered by Pearson correlation coefficient with area. Each network centrality was applied on the unweighted, weighted by area, shared cell wall, and distance network scenario. The Pearson correlation coefficient with area was calculated and were ordered from highest to lowest.

Feature name	Pearson correlation coefficient	p-value
weighted node degree weighted by distance	0.753377502	0
betweenness centrality weighted by area	0.707512442	1.0591074960896e-311
avg path length on 2 neighborhood weighted by distance	0.682352679	1.596195111956464e-281
weighted node degree	0.620070225	2.0669662274793384e-218
current flow betweenness centrality weighted by distance	0.611707287	4.812371330617275e-211
size on 2 neighborhood	0.594935365	5.057021194729162e-197
estrada index on 2 neighborhood	0.583528003	5.997153229184418e-188
betweenness centrality	0.5705616	6.373117152528312e-178
avg path length on 2 neighborhood	0.568504991	2.158701190828338e-176
current flow betweenness centrality	0.567399589	1.3876677872553513e-175
load centrality	0.565921643	1.6842614226438536e-174
communicability betweenness centrality	0.56005322	3.0531146350824394e-170
betweenness centrality weighted by shared wall	0.555983135	2.550625705762799e-167
eigenvector centrality weighted by distance	0.447873742	4.982402890283743e-102
rel graph density on 2 neighborhood	0.441796006	4.698909119359268e-99
current flow betweenness centrality weighted by shared wall	0.414482956	3.222728406558215e-86
page rank weighted by distance	0.359448329	1.0059855816964388e-63
harmonic centrality weighted by area	0.34300905	8.090933479758919e-58
current flow betweenness centrality weighted by area	0.341238842	3.2637221377844e-57
load centrality weighted by distance	0.339933021	9.025353613645715e-57
betweenness centrality weighted by distance	0.339062411	1.7603904997155677e-56
eigenvector centrality	0.300908478	2.868592129312282e-44
katz centrality	0.296715935	4.5579982830626784e-43
page rank	0.24762875	4.402906698088474e-30
page rank weighted by shared wall	0.170029847	9.015304252249096e-15
harmonic centrality weighted by shared wall	0.162714549	1.2198443119058207e-13
information centrality weighted by distance	0.162353154	1.3608683154738197e-13
weighted node degree weighted by shared wall	0.116561928	1.2480863722759725e-07
information centrality	0.113964977	2.343018162399504e-07
information centrality weighted by shared wall	0.102419688	3.4450652871124502e-06
harmonic centrality	0.089330332	5.2274644865479356e-05

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

Table S3. 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 six-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 0.3 (cor < 0.3) and unweighted topological features (unweighted).

topo

	train F1	train Acc	train precision	train Auc	val F1	val Acc	val precision	val Auc
split 0	85.07	84.63	82.68	0.91	43.38	61.01	79.19	0.81
split 1	83.36	82.90	81.16	0.91	80.85	76.32	67.86	0.90
split 2	83.91	83.62	82.42	0.91	70.79	73.56	79.06	0.83
split 3	75.64	75.85	76.29	0.82	79.12	76.25	70.59	0.84
split 4	81.57	80.89	78.76	0.89	79.90	80.08	80.65	0.86
split 5	82.21	81.79	80.35	0.89	79.20	77.43	73.45	0.86
train mean	81.96±3.04	81.61±2.84	80.28±2.21	0.89±0.03	72.21±13.32	74.11±6.17	75.13±4.81	0.85±0.03
test WT SAM	81.48±11.86	81.85±8.61	80.91±8.64	0.91±0.06	70.42±16.86	75.96±9.39	85.96±9.2	0.92±0.05
test ktn SAM					64.19±4.87	65.78±3.51	69.88±8.39	0.75±0.03
test WT floral					70.33±20.66	75.94±11.89	85.59±12.5	0.89±0.08
test ktn floral					65.16±12.61	72.77±6.44	91.65±13.37	0.9±0.05

area

	train F1	train Acc	train precision	train Auc	val F1	val Acc	val precision	val Auc
split 0	64.72	69.84	77.94	0.82	88.00	87.78	86.45	0.93
split 1	73.20	75.66	81.42	0.85	78.20	76.34	72.51	0.83
split 2	74.83	76.98	82.53	0.87	50.26	63.76	80.11	0.79
split 3	72.77	74.00	76.37	0.82	81.01	81.25	82.05	0.89
split 4	72.35	74.39	78.61	0.82	78.95	80.00	83.33	0.91
split 5	66.45	71.95	82.66	0.83	55.03	68.07	93.06	0.90
train mean	70.72±3.75	73.8±2.35	79.92±2.41	0.83±0.02	71.91±14.05	76.2±8.11	82.92±6.22	0.88±0.05
test WT SAM	76.74±14.89	79.83±9.36	84.16±8.02	0.91±0.05	72.54±16.77	78.3±8.64	90.45±7.93	0.92±0.02
test ktn SAM					66.78±8.76	71.43±4.70	78.43±3.45	0.82±0.04
test WT floral					67.21±14.66	75.24±8.54	93.49±7.19	0.93±0.04
test ktn floral					55.75±14.00	69.87±6.97	98.8±3.40	0.93±0.03

topo with area

	train F1	train Acc	train precision	train Auc	val F1	val Acc	val precision	val Auc
split 0	82.79	82.20	80.16	0.89	69.27	74.65	87.92	0.87
split 1	82.28	81.99	80.98	0.91	81.43	77.19	68.67	0.89
split 2	83.88	83.50	82.02	0.91	70.05	72.90	78.29	0.83
split 3	75.99	76.30	77.00	0.83	79.70	77.08	71.52	0.85
split 4	81.23	80.76	79.31	0.89	78.71	78.58	78.25	0.86
split 5	82.04	81.57	80.01	0.90	75.75	74.98	73.51	0.85
train mean	81.37±2.54	81.06±2.28	79.91±1.56	0.89±0.03	75.82±4.67	75.90±1.90	76.36±6.21	0.86±0.02
test WT SAM	81.71±11.89	81.98±8.61	80.64±8.49	0.91±0.06	71.78±16.57	76.75±8.94	85.76±8.95	0.92±0.04
test ktn SAM					64.86±5.92	66.75±2.47	70.26±7.28	0.74±0.03
test WT floral					72.16±18.17	77.33±10.81	87.32±11.27	0.89±0.08
test ktn floral					69.88±10.61	76.22±6.58	94.97±10.74	0.91±0.05

cor < 0.3

	train F1	train Acc	train precision	train Auc	val F1	val Acc	val precision	val Auc
split 0	83.74	83.21	81.19	0.89	56.72	67.30	83.83	0.83
split 1	81.72	81.29	79.88	0.90	80.85	76.32	67.86	0.91
split 2	83.02	82.22	79.48	0.90	70.64	72.49	75.73	0.82
split 3	82.62	82.08	80.22	0.88	78.74	77.50	74.63	0.87
split 4	81.26	80.73	79.12	0.89	75.17	74.54	73.35	0.83
split 5	81.81	81.22	79.33	0.89	79.40	78.11	74.98	0.87
train mean	82.36±0.85	81.79±0.82	79.87±0.69	0.89±0.01	73.59±8.25	74.38±3.68	75.06±4.70	0.85±0.03
test WT SAM	81.08±9.31	81.28±7.01	81.13±8.36	0.91±0.06	71.31±19.30	76.43±11.50	84.16±11.85	0.90±0.05
test ktn SAM					69.25±11.41	69.83±6.35	71.36±7.96	0.82±0.03
test WT floral					63.41±27.92	72.62±11.45	69.92±27.19	0.87±0.05
test ktn floral					59.27±15.65	69.73±6.99	91.13±14.29	0.87±0.07

unweighted

	train F1	train Acc	train precision	train Auc	val F1	val Acc	val precision	val Auc
split 0	68.40	69.52	71.00	0.76	59.70	61.42	62.49	0.68
split 1	65.15	66.61	68.14	0.74	65.14	64.33	63.69	0.69
split 2	67.92	68.84	70.00	0.75	57.19	62.58	66.80	0.71
split 3	63.39	66.20	69.14	0.73	76.61	73.54	68.65	0.78
split 4	64.66	65.56	66.38	0.72	72.46	71.50	70.09	0.81
split 5	65.76	66.29	66.82	0.73	63.47	64.93	66.23	0.72
train mean	65.88±1.77	67.17±1.47	68.58±1.65	0.74±0.01	65.76±6.81	66.38±4.52	66.33±2.63	0.73±0.05
test WT SAM	68.83±10.35	70.22±7.12	70.79±6.23	0.77±0.09	68.35±12.59	69.76±8.43	71.23±9.47	0.80±0.06
test ktn SAM					68.06±6.35	68.88±3.96	69.15±1.30	0.76±0.02
test WT floral					60.51±6.18	65.36±6.23	74.31±15.6	0.76±0.09
test ktn floral					58.23±13.71	66.38±8.72	75.45±14.33	0.74±0.10

Table S4. 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 six-fold cross-validations, as well as retraining on train+val data testing on test data (testing) (A) excluding or (B) including dividing neighbors, 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 smaller than 0.3 (cor < 0.3), respectively, and unweighted topological features (unweighted).

A

topo

	train F1	train c0 F1	train c1 F1	train c2 F1	train Acc	train precision	train c0 precision	train c1 precision	train c2 precision	train Auc	val F1	val c0 F1	val c1 F1	val c2 F1	val Acc	val precision	val c0 precision	val c1 precision	val c2 precision	val Auc
split 0	66.81	69.10	72.46	58.94	66.94	66.75	67.71	71.53	61.02	0.82	51.25	53.90	55.04	44.82	51.30	51.66	49.78	59.32	45.90	0.71
split 1	62.47	62.96	68.53	55.94	62.53	62.47	64.31	66.98	56.11	0.80	62.21	65.99	71.30	49.35	62.69	62.19	67.78	66.12	52.67	0.81
split 2	65.77	67.89	72.70	56.73	65.94	65.69	66.89	71.31	58.86	0.83	51.66	55.63	53.42	45.92	51.70	51.77	53.20	56.01	46.10	0.72
split 3	60.81	62.26	66.72	53.44	60.94	60.75	62.57	64.72	54.96	0.79	58.68	61.73	63.46	50.86	58.43	59.71	58.85	71.85	48.42	0.78
split 4	62.81	64.84	68.51	55.08	62.95	62.75	64.74	66.58	56.92	0.80	59.56	58.62	68.68	51.38	59.85	59.47	58.65	65.53	54.23	0.78
split 5	61.58	62.90	66.82	55.01	61.60	61.55	63.04	66.26	55.36	0.79	60.74	60.52	66.15	55.57	60.93	61.48	68.43	60.58	55.44	0.77
train mean	63.38±2.18	64.99±2.62	69.29±2.44	55.84±1.68	63.48±2.21	63.33±2.17	64.88±1.88	67.90±2.59	57.21±2.12	0.80±0.02	57.35±4.31	59.43±3.98	63.01±6.66	49.65±3.58	57.48±4.42	57.71±4.34	59.45±6.87	63.23±5.20	50.46±3.83	0.76±0.04
test WT SAM	63.00±7.77	64.83±8.97	68.64±11.74	55.53±8.14	63.53±7.69	64.16±7.07	65.37±8.55	69.25±7.87	57.86±9.99	0.81±0.06	51.51±8.97	53.65±19.46	60.21±8.55	40.67±9.82	53.29±7.64	53.73±7.32	54.59±6.47	61.47±8.77	45.12±9.79	0.74±0.06
test ktn SAM											43.54±3.50	44.16±4.50	41.22±8.61	45.24±4.16	44.01±3.21	44.54±3.38	42.99±4.84	43.73±4.68	46.92±6.27	0.61±0.02
test WT floral											45.94±7.59	48.92±5.20	39.64±17.2	49.27±3.49	46.91±6.26	46.96±7.0	47.66±8.39	45.48±14.94	47.73±3.60	0.66±0.07
test ktn floral											42.91±6.34	47.38±6.51	36.46±12.78	44.88±4.11	44.15±5.31	45.85±4.69	44.44±6.72	50.57±8.46	42.52±5.36	0.62±0.06

bio

	train F1	train c0 F1	train c1 F1	train c2 F1	train Acc	train precision	train c0 precision	train c1 precision	train c2 precision	train Auc	val F1	val c0 F1	val c1 F1	val c2 F1	val Acc	val precision	val c0 precision	val c1 precision	val c2 precision	val Auc
split 0	42.32	2.27	55.01	69.67	51.18	49.84	41.40	42.43	65.68	0.67	38.31	2.00	51.64	61.28	46.37	44.54	34.01	39.11	60.49	0.64
split 1	48.35	23.41	53.67	67.97	52.30	55.77	58.68	42.77	65.85	0.71	48.62	29.03	47.49	69.34	50.45	50.77	44.98	39.97	67.38	0.72
split 2	54.37	51.31	44.21	67.59	55.02	54.21	52.16	47.95	62.51	0.74	52.91	53.99	36.11	68.62	54.53	53.42	49.51	47.63	63.13	0.76
split 3	52.09	38.10	50.53	67.65	53.27	54.52	54.94	43.37	65.36	0.72	44.78	15.45	51.66	67.22	49.79	50.69	46.33	40.78	64.95	0.70
split 4	41.60	3.86	53.97	66.97	49.79	51.33	48.24	41.39	64.37	0.67	43.63	2.00	56.18	73.30	53.46	70.11	100.00	44.16	66.18	0.68
split 5	42.08	3.28	54.68	68.29	50.47	48.76	38.80	41.98	65.50	0.69	40.46	4.07	51.51	65.80	48.19	49.73	46.29	39.48	63.43	0.68
train mean	46.80±5.12	20.37±19.03	52.01±3.78	68.02±0.84	52.00±1.77	52.40±2.58	49.02±7.07	43.32±2.16	64.88±1.16	0.70±0.03	44.82±4.87	17.76±18.84	49.10±6.33	67.59±3.65	50.46±2.83	53.21±8.01	53.52±21.35	41.85±3.07	64.26±2.24	0.69±0.04
test WT SAM	52.92±4.88	46.80±7.81	43.06±8.37	68.90±7.58	53.93±4.89	53.46±5.19	50.30±7.93	45.17±6.24	64.91±7.03	0.74±0.05	48.23±4.05	38.72±10.84	40.23±9.50	65.76±6.54	50.16±4.88	49.60±5.43	45.54±10.80	42.65±7.24	60.62±3.58	0.71±0.05
test ktn SAM											40.02±7.25	34.72±9.57	39.26±4.76	46.08±12.19	40.45±7.25	40.28±7.16	37.52±7.80	40.82±6.40	42.51±12.20	0.57±0.08
test WT floral											52.35±4.51	44.69±5.19	39.41±11.24	72.95±4.19	53.75±4.59	53.12±5.80	47.80±8.75	42.41±6.90	68.15±6.74	0.75±0.03
test ktn floral											50.69±2.97	34.95±10.66	41.48±10.73	75.64±4.06	52.96±3.14	51.43±3.37	41.00±6.85	42.39±3.66	70.91±6.91	0.74±0.03

topoAndBio

	train F1	train c0 F1	train c1 F1	train c2 F1	train Acc	train precision	train c0 precision	train c1 precision	train c2 precision	train Auc	val F1	val c0 F1	val c1 F1	val c2 F1	val Acc	val precision	val c0 precision	val c1 precision	val c2 precision	val Auc
split 0	71.76	70.59	70.65	74.02	71.74	71.82	68.75	72.01	74.71	0.87	58.17	55.48	56.45	62.58	58.31	58.27	57.48	58.35	58.97	0.77
split 1	72.38	71.00	72.35	73.80	72.37	72.41	70.42	71.97	74.83	0.87	76.77	73.70	81.10	75.50	76.94	77.65	73.13	74.72	85.09	0.90
split 2	74.09	72.16	75.79	74.33	74.10	74.10	71.95	75.23	75.12	0.88	63.47	63.51	58.73	68.16	63.58	64.00	58.46	64.51	69.02	0.82
split 3	68.97	67.43	67.72	71.76	68.95	69.01	66.57	67.58	72.89	0.85	67.44	68.82	64.68	68.82	67.61	68.79	61.48	75.08	69.81	0.84
split 4	71.97	70.39	72.66	72.85	71.97	72.00	70.82	71.07	74.13	0.87	65.08	60.03	66.46	68.74	65.21	65.65	66.28	60.71	70.95	0.85
split 5	69.95	68.39	68.39	72.19	69.95	69.97	67.47	70.13	72.33	0.86	63.63	62.45	62.39	66.04	63.80	63.96	65.62	58.82	67.60	0.83
train mean	71.52±1.66	69.99±1.60	71.41±2.59	73.16±0.96	71.51±1.67	71.55±1.66	69.33±1.90	71.33±2.30	74.00±1.04	0.86±0.01	65.76±5.66	64.00±5.89	64.97±7.97	68.31±3.88	65.88±5.68	66.36±5.93	63.57±5.26	65.33±7.06	70.24±7.71	0.84±0.04
test WT SAM	71.89±9.01	70.12±11.09	70.84±13.78	74.70±8.27	72.12±8.71	72.40±9.04	70.17±10.77	70.69±11.55	76.34±9.87	0.87±0.05	64.26±5.48	64.39±10.76	64.92±6.61	63.49±7.32	64.79±5.17	65.64±5.24	64.62±4.33	67.14±10.58	65.16±6.00	0.83±0.03
test ktn SAM											41.24±5.25	42.86±4.19	44.08±7.07	36.79±7.05	41.66±5.41	42.66±6.78	40.31±5.24	45.08±3.84	42.59±13.55	0.60±0.06
test WT floral											59.58±8.83	55.18±10.86	53.14±14.40	70.42±3.59	60.08±8.29	60.45±9.08	56.35±12.53	56.36±13.54	68.63±7.79	0.79±0.07
test ktn floral											60.79±6.51	56.81±7.75	53.76±9.90	71.80±3.79	61.41±6.24	61.53±6.78	56.15±7.85	61.02±10.30	67.42±6.01	0.78±0.05

r < 0.3

186

187

	train F1	train c0 F1	train c1 F1	train c2 F1	train Acc	train precision	train c0 precision	train c1 precision	train c2 precision	train Auc	val F1	val c0 F1	val c1 F1	val c2 F1	val Acc	val precision	val c0 precision	val c1 precision	val c2 precision	val Auc
split 0	62.57	63.01	69.35	55.34	62.60	62.54	63.13	68.74	55.73	0.80	52.55	63.88	57.45	46.31	52.44	54.16	58.97	44.99	0.71	
split 1	61.95	64.28	68.02	53.54	62.09	61.88	64.42	66.00	55.21	0.79	64.49	68.86	73.98	50.63	65.07	64.34	66.48	70.88	55.66	0.83
split 2	62.12	63.88	69.57	52.93	62.30	62.04	64.46	67.18	54.49	0.81	53.90	55.54	56.43	49.72	53.89	53.93	54.45	57.58	49.77	0.72
split 3	59.56	61.12	66.65	50.92	59.83	59.49	60.61	63.90	53.97	0.78	61.00	65.02	66.75	50.34	61.11	61.67	60.19	73.18	51.63	0.79
split 4	60.85	62.25	67.77	52.55	61.03	60.78	62.43	65.50	54.40	0.78	60.25	60.68	65.82	54.26	60.36	60.22	59.78	64.60	56.28	0.78
split 5	61.18	63.04	67.42	53.09	61.28	61.12	63.03	66.07	54.25	0.79	59.74	57.04	68.46	53.74	59.99	60.95	68.35	63.63	50.86	0.77
train mean	61.37±0.99	62.93±1.04	68.13±1.03	53.06±1.31	61.52±0.94	61.31±1.00	63.01±1.31	66.23±1.49	54.68±0.60	0.79±0.01	58.66±1.15	60.32±5.47	64.91±1.15	50.83±2.65	58.91±4.34	58.97±4.21	60.57±5.40	64.81±5.70	51.53±3.79	0.77±0.04
test WT SAM	60.58±6.02	61.23±12.04	66.67±12.77	53.83±10.76	61.18±5.00	61.52±9.22	62.36±10.97	65.29±10.33	56.91±11.63	0.78±0.08	53.82±5.58	54.50±18.26	63.16±8.82	43.79±10.23	56.02±7.60	57.63±9.45	54.10±5.67	61.20±10.86	57.60±15.82	0.75±0.07
test ktn SAM											44.46±2.99	45.22±6.94	42.17±7.75	45.99±3.86	44.89±3.17	45.06±2.96	45.06±5.32	43.10±3.84	47.03±5.74	0.62±0.02
test WT floral											48.11±8.21	46.22±10.10	47.72±11.51	50.39±7.13	48.39±8.17	49.03±8.15	49.12±11.75	51.21±10.37	46.76±5.87	0.67±0.09
test ktn floral											47.17±6.91	49.01±10.33	44.24±10.71	48.25±3.87	47.79±6.46	48.45±7.27	48.91±9.10	51.02±10.94	45.41±7.75	0.65±0.07

	train F1	train c0 F1	train c1 F1	train c2 F1	train Acc	train precision	train c0 precision	train c1 precision	train c2 precision	train Auc	val F1	val c0 F1	val c1 F1	val c2 F1	val Acc	val precision	val c0 precision	val c1 precision	val c2 precision	val Auc
split 0	60.84	64.14	65.99	52.38	60.91	60.79	63.08	65.94	53.35	0.78	48.03	45.25	57.32	41.53	48.69	48.58	51.73	50.65	43.35	0.68
split 1	58.04	59.69	64.49	48.93	58.21	58.02	61.11	61.53	51.42	0.76	60.07	67.31	68.67	44.22	60.96	59.92	64.81	64.38	50.58	0.81
split 2	60.48	65.46	67.10	48.87	60.99	60.43	63.59	63.69	54.01	0.79	51.23	55.04	52.44	46.22	51.27	51.39	52.57	55.57	46.04	0.71
split 3	57.54	58.98	62.56	51.09	57.53	57.56	59.91	62.02	50.75	0.75	56.77	63.29	66.98	40.05	57.86	56.63	59.53	63.23	47.11	0.79
split 4	58.49	61.48	64.42	49.55	58.70	58.40	61.05	62.22	51.93	0.77	60.58	64.96	66.95	50.23	60.99	60.55	63.50	63.43	54.72	0.77
split 5	58.88	61.72	64.52	50.41	59.07	58.82	61.82	62.05	52.60	0.77	59.34	64.37	65.32	48.32	59.80	59.30	62.85	61.91	53.14	0.77
train mean	59.04±1.22	61.91±2.28	64.85±1.42	50.37±1.13	59.24±1.30	59.00±1.20	61.76±1.26	62.91±1.51	52.34±1.11	0.77±0.01	56.00±4.75	59.97±7.60	62.95±5.96	45.10±3.58	56.59±4.85	56.06±4.54	59.17±5.21	59.86±5.03	49.15±4.01	0.75±0.04
test WT SAM	56.38±6.67	59.30±12.24	61.59±15.10	48.26±10.08	57.59±8.43	57.56±8.60	60.01±9.22	60.29±12.26	52.38±9.73	0.76±0.08	55.25±6.87	59.46±12.04	63.50±7.20	42.81±9.51	56.57±7.43	57.42±7.85	59.37±8.32	61.12±3.83	51.76±15.42	0.76±0.05
test ktn SAM											46.30±5.26	43.52±4.97	44.02±5.17	33.65±11.32	41.51±4.82	41.77±4.67	40.67±4.49	40.13±4.55	44.52±5.92	0.59±0.04
test WT floral											46.00±6.12	45.10±9.37	47.98±11.46	46.73±8.00	46.75±8.06	47.44±8.11	47.03±10.66	50.94±12.66	45.24±6.12	0.66±0.07
test ktn floral											49.88±5.83	49.71±10.13	51.83±10.18	48.10±3.81	50.04±6.74	50.32±6.84	50.96±4.83	54.41±10.29	45.59±3.71	0.67±0.07

B

topo

	train F1	train c0 F1	train c1 F1	train c2 F1	train Acc	train precision	train c0 precision	train c1 precision	train c2 precision	train Auc	val F1	val c0 F1	val c1 F1	val c2 F1	val Acc	val precision	val c0 precision	val c1 precision	val c2 precision	val Auc
split 0	61.68	64.86	67.47	52.70	62.02	61.71	62.29	65.28	57.54	0.79	51.17	55.35	59.84	38.32	52.30	51.71	51.90	54.79	48.44	0.73
split 1	60.28	63.41	66.50	50.94	60.67	60.30	61.42	63.57	55.92	0.78	61.87	68.37	65.97	51.27	61.92	61.89	69.82	63.85	52.01	0.89
split 2	64.63	68.64	70.61	54.62	64.95	64.68	66.09	68.84	58.82	0.81	54.35	58.22	56.28	49.57	54.38	54.49	55.75	59.20	48.53	0.73
split 3	59.59	63.44	65.08	50.24	59.92	59.56	61.06	63.17	54.44	0.78	59.49	64.72	68.42	45.33	60.37	59.81	58.66	67.28	53.90	0.80
split 4	59.83	64.03	65.89	49.58	60.27	59.82	61.63	63.22	54.61	0.78	61.56	65.44	69.89	49.34	62.66	62.61	61.89	63.46	62.47	0.80
split 5	60.10	64.23	65.93	50.12	60.54	60.13	62.47	62.46	55.48	0.78	61.60	62.66	65.81	56.34	61.55	61.81	65.86	64.97	54.58	0.78
train mean	61.02±1.75	64.77±1.80	66.91±1.80	51.37±1.76	61.40±1.72	61.02±1.73	62.49±1.68	64.42±2.15	56.13±1.57	0.79±0.01	58.34±1.12	62.46±4.43	64.37±4.79	48.20±5.52	58.86±4.01	58.72±4.14	60.65±5.02	62.26±4.12	53.26±4.72	0.77±0.03
test WT SAM	60.22±5.09	65.18±5.73	66.67±6.58	48.80±8.26	61.04±5.09	60.93±5.17	62.78±5.86	63.60±6.22	56.40±8.04	0.79±0.04	56.24±7.47	62.04±6.83	60.30±6.89	46.37±10.75	56.92±6.89	56.68±7.61	59.20±4.68	61.61±5.65	49.22±10.42	0.77±0.03
test ktn SAM											43.50±3.27	46.34±2.64	46.25±4.47	37.91±5.57	44.11±3.17	44.74±3.29	43.90±5.65	44.15±3.34	46.17±3.67	0.62±0.02
test WT floral											47.47±7.39	50.76±7.31	44.87±16.48	46.80±3.20	48.30±6.49	48.14±7.38	48.26±7.61	49.03±13.77	47.14±2.29	0.66±0.08
test ktn floral											47.33±5.58	50.80±6.94	45.10±6.57	46.10±3.01	47.67±4.48	48.29±4.61	48.33±6.52	52.02±7.08	44.52±2.40	0.67±0.05

bio

	train F1	train c0 F1	train c1 F1	train c2 F1	train Acc	train precision	train c0 precision	train c1 precision	train c2 precision	train Auc	val F1	val c0 F1	val c1 F1	val c2 F1	val Acc	val precision	val c0 precision	val c1 precision	val c2 precision	val Auc
split 0	48.28	56.31	18.45	70.07	53.34	53.81	44.91	49.31	67.21	0.75	45.43	56.47	11.80	68.02	51.87	52.61	43.70	47.13	67.01	0.73
split 1	42.27	2.64	55.04	69.12	50.88	46.20	30.57	42.48	65.57	0.70	42.86	3.70	56.55	68.31	51.55	70.05	100.00	42.74	67.42	0.70
split 2	53.02	47.57	42.49	69.00	53.62	52.72	48.49	45.02	64.66	0.74	52.68	45.81	41.56	70.68	53.70	52.41	46.45	46.33	64.46	0.75
split 3	42.33	1.47	55.74	69.77	51.41	48.09	34.95	42.70	66.63	0.69	40.33	0.00	55.09	65.90	49.45	35.38	0.00	41.68	64.45	0.67
split 4	43.25	6.07	54.86	68.83	51.01	49.63	41.00	42.64	65.24	0.70	42.73	0.00	56.37	72.12	52.36	37.00	0.00	43.12	67.98	0.72
split 5	42.67	3.24	55.59	69.19	51.34	49.42	40.12	43.04	65.10	0.71	41.92	3.57	54.46	67.72	50.51	59.44	72.76	42.11	63.45	0.70
train mean	45.30±0.03	19.55±23.08	47.03±13.61	69.33±4.44	51.93±1.11	49.98±2.60	40.01±5.94	44.20±2.44	65.73±0.89	0.72±0.02	44.32±0.03	18.26±23.50	45.92±16.11	68.79±2.04	51.59±1.32	51.15±12.10	43.82±36.15	43.85±2.10	65.78±1.71	0.71±0.02
test WT SAM	42.05±2.55	1.52±3.50		55.59±2.81	69.04±4.94	51.10±2.79	42.81±12.15	19.03±36.58	42.55±2.27	0.70±0.03	41.31±2.68	53.50±4.95	55.63±2.19	67.79±5.68	50.78±2.95	39.30±8.67	11.70±22.12	42.99±1.80	63.21±5.05	0.68±0.05
test ktn SAM											31.23±8.82	3.89±7.36	46.26±6.13	43.54±13.09	37.99±8.41	31.01±4.13	16.25±26.67	36.03±5.23	40.74±12.15	0.54±0.10
test WT floral											43.76±1.21	0.84±0.81	56.60±1.64	73.84±2.22	52.97±1.31	45.82±11.24	21.75±31.29	42.98±1.39	72.75±5.69	0.72±0.01
test ktn floral											44.91±9.89	1.81±2.44	58.31±0.92	74.59±1.65	54.34±0.84	48.19±11.24	28.44±31.59	44.79±0.80	71.34±3.73	0.72±0.01

topoAndBio

	train F1	train c0 F1	train c1 F1	train c2 F1	train Acc	train precision	train c0 precision	train c1 precision	train c2 precision	train Auc	val F1	val c0 F1	val c1 F1	val c2 F1	val Auc	val precision	val c0 precision	val c1 precision	val c2 precision	val Auc
split 0	73.10	72.18	72.24	74.88	73.08	73.12	71.54	72.28	75.52	0.88	64.07	60.45	64.05	67.70	64.04	64.12	60.01	63.32	69.05	0.81
split 1	70.33	68.57	68.90	73.51	70.33	68.92	68.57	73.51	0.86	70.92	72.49	71.55	68.72	70.92	70.94	73.29	70.21	69.32	68.89	0.88
split 2	71.64	70.53	71.79	72.61	71.64	71.65	70.71	71.00	73.26	0.87	66.07	64.47	64.84	68.90	66.03	66.16	63.48	64.19	70.80	0.83
split 3	71.08	69.23	70.05	73.97	71.08	71.10	69.98	69.18	74.13	0.86	70.03	70.84	70.16	69.08	70.06	70.35	66.38	71.90	72.78	0.87
split 4	69.63	67.89	68.71	72.30	69.64	69.64	68.55	68.05	72.32	0.86	71.57	68.37	70.73	75.61	71.64	71.69	72.15	68.58	74.34	0.88
split 5	71.27	69.48	70.44	73.74	71.27	71.27	69.72	70.45	73.64	0.87	69.30	66.82	68.73	72.34	69.37	69.38	68.86	68.71	69.57	0.85
train SAM	71.181.08	69.671.13	70.351.32	73.501.06	71.181.08	69.901.02	69.921.47	73.730.97	0.8740.10	66.662.70	67.242.39	68.342.89	70.392.74	68.682.73	68.772.72	67.534.73	67.823.09	70.081.06	68.503.03	0.8503.03
test W/T test	72.5845.50	71.326.24	72.012.37	74.095.49	72.160.51	71.821.05	71.385.93	71.697.03	75.407.37	0.8840.04	66.274.24	67.782.89	65.444.85	65.599.55	66.682.34	66.953.83	65.394.50	67.433.42	68.084.36	68.502.02
test kin SAM											42.476.30	43.386.16	45.094.58	38.955.27	42.994.75	45.187.22	41.315.74	42.494.44	43.9514.43	0.8103.03
test W/T test kin							12.1745.83	88.635.24	93.514.12	93.950.83	62.704.37	62.584.41	62.584.41	62.584.41	62.584.41	62.584.41	62.584.41	62.584.41	62.584.41	62.584.41
test kin fMRIal							62.584.41	98.635.81	55.547.78	72.591.87	62.924.76	63.305.23	58.742.28	62.404.79	62.404.79	62.404.79	62.404.79	62.404.79	62.404.79	62.404.79

unweighted

	train F1	train c0 F1	train c1 F1	train c2 F1	train Acc	train precision	train c0 precision	train c1 precision	train c2 precision	train Auc	val F1	val c0 F1	val c1 F1	val c2 F1	val Acc	val precision	val c0 precision	val c1 precision	val c2 precision	val Auc
split 0	57.50	62.75	64.30	45.43	58.19	57.57	59.52	60.99	52.21	0.76	51.22	54.82	57.49	41.37	51.70	51.26	54.83	53.18	45.78	0.70
split 1	55.18	60.08	61.94	43.53	55.85	55.24	57.73	58.09	49.90	0.74	61.80	69.88	67.79	47.72	62.50	61.77	68.15	63.32	53.84	0.80
split 2	56.87	63.94	63.57	43.10	57.78	56.97	60.39	59.69	50.83	0.76	52.97	56.53	57.81	44.58	53.30	53.01	53.41	56.81	48.82	0.72
split 3	55.91	60.29	62.33	45.10	56.45	55.93	58.16	58.90	50.73	0.75	54.60	64.25	66.10	33.46	56.33	54.90	58.21	60.82	45.66	0.77
split 4	55.43	60.58	62.21	43.51	56.18	55.57	58.19	57.90	50.64	0.74	57.12	62.12	66.70	42.53	58.32	57.64	57.06	62.56	53.29	0.78
split 5	55.40	60.82	62.40	42.97	56.20	55.53	58.13	58.15	50.32	0.75	56.10	62.77	62.87	42.66	56.99	56.26	60.48	57.92	50.38	0.76
train mean	56.05±0.85	61.41±1.43	62.79±0.85	43.94±0.97	56.77±0.88	56.14±0.84	58.69±0.94	58.95±1.09	50.77±0.72	0.75±0.01	55.64±3.37	61.73±4.98	63.13±4.15	42.05±4.35	56.61±3.49	55.81±3.38	58.69±4.80	59.10±3.52	49.63±3.24	0.75±0.04
test WT SAM	55.72±5.03	62.24±7.94	63.63±7.76	41.29±7.27	57.13±5.58	56.70±5.38	59.00±6.68	58.67±6.91	52.44±8.01	0.76±0.06	55.44±4.86	62.80±6.12	64.01±5.59	39.52±10.60	57.08±5.02	56.58±5.58	59.56±4.18	59.03±5.86	51.16±10.51	0.77±0.04
test kn SAM											41.23±4.11	46.55±4.59	44.43±3.81	32.70±7.63	42.49±3.93	44.14±4.22	42.92±4.59	40.03±3.59	49.47±6.31	0.60±0.03
test WT floral											48.89±7.33	50.89±8.20	49.38±10.92	46.41±5.60	49.06±7.33	49.10±7.18	49.24±9.00	49.86±8.94	48.41±6.13	0.67±0.08
test kn floral											49.54±4.33	51.99±6.35	51.05±6.97	45.58±2.41	49.70±4.43	49.66±4.33	50.95±6.14	50.98±6.39	47.05±1.92	0.68±0.05

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.
- Eng, R.C., Schneider, R., Matz, T.W., Carter, R., Ehrhardt, D.W., Jönsson, H., Nikoloski, Z., Sampathkumar, A., 2021. KATANIN and CLASP function at different spatial scales to mediate microtubule response to mechanical stress in Arabidopsis cotyledons. *Current biology : CB*. <https://doi.org/10.1016/j.cub.2021.05.019>.
- Gibson, W.T., Veldhuis, J.H., Rubinstein, B., Cartwright, H.N., Perrimon, N., Brodland, G.W., Nagpal, R., Gibson, M.C., 2011. Control of the mitotic cleavage plane by local epithelial topology. *Cell* 144, 427–438.
- Hartig, K., Beck, E., 2006. Crosstalk between auxin, cytokinins, and sugars in the plant cell cycle. *Plant biology (Stuttgart, Germany)* 8, 389–396.
- Jackson, M.D.B., Duran-Nebreda, S., Kierzkowski, D., Strauss, S., Xu, H., Landrein, B., Hamant, O., Smith, R.S., Johnston, I.G., Bassel, G.W., 2019. Global Topological Order Emerges through Local Mechanical Control of Cell Divisions in the Arabidopsis Shoot Apical Meristem. *Cell systems* 8, 53–65.e3.
- Jones, A.R., Forero-Vargas, M., Withers, S.P., Smith, R.S., Traas, J., Dewitte, W., Murray, J.A.H., 2017. Cell-size dependent progression of the cell cycle creates homeostasis and flexibility of plant cell size. *Nature communications* 8, 15060.
- Katz, L., 1953. A new status index derived from sociometric analysis. *Psychometrika* 18, 39–43.
- Kitagawa, M., Jackson, D., 2017. Plasmodesmata-Mediated Cell-to-Cell Communication in the Shoot Apical Meristem: How Stem Cells Talk. *Plants (Basel, Switzerland)* 6.
- Komis, G., Luptovčiak, I., Ovečka, M., Samakovli, D., Šamajová, O., Šamaj, J., 2017. Katanin Effects on Dynamics of Cortical Microtubules and Mitotic Arrays in Arabidopsis thaliana Revealed by Advanced Live-Cell Imaging. *Frontiers in plant science* 8, 866.
- Li, Y., Liu, D., López-Paz, C., Olson, B.J., Umen, J.G., 2016. A new class of cyclin dependent kinase in Chlamydomonas is required for coupling cell size to cell division. *eLife* 5, e10767.

- Louveaux, M., Julien, J.-D., Mirabet, V., Boudaoud, A., Hamant, O., 2016. Cell division plane orientation based on tensile stress in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* 113, E4294-303.
- Marchiori, M., Latora, V., 2000. Harmony in the small-world. *Physica A: Statistical Mechanics and its Applications* 285, 539–546.
- Nowak, J., Eng, R.C., Matz, T., Waack, M., Persson, S., Sampathkumar, A., Nikoloski, Z., 2021. A network-based framework for shape analysis enables accurate characterization of leaf epidermal cells. *Nature communications* 12, 458.
- Pisner, D.A., Schnyer, D.M., 2020. Support vector machine. In: *Machine Learning*. Elsevier, 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.
- Wang, Y., Sampathkumar, A., 2020. Live Cell Imaging of Microtubule Cytoskeleton and Micromechanical Manipulation of the *Arabidopsis* Shoot Apical Meristem. *Journal of visualized experiments : JoVE*.
- Wangenheim, D. von, Fangerau, J., Schmitz, A., Smith, R.S., Leitte, H., Stelzer, E.H.K., Maizel, A., 2016. Rules and Self-Organizing Properties of Post-embryonic Plant Organ Cell Division Patterns. *Current biology : CB* 26, 439–449.
- Willis, L., Refahi, Y., Wightman, R., Landrein, B., Teles, J., Huang, K.C., Meyerowitz, E.M., Jönsson, H., 2016. Cell size and growth regulation in the *Arabidopsis thaliana* apical stem cell niche. *Proceedings of the National Academy of Sciences of the United States of America* 113, E8238-E8246.