ORIGINAL ARTICLE

Cytometry PART A
Journal of the International Society for Advancement of Cytometry

A Method for Characterizing Phenotypic Changes in Highly Variable Cell Populations and its Application to High Content Screening of *Arabidopsis thaliana* Protoplasts

Gregory R. Johnson,[1] Joshua D. Kangas,[1] Alexander Dovzhenko,[2] Rüdiger Trojok,[3] Karsten Voigt,[2] Timothy D. Majarian,[1] Klaus Palme,[2,4] Robert F. Murphy[1,4,5]*

[1]Computational Biology Department, Carnegie Mellon University, Pittsburgh

[2]Institute for Biology II/Molecular Plant Physiology, Faculty of Biology, Albert Ludwig University of Freiburg, Freiburg, Germany

[3]Centre for Biological Systems Analysis (ZBSA), Albert Ludwig University of Freiburg, Freiburg, Germany

[4]Freiburg Institute for Advanced Studies (FRIAS), Albert Ludwig University of Freiburg, Freiburg, Germany

[5]Departments of Biological Sciences, Biomedical Engineering and Machine Learning, Carnegie Mellon University, Pittsburgh

Received 21 July 2016; Revised 22 December 2016; Accepted 19 January 2017

Grant sponsor: U.S. National Institutes of Health; Grant numbers: GM103712 and EB009403; Grant sponsor: Excellence Initiative of the German Federal and State Governments through the Freiburg Institute for Advanced Studies and the Freiburg Centre for Biological Signalling Studies; Grant number: EXC 294, ZUK 43; Grant sponsor: Research Award to RFM from the Alexander von Humboldt Foundation; Grant Sponsor: German Research Foundation; Grant number: SFB746; Grant sponsor: German Federal Ministry of Education and Research; Grant numbers: 0315329B and 0316185B.

• **Abstract**

Quantitative image analysis procedures are necessary for the automated discovery of effects of drug treatment in large collections of fluorescent micrographs. When compared to their mammalian counterparts, the effects of drug conditions on protein localization in plant species are poorly understood and underexplored. To investigate this relationship, we generated a large collection of images of single plant cells after various drug treatments. For this, protoplasts were isolated from six transgenic lines of *A. thaliana* expressing fluorescently tagged proteins. Eight drugs at three concentrations were applied to protoplast cultures followed by automated image acquisition. For image analysis, we developed a cell segmentation protocol for detecting drug effects using a Hough transform-based region of interest detector and a novel cross-channel texture feature descriptor. In order to determine treatment effects, we summarized differences between treated and untreated experiments with an L_1 Cramér-von Mises statistic. The distribution of these statistics across all pairs of treated and untreated replicates was compared to the variation within control replicates to determine the statistical significance of observed effects. Using this pipeline, we report the dose dependent drug effects in the first high-content *Arabidopsis thaliana* drug screen of its kind. These results can function as a baseline for comparison to other protein organization modeling approaches in plant cells.

• **Key terms**

high content screening; cellular heterogeneity; subcellular location; fluorescence microscopy

Introduction

Given the complexity of biological systems it has become well understood that automated approaches to biological discovery are needed to reduce the time, effort and cost of basic research (1,2). Many high-content screens for drug effects have been performed, although many have suffered from irreproducibility, raising questions of how to deal with variation among experimental replicates, and how to compare the results of screens carried out with different equipment and in different laboratories (3). Heterogeneity in individual cellular responses and across experiments is particularly confounding in such studies. Although several approaches to addressing heterogeneous cellular responses in high-content screens have been described (4), few have been specifically designed to directly model replicate variation (which can result from experimental noise, image-field-to-field or even cell-to-cell heterogeneity), other than to show that a measured trend holds across pairs of experimental replicates.

*Correspondence to: Robert F. Murphy; Computational Biology Department, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213. E-mail: murphy@cmu.edu

Published online 28 February 2017 in Wiley Online Library (wileyonlinelibrary.com)

DOI: 10.1002/cyto.a.23067

Arabidopsis thaliana has become a popular plant model (5) and system to study protein gene expression heterogeneity (6). While plant-based protein and gene studies have been gaining popularity for agricultural and basic science applications, comparatively little research toward high-content, cell-based assays for plants has been described. This is in contrast to the widespread use of mammalian cell-based assays used for drug development and the number of mammalian cell-based projects that have been undertaken, such as the Human Protein Atlas (7), Mitocheck (8), and CellMorph (9).

High-content screens typically focus on measuring a particular cellular activity, often using a combination of image features and supervised machine learning to learn to distinguish positive and negative controls (10,11). When positive controls are not available (e.g., for discovering new phenotypes), unsupervised or semi-supervised machine learning can be used (9,11,12). However, a classification-accuracy based approach via comparison of control and treated experiments (in absence of a positive control) can be preferable as it is readily interpretable as a difference between two sample populations (13). One of the major caveats of this approach is that it may discard heterogeneous population information, summarizing the relationship between two populations as simply an error statistic. Furthermore, it does not detect the magnitude of differences (measured by accuracy). In other words, as phenotypes expressed in experiments become more different, eventually a classifier may become 100% accurate. At this point, the magnitude of effect sizes can no longer be distinguished between alternate conditions.

Here we present an automated protocol for screening for drug effects on protein subcellular location in Arabidopsis protoplasts that simultaneously considers both cellular and experimental variation. It improves upon previous methods by providing an unbounded indicator of magnitude-of-effect by considering the distribution of differences between experimental replicates, rather than the proportion of cells correctly classified. Instead of implementing separate procedures for evaluating the quality of control or experimental replicates, we base our method on the assumption that an effect is observed if the difference between control and treated groups is greater than the difference between control groups alone given all sources of error being equal. Our results demonstrate the ability to detect drug effects in highly heterogeneous populations and with differing numbers of experimental replicates between cohorts. The combination of experimental protocol and analytical approach we have developed is expected to be useful for investigating cellular and subcellular phenotypes in plants.

Results

High-Throughput Imaging of Protoplasts

In order to assess small molecule effects on plant protein organization we used protoplasts from aseptically grown shoots of *Arabidopsis thaliana*. Protoplasts were isolated from above ground tissues from one of six plant lines, each expressing a fluorescently-tagged protein. The tagged proteins were chosen to provide a diverse set of subcellular patterns. A set of widefield images for each cell line was generated for treatment with one of eight drugs, each at three concentrations (0.1, 0.5, 2.0 μM). The drugs are listed in Table 1; they were chosen since they are frequently used in plant studies and their full effects have not been extensively characterized. Dimethyl sulfoxide (DMSO), in which compounds were dissolved, was used as a control. This set of compounds and proteins were expected to yield interesting phenotypes that would illustrate the power of the method and could challenge existing methods. At least two replicates were performed for each combination of drug, concentration, and plant line, with as many as 16 replicates being performed and 4 replicates being the mode. There were between 52 and 62 replicates performed for control conditions, with 1200 wells total. A small fraction of the images generated in this process were removed as a result of camera error yielding a total of 18,671 fields.

Analysis Pipeline Overview

We constructed a pipeline to rank drug effects on protein location patterns from cell populations in high-throughput 2D wide-field images. The pipeline (Fig. 1) has four steps; the first is to detect regions-of-interest (ROI) in each image. Due to the fact that plant protoplasts take a spherical shape after removal of the cell wall, we used a circle detection method. The second step consists of texture feature calculation for each region of interest (ROI). These features are designed to capture the spatial relationship between pixel intensities within and across protein patterns. In the third step, given the ROI features, all pairs of untreated (control) and treated experiments are compared to determine the magnitude of phenotypic variation as a result of drug addition, and the variation of differences expected between untreated and treated experimental pairs. Lastly, given the results of the pairwise tests, we determine if the measured differences are greater than what would be expected between pairs of control replicates to identify statically significant drug effects. This pipeline was applied to all protein–drug combinations as described in Methods.

Region of Interest Detection and Associated Features

For each image, we found regions of interest corresponding to individual protoplasts by detecting circular regions of appropriate sizes (see "Methods"). Examples of the patch-

Table 1. Drugs used in this study

DRUG NAME	ABBREVIATION	KNOWN ACTIVITY (PUBCHEM)
Brefeldin-A	BFA	TDP1, Arf-GDP-ARNO, delta17 Arf1, and TGF-b inhibitor. Antiproliferative activity against human UACC62, MCF7, NCI60, UACC-62, HCT116 and other cell lines.
Benzylphosphonic acid	BPA	No activity results found
Damnacanthal	DMN	p56lck autophosphorylation, ALDH1A1, Histone Lysine Methyltransferase G9a, VDR, Tau fibril formation inhibitor. Cytotoxic to MCF8, H460, DU145 cells.
Endothall	Endothall	Agonist to ROR-gamma, ARE, AR, ROR-gamma, RAR, AhR signaling pathways.
N9-isopropyl olomoucine	N9	cdc2 p34/Cyclin B, Cyclin-dependent kinase 1/2, and STK33 inhibitor
Oryzalin	Oryzalin	*Toxoplasma gondii* and *Cryptosporidium parvum* antimicrobial. Cytotoxic to DT40 cells. Disruptive to cell membrane potential. Ebola virus entry blocker. Activator of PXR signaling pathway.
Tyrphostin	TRP	*Plasmodium falciparum* antimicrobial. PYRK, Aldose reductase, Janus kinase 2 mutant, GLS, CDK2/CycE, Schistosoma Mansoni Peroxiredoxin and JMJD2E inhibitor.
ZM-449829	ZM	Human Muscle isoform 2 Pyruvate Kinase Activator. ALDH1A1, ROR gamma transcriptional activity, Histone Lysine Methyltransferase G9a, VDR, 15-hLO-2, TDP1, Cruzain, ERK signaling pathway inhibitor.

based region of interest detection are shown in Figure 2. Although the segmentation method is not perfect, it has a low false positive rate, detecting few patches with no or partial protoplast content. Furthermore, given the use of the texture features described below, perfect segmentation is not necessary.

For each region of interest, we compute a co-occurrence matrix of pixel intensities at a range of specified pixel offsets and channel pairs, and calculate 12 Haralick texture features (14) on each co-occurrence matrix (see "Methods"). We categorize the features in three groups, protein texture features, protein–chloroplast features, and chloroplast features, depending on which fluorescence channel or channel pairs the features were computed on.

Drug–Effect Detection

To accurately evaluate the effect of drugs, it is important to consider differences not only between control and drug experiments, but to also consider the variation among control experiments themselves. This was especially relevant for our protoplast images due to the extensive variation in cell morphology, the presence of damaged cells, and the relatively low number of cells per field during imaging at high magnification. In other words, to confidently say that there is an effect of a drug condition, the variation between control and condition experiments must be significantly greater than the variation between replicates of identical controls. Additional challenges arise when comparing different numbers of experiments, images, and regions of interest. With this in mind, we constructed a procedure to simultaneously consider these differences in the evaluation of drug-effect detection.

Given segmented regions of interest and their associated descriptors, we can determine the variation between experiments via classification-based methods. Here we introduce a measure that is the difference between the cumulative distribution functions with respect to the distance from a separating hyperplane that we call *deltaF* (see "Methods"). This statistic is an L_1 version of the Cramér-von Mises statistic (15), and has advantages over similar previously described methods (13), by not only describing the difference between two conditions, but also not being constrained with an upper bound. It provides additional information on how different two images are even if classification accuracy is 100%, which is useful for the comparison of large effects. The underlying intuition is that the ability to predict the experiment from which a cell or region of interest arose should be difficult between very similar conditions but will be very easy when there is a clear change in phenotype. To reduce classification bias due to different sized populations (i.e., a control replicate having more regions of interest than an experimental replicate) the data from each experiment were weighted equally. We compared all pairs of untreated (control) and treated experiments for a given cell line and specific condition to construct a distribution of observed effect sizes by measuring the distances to a support vector machine (SVM) decision boundary via five-fold cross-validation, and comparing the difference between control and treated distances (the *deltaF* measure). An additional *null distribution* was constructed by measuring the *deltaF* between all pairs of control experiments against each other. This process was repeated for each experimental condition (drug, protein, and concentration combination). The resulting histograms of *deltaF* responses for the images are shown in Figure 3 using the complete set of texture features. The *null* distribution (control–control experimental comparisons) is shown in red, and represents the maximum variation we would expect across these experiments.

By comparing the treated and untreated dose response distributions we can determine if there exists a statistical difference between conditions and thus identify whether a drug has had an effect. Here we compared dose response distributions via a single-sided, two-sample Kolmogorov–Smirnov Test (16) versus the distribution of control samples from all

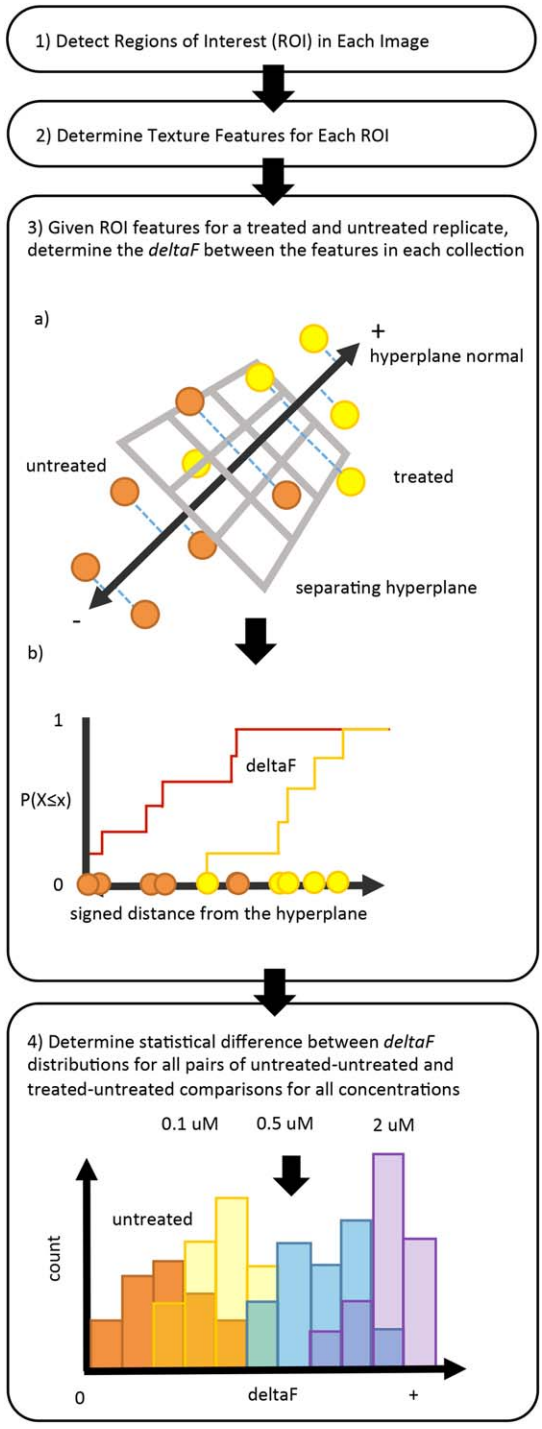

Figure 1. Pipeline overview. Four steps were performed: (1) detection of spherical regions of interest, (2) computation of intra and inter-channel texture features, (3) determination of a measure of discrimination, *deltaF*, between sets of replicates (e.g., between untreated and untreated or between untreated and treated with a given drug concentration), and (4) determination of the area of overlap between the CDFs for different comparisons. The third step consists of two parts: (a) Fitting a SVM to separate the two replicates, and project every point on to the hyperplane normal, and (b) Determining the CDF with respect to signed distance from the hyperplane. A condition was considered to have an effect if the area of overlap in step 4 was statistically significant. [Color figure can be viewed at wileyonlinelibrary.com]

experiments across all lines. The results of the test are adjusted via Bonferroni correction (17).

We ran the pipeline for different sets of texture features to determine which channels provided the most information about drug-induced phenotypes. The results are shown in Figure 4 for protein texture features (green bars), protein and protein–chloroplast features (orange bars), as well as chloroplast features alone (light blue bars); an asterisk indicates that the result is statistically different from control at an $\alpha < 0.001$. To ensure we have both sufficient numbers of regions of interest and experimental replicates to avoid detecting spurious relationships, we constructed a random-relationship baseline with the complete feature set by evaluating a "sham" experiment. Each ROI in a replicate was randomly assigned a "treated" or "untreated" label, and the *deltaF* was determined. This process was repeated for each experimental condition, each untreated and treated group was compared, and a significance level determined (dark blue bars).

Drug Responses

As to be expected, the tagged protein features alone were quite similar in effects detected (Fig. 4, green bars) compared to the complete feature set (yellow bars), but of diminished effect size compared to the results with all features in many cases. Despite this fact, the protein features alone were sufficient to determine drug effect in virtually all cases in which a drug effect was detected. The addition of chloroplasts as a fiduciary marker (cross channel protein-chloroplast features, orange bars), increases the detected effect virtually across the board. We believe that this may be an example of the *curse of dimensionality*, where spurious correlations are detected between training sets when training the SVM. Interestingly some effects can be detected with chloroplast features alone (light blue bars), but those effects are not consistent across all lines. This suggests that chloroplast behavior may have been affected by the creation of the different tagged lines.

Comparison with Current Methods

The method we have described was designed to handle the dramatic variation both within samples and between control replicates. It is naturally of interest to compare our results with those that might have been obtained with methods commonly in use. For this, we chose two simple, illustrative approaches. The first uses a single image feature, the total protein intensity, and asks whether the distributions of that feature among cells differ between control and experimental conditions. The second defines the unperturbed phenotype of each cell line by training a classifier to distinguish the six cell lines and considers a drug to cause a perturbation for a particular cell line if the most frequent cell classification in the presence of the drug is different from the most frequent cell classification in the unperturbed phenotype of that cell line. Table 2 shows the results of these two approaches compared to the results shown in Figure 4. The results indicate that "simple" methods identify many more conditions as being perturbed than our method. Our

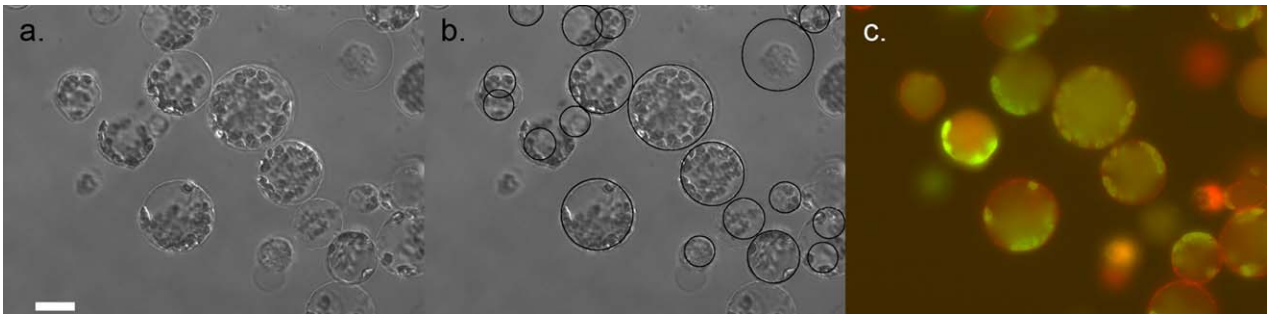

Figure 2. Example images and patch based segmentation. (A) Example 2D, widefield, DIC image. (B) 2D DIC image with detected circular regions of interest. (C) Fluorescence channels corresponding 2D image, with chloroplast auto fluorescence in green and NCRK-313-GFP in red. Bar, 20 μm. [Color figure can be viewed at wileyonlinelibrary.com]

method finds only 36 conditions to be perturbed, while the other methods find 110 or 80 (out of 134) to be perturbed. Since we consider it unlikely for well over half of the combinations of drugs and targets to be perturbed, we interpret these results as support for the more robust nature of our new approach.

Examples of Effects Observed

Given the results in Figure 4 we see that detected effects generally scale with dose size, where Tyrphostin and Damnacanthal have the strongest effect across all dose concentrations. Tyrphostin is an inhibitor of tyrosine kinase and clathrin-coated endocytosis (18). Damnacanthal has similarly

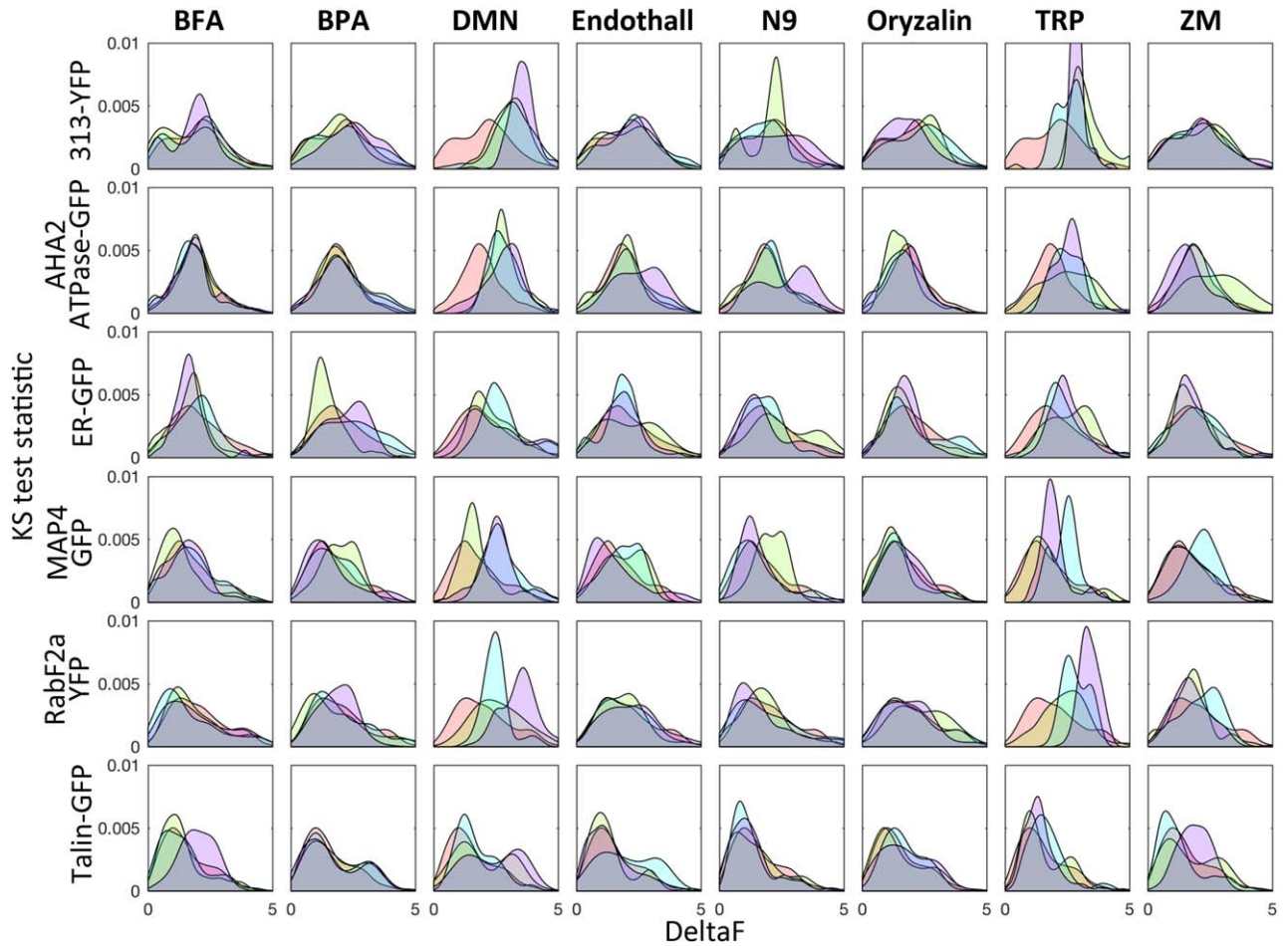

Figure 3. Normalized dose response histograms for experiments using all features. Red indicates the distribution of *deltaF* measures across control experiments, yellow, blue and purple distributions indicate 0.1, 0.5, and 2.0 μM drug treatment, respectively. Note for example the differences between the histograms for DMN and TRP. [Color figure can be viewed at wileyonlinelibrary.com]

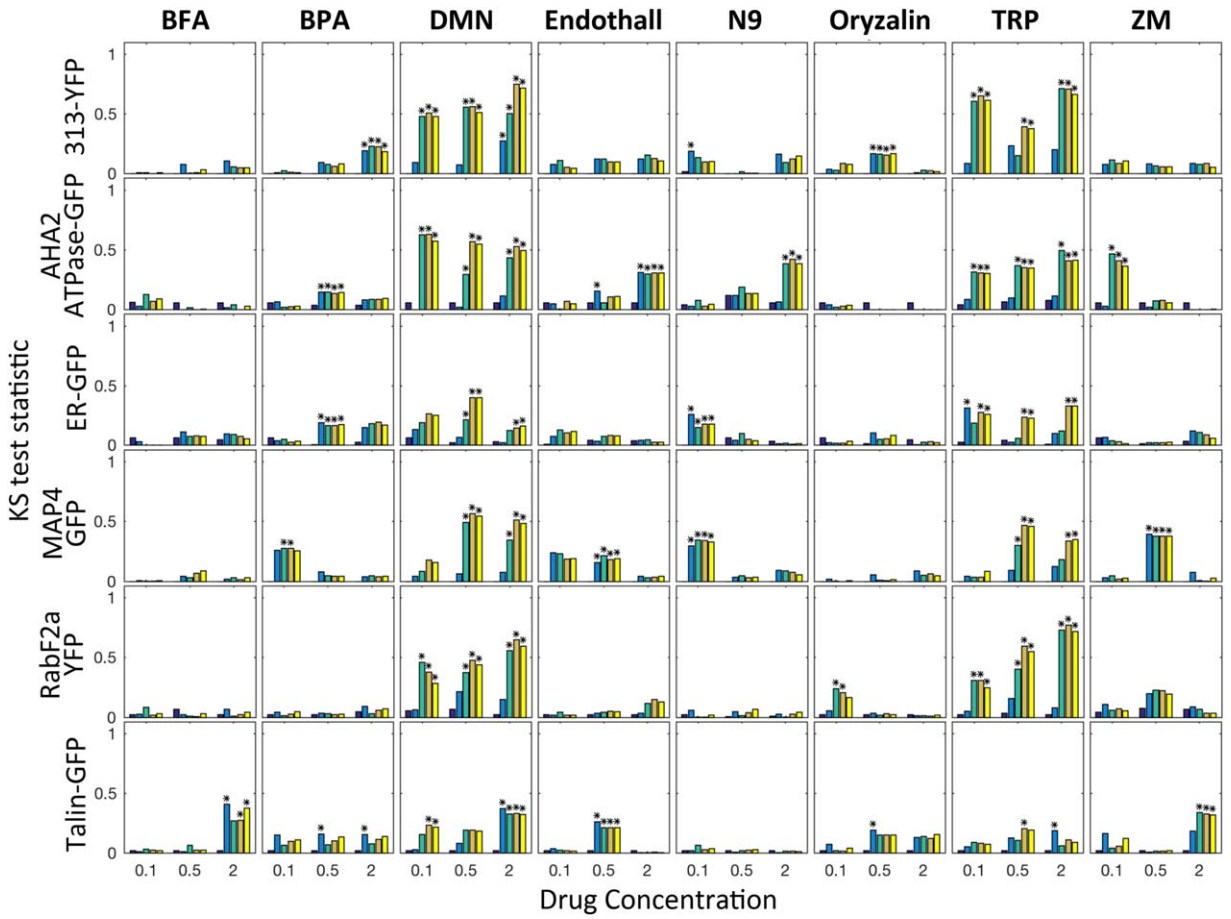

Figure 4. Analysis of statistic values for differences between control and experiment dose response histograms for increasing dosage using all feature sets on 2D images. Bars represent sham comparison (dark blue), chloroplast features (light blue), protein features only (green), protein and protein–chloroplast features (orange), and chloroplast, protein, and protein–chloroplast features (yellow). Asterisk indicates statistical difference from control at α=0.001 Bonferroni-corrected across all comparisons, 720 in total. [Color figure can be viewed at wileyonlinelibrary.com]

been reported to be a tyrosine kinase inhibitor. Our results suggest that both of these drugs result in large reorganizations of protein localization patterns across many proteins. We therefore asked whether they produced similar effects on those proteins by directly comparing the samples treated with each using the same method used to compare them to their respective controls. As shown in Figure 5, the effects on each protein were statistically different, with the exception of talin-GFP. This is consistent with the weaker effect of both drugs upon talin-GFP in comparison with controls.

Relative to other proteins, the results indicate a strong sensitivity of AHA2–GFP localization to N9-isopropyl olomoucine at high doses. There may also be some dose-specific effects with treatment of NCRK-313-GFP with ZM and BPA. We also present typical images for the highest dose of the three drug–target combinations for which the confidences of a drug-induced difference are highest (Fig. 6). Each row of images shows the untreated and treated condition for a given drug–target combination. Although these images represent typical treated and untreated images (and are a subset of the images for each replicate), some differences may be more clear than others. With NCRK-313-YFP treated with Damnacanthal, we see a clear increase of NCRK-313-YFP fluorescence, while the difference detected for AHA2-GFP treated with Damnacanthal does not appear to be detectable by eye [computational analysis has been shown to detect real but subtle differences in subcellular patterns that are not visually distinguishable (19)]. For RabF2a-YFP, the effect of Tyrphostin

Table 2. Comparison of results from the method described in this article and simpler methods

	INTENSITY T TEST		RANDOM FOREST	
FIGURE 4 PF	UNPERTURBED	PERTURBED	UNPERTURBED	PERTURBED
Unperturbed	34	74	59	49
Perturbed	0	36	5	31

The values show the number of conditions considered to be significantly perturbed or unperturbed in the results shown in Figure 4 (using protein features) that were also considered to be perturbed or unperturbed by either of the simpler methods described in the text. Note the large number of conditions considered to be perturbed by either of those methods that are not considered to be perturbed by our new method.

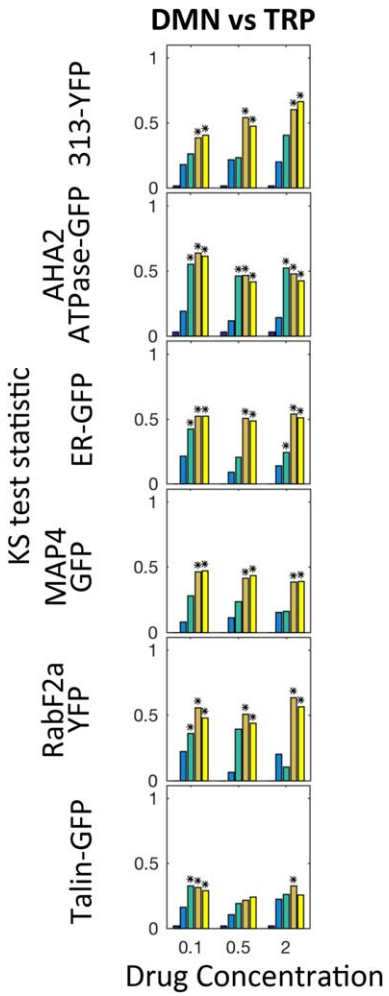

Figure 5. Comparison of effects of Damnacanthal and Tyrphostin. The dose response histograms for Damnacanthal and Tyrphostin were directly compared using the same method as in Figure 4. Bars represent sham comparison (dark blue), chloroplast features (light blue), protein features only (green), protein and protein–chloroplast features (orange), and chloroplast, protein, and protein–chloroplast features (yellow). Asterisk indicates statistical difference from control at α=0.001 Bonferroni-corrected across all comparisons, 90 in total. [Color figure can be viewed at wileyonlinelibrary.com]

appears to be to shift it from a diffuse to a punctate pattern. Based on our results, the developed method provides a potent platform for future pharmacological or pharmacogenetic studies in single plant cell models.

Discussion

Here we presented the results of the first high-content drug screen for *Arabidopsis thaliana* shoot protoplasts. We developed a novel drug–effect detection method to compare the variation in protein patterns between untreated and drug treated replicates to control replicate variation and used it to determine the statistical significance of drug effects based on differences in protein pattern. This method improves upon previous classification-based methods by providing an unbounded measure of difference between control and treated experiments. Compared to previous effect–detection screening methods, we do not use a separate procedure for evaluating the quality of control experiments, but rather use the same comparison to measure differences between treated and control replicates and intercontrol replicates themselves. This intercontrol comparison becomes the baseline for evaluation, given the assumption that any observed effect should be greater than the greatest difference between pairs of control replicates. Our method allows for flexibility with varying numbers of replicates, whereas other methods may require the same number of replicates across all experiments. Similar to other classification-based approaches, we use a linear boundary as the discriminating element, but rather than consider the fraction of correctly classified regions of interest, we use the distance to the hyperplane as a maximum discrimination descriptor, and determine the differences between distributions via the L_1 Cramér-von Mises statistic (20). It is worth noting that our method does not necessarily provide better detection of subtle or complex phenotypes, but does permit better detection of different phenotypes in the presence of significant within and between sample variation. By using this method in conjunction with protein-specific and protein–chloroplast features, we were able to detect dosage specific drug responses in a range of conditions. The methods and results presented here provide a baseline for comparing effects to those of generative models, and other cell modeling methods.

Methods

Experiments

Protoplasts were isolated from six *Arabidopsis thaliana* lines expressing fluorescently tagged proteins as previously described (21). The following lines were used: AHA2-GFP (22), ER-GFP (23), talin-GFP (24), NCRK-313-YFP (25), RabF2a-YFP (26), or GFP-MAP4 (27). Three concentrations (2.0, 0.5, 0.1 μM) of eight drugs (benzylphosphonic acid, Brefeldin-A, Damnacanthal, endothall, N9-isopropyl olomoucine, oryzalin, tyrphostin, ZM-449829) were used. Information about each drug is provided in Table 1. The drugs were added to protoplast samples to yield a total of 150 unique experimental conditions including untreated controls. The robotic addition of compounds to each experimental well was synchronized with the timing of imaging on the microscope such that drug exposure times for all experiments were approximately equal (4 h) across all wells in the multiwell plate. 2D image series were acquired using an Axio Observer.Z1 microscope (Carl Zeiss, Germany) with a 40× magnification air objective with a numerical aperture of 0.95 (Carl Zeiss, Germany), and a pixel size of 0.161 × 0.161 μm, with a total image size of 1388 × 1040 with 16 fields per well. Each image was taken at a constant distance from the bottom of the 96-well plate (Ibidi, Germany). DIC, fluorescent-tagged protein (470 nm), and chloroplast autofluorescence (558 nm) channels were taken.

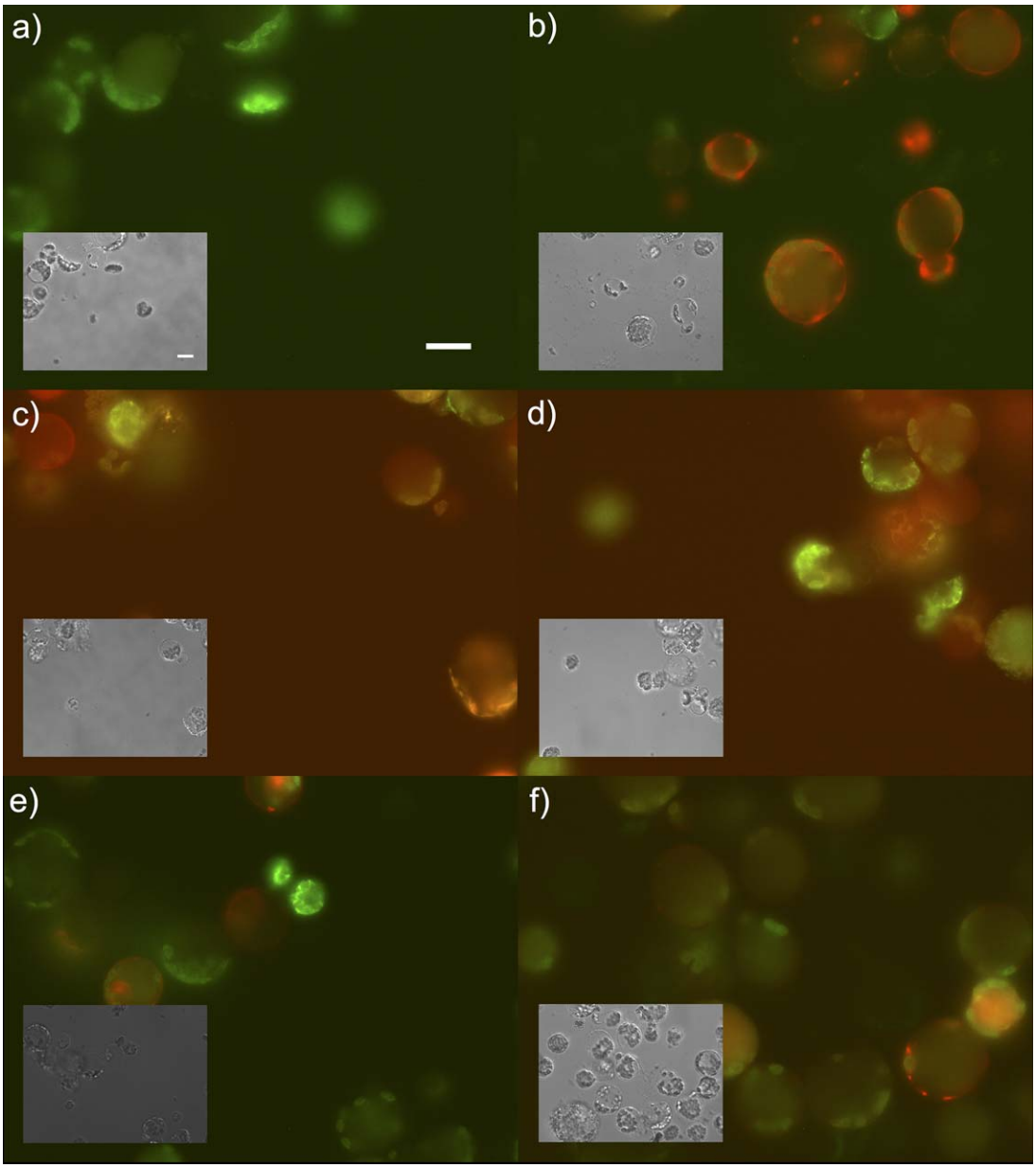

Figure 6. Example treated and untreated images for the three effects detected with the highest confidence. (A) and (B) show untreated and Damnacanthal treated NCRK-313-YFP tagged cells. (C) and (D) show untreated and Damnacanthal treated AHA2-ATPase tagged cells. (E) and (F) show untreated and Tyrphostin treated RabF2a-YFP tagged cells. The images were chosen from those for highest concentration for each drug; the specific image chosen was that with the closest match to the average number of regions per field. Within each row, the chloroplast (green) and protein (red) channels have been contrast stretched to the same dynamic range. Bars, 20 μm. [Color figure can be viewed at wileyonlinelibrary.com]

Region-of-Interest Detection

Due to the spherical shape taken by plant protoplasts after isolation, we implemented a circular Hough transform (28) to create local regions expected to contain a single cell. Each image is first edge-filtered (29) to find circle-edge candidate pixels and all contiguous edge-pixel regions <10 pixels in size are removed. Given the edge image, likelihood that a pixel, $\vec{x}$, is the center of a circle of radius r, is approximated by the response of the image cross-correlated with a filter containing an image of a circle of radius r,

$$\mathcal{L}_r(\vec{x}) = (I \star f_r(\vec{x}))/r. \quad (1)$$

where I is the image, f_r is the filter, and $\star$ indicates the correlation operation. The edge-image is correlated with circles of varying size, producing a likelihood image for a circle of specified radius, where the value of each pixel in the likelihood image corresponds to the confidence of a circle of radius r at that location. Larger circles will have a greater measure of confidence due to the correlation measure, so it is normalized by the radius of the circle detector.

The likelihood images, $\mathcal{L}_{r_1}, \ldots, \mathcal{L}_{r_n}$, are combined to a single 3-dimensional (x, y, r) image with coordinates corresponding to the likelihood of a circle at a pixel (x, y) with a particular radius (r). A search was performed over the (x, y, r) image, identifying potential circle locations in order of decreasing likelihood. We performed the search on a DIC image over radii of 8.05–32.2 μm in steps of 1.61 μm on a 5× downsized image. Circles containing cells were found in order of decreasing confidence until a threshold was reached; in each step, the highest likelihood circle remaining was found, the circular region was used for feature calculation (below), and all circles containing it were removed from the image (set to zero). Figure 2 shows typical segmentation results for 2D images.

Under varying drug-induced perturbations, some cells may no longer exhibit a near-perfect spherical profile. Our ROI detection method still performs well under these conditions. To attain a high likelihood of a circle centered at a given location in the image, the pixel data need not contain a contiguous circle. It suffices that possibly many disjoint arcs align well. Further, to increase the probability of detection for a deformed cell, we use a Gaussian filter to widen identified edges. After ROI identification, active contouring was used to attain the true shape of a cell. To control the false-positive identification rate, any identified ROIs without fluorescent signal within the cell were removed. All of the above procedures ensure our ROI detection pipeline limits false positives and false negatives under perturbed cell shape.

Texture Features

For each circular region found as above, we computed features to describe the spatial relationship between protein patterns in the image (the features were only calculated on pixels within the spherical region). We used Haralick texture features (14,28) computed from different co-occurrence matrices generated for different pairs of fluorescence channels and different spatial offsets. A separate co-occurrence matrix was calculated for each of three combinations of fluorescence channels (protein and protein, protein and chloroplast, and chloroplast and chloroplast) and five different x and y offsets (1, 2, 4, 8, and 16 pixels in each direction). (Co-occurrence matrices for different channels were calculated as the frequency of having a pixel in one channel have one value while the same pixel in the other channel had a second value.) For a given covariance matrix, we computed 12 features as previously described (10). Therefore each region was represented by a set of 12 features for each offset and channel-pair, a total of 60 features per channel-pair, and a total of 180 features for all channel pairs. The values of each feature across all patches and all experiments were z-scored.

Drug Effect-Size Measure

Given a collection of patch features from images corresponding to treated and untreated cell populations, we trained an SVM using fivefold cross validation, weighting data equally in the case of unbalanced classes. For each fold, we determined the signed distance of the members of a test set from the hyperplane where points were assigned a positive distance if they were on the "treated" side of the hyperplane, and a negative distance if they were on the "untreated" side. The distances of all the points are recorded across the fivefolds, and an empirical cumulative distribution function (CDF) with respect to distance from the hyperplane was determined for each the treated and untreated groups. Given the two cumulative distribution functions, we measured the difference between two experiments as

$$\Delta F = \int_{-\infty}^{\infty} |F_{\text{untreated}}(x) - F_{\text{treated}}(x)| \, dx, \qquad (3)$$

where $F_{\text{untreated}}$ and F_{treated} are the CDFs for the untreated and treated groups respectively.

Other Approaches for Detecting Perturbation

We performed two additional comparison analysis using simple, illustrative examples of existing methods. In the first, we determined the average pixel intensity for each region of interest in the GFP channel across all images. Using these values, we performed a two-sample t test for difference between each control–drug pair to detect dose-dependent drug effects. The test statistic for each test was Bonferroni corrected and the number of conditions that were below an alpha level of 0.001 were tabulated.

For the second method, we trained a six class Random Forest classifier to recognize the patterns of the untreated lines using protein features all regions for all controls. With this, a class for each region for each experimental condition was assigned. An experimental condition for a given line was considered perturbed if a plurality of the regions were not assigned to the most frequent cell classification in the unperturbed phenotype of that line.

Availability

A Reproducible Research Archive containing all raw data, software, and processed results is available from http://murphylab.cbd.cmu.edu/software.

Acknowledgments

We thank Dr. Armaghan Naik for helpful discussions, Dr. Roland Nitschke for advice on microscopy, and Katja Rapp for technical support.

Literature Cited

1. Giuliano KA, De Biasio RL, Dunlay RT, Gough A, Volosky JM, Zock J, Pavlakis GN, Taylor DL. High-content screening: A new approach to easing key bottlenecks in the drug discovery process. J Biomol Screen 1997;2:249–259.
2. Abraham VC, Taylor DL, Haskins JR. High content screening applied to large-scale cell biology. Trends Biotechnol 2004;22:15–22.
3. Prinz F, Schlange T, Asadullah K. Believe it or not: How much can we rely on published data on potential drug targets? Nat Rev Drug Discov 2011;10:712.
4. Gough AH, Chen N, Shun TY, Lezon TR, Boltz RC, Reese CE, Wagner J, Vernetti LA, Grandis JR, Lee AV. Identifying and quantifying heterogeneity in high content analysis: Application of heterogeneity indices to drug discovery. PLoS ONE 2014;9:e102678.
5. Meinke DW, Cherry JM, Dean C, Rounsley SD, Koornneef M. *Arabidopsis thaliana*: A model plant for genome analysis. Science 1998;282:662–682.
6. Lee CP, Taylor NL, Millar AH. Recent advances in the composition and heterogeneity of the Arabidopsis mitochondrial proteome. Front Plant Sci 2013;4:4.
7. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, Zwahlen M, Kampf C, Wester K, Hober S. Towards a knowledge-based human protein atlas. Nat Biotechnol 2010;28:1248–1250.
8. Neumann B, Walter T, Heriche JK, Bulkescher J, Erfle H, Conrad C, Rogers P, Poser I, Held M, Liebel U, et al. Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. Nature 2010;464:721–727.

9. Fuchs F, Pau G, Kranz D, Sklyar O, Budjan C, Steinbrink S, Horn T, Pedal A, Huber W, Boutros M. Clustering phenotype populations by genome-wide RNAi and multiparametric imaging. Mol Syst Biol 2010;6:370.
10. Boland MV, Markey MK, Murphy RF. Automated recognition of patterns characteristic of subcellular structures in fluorescence microscopy images. Cytometry 1998;33:366–375.
11. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, Guertin DA, Chang JH, Lindquist RA, Moffat J. CellProfiler: Image analysis software for identifying and quantifying cell phenotypes. Genome Biol 2006;7:R100.
12. Bakal C, Aach J, Church G, Perrimon N. Quantitative morphological signatures define local signaling networks regulating cell morphology. Science 2007;316:1753–1756.
13. Loo L-H, Wu LF, Altschuler SJ. Image-based multivariate profiling of drug responses from single cells. Nat Methods 2007;4:445–453.
14. Haralick RM. Statistical and structural approaches to texture. Proc IEEE 1979;67: 786–804.
15. Schmid F, Trede M. A distribution free test for the two sample problem for general alternatives. Comput Stat Data Anal 1995;20:409–419.
16. Massey Jr. FJ. The Kolmogorov–Smirnov test for goodness of fit. J Am Stat Assoc 1951;46:68–78.
17. Dunn OJ. Multiple comparisons among means. J Am Stat Assoc 1961;56:52–64.
18. Ortiz-Zapater E, Soriano-Ortega E, Marcote MJ, Ortiz-Masiá D, Aniento F. Trafficking of the human transferrin receptor in plant cells: Effects of tyrphostin A23 and brefeldin A. Plant J 2006;48:757–770.
19. Murphy RF, Velliste M, Porreca G. Robust numerical features for description and classification of subcellular location patterns in fluorescence microscope images. J VLSI Sig Proc 2003;35:311–321.
20. Schmid F. Trede M. An L 1-variant of the Cramer–von Mises test. Stat Prob Lett 1996;26:91–96.
21. Dovzhenko A, Dal Bosco C, Meurer J, Koop H. Efficient regeneration from cotyledon protoplasts in *Arabidopsis thaliana*. Protoplasma 2003;222:107–111.
22. Młodzińska E, Kłobus G, Christensen MD, Fuglsang AT. The plasma membrane H+-ATPase AHA2 contributes to the root architecture in response to different nitrogen supply. Physiol Plant 2015;154:270–282.
23. Nelson BK, Cai X, Nebenführ A. A multicolored set of *in vivo* organelle markers for co-localization studies in Arabidopsis and other plants. Plant J 2007;51:1126–1136.
24. Kost B, Spielhofer P, Chua NH. A GFP–mouse talin fusion protein labels plant actin filamentsin vivoand visualizes the actin cytoskeleton in growing pollen tubes. Plant J 1998;16:393–401.
25. Molendijk AJ, Ruperti B, Singh MK, Dovzhenko A, Ditengou FA, Milia M, Westphal L, Rosahl S, Soellick TR, Uhrig JA. Cysteine-rich receptor-like kinase NCRK and a pathogen-induced protein kinase RBK1 are Rop GTPase interactors. Plant J 2008;53: 909–923.
26. Ueda T, Uemura T, Sato MH, Nakano A. Functional differentiation of endosomes in Arabidopsis cells. Plant J 2004;40:783–789.
27. Marc J, Granger CL, Brincat J, Fisher DD, Kao T-H, McCubbin AG, Cyr RJ. A GFP–MAP4 reporter gene for visualizing cortical microtubule rearrangements in living epidermal cells. Plant Cell 1998;10:1927–1939.
28. Ballard DH. Generalizing the Hough transform to detect arbitrary shapes. Pattern Recognit 1981;13:111–122.
29. Canny J. A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell 1986;6:679–698.