Image Analysis in the Life Sciences: Computational Methods and Tools

Dr. rer. nat. Birgit Möller

Dr. rer. nat. Birgit Möller
Pattern Recognition and Bioinformatics
Institute of Computer Science
Faculty of Natural Sciences III
Martin Luther University Halle-Wittenberg

Office address:

Institute of Computer Science
Von-Seckendorff-Platz 1
06120 Halle (Saale), Germany

Habilitation thesis,
submitted to the Faculty of Natural Sciences III
of the Martin Luther University Halle-Wittenberg

Reviewers:

1.　　Prof. Dr.-Ing. Stefan Posch
　　　_____

2.　　Prof. Dr.-Ing. Joachim Denzler
　　　_____

3.　　Prof. Dr. Michael Pound
　　　_____

Date of defense: October 27, 2022

Image Analysis in the Life Sciences:
Computational Methods and Tools

Habilitation thesis

to obtain the academic degree

Dr. rer. nat. habil.

submitted to the

Faculty of Natural Sciences III
of the
Martin Luther University Halle-Wittenberg

by

Dr. rer. nat. Birgit Möller

born on September 20, 1976 in Bielefeld, Germany

Literature references printed in bold within the text point to papers being part of this thesis. The original papers can be found in Chapters 2 to 5, and, where applicable, supplemental material which is suitable for printing is included in Appendix B.

Contents

Chapter 1

Summary

In numerous research areas, scientific progress and the discovery of new knowledge is nowadays unthinkable without support of powerful computational methods and software tools [Nat. Meth. Edit., 2019]. Often exponentially growing amounts and completely new types of data can only be turned into meaningful interpretations by custom-fit computational solutions.

The life sciences do not constitute an exception. In particular for research approaches relying on visual observations and imaging techniques, substantial upheavals occurred in the past. Observations have always been central to science in general [Daston & Lunbeck, 2011], and in particular to the natural sciences. Here already for ancient times first discoveries of basic natural laws from astronomical observations are reported [Wilson, 2017]. While until the end of the Middle Ages only rather rudimentary technical tools were available to support these efforts, the groundbreaking inventions of telescopes and microscopes towards the end of the 16th century [Croft, 2006; Wilson, 2017] marked the transition into a new era. For biology and medicine, microscopes opened insights into the so far unknown microscopic world. They paved the way for pioneering findings, like the first description of cells by Robert Hooke in 1665, and later on, the formulation of the cell theory with cells as the smallest basic units of all living organisms by Matthias Jakob Schleiden and Theodor Schwann in 1838-39 [Mazzarello, 1999].

Since then steady progress in imaging technology, like the combination of microscopy and photography to preserve microscope images beyond the moment [Hazelwood et al., 2007], or the advent of fluorescence microscopy [Dunst & Tomancak, 2019], pushed forward the frontiers of what could be made visible by microscopes. These advances came along not only with improved image quality, but – not least due to the emergence of digital imaging technologies – also with a massive rise in the amounts of acquired image data. This trend is still ongoing today as new techniques regularly emerge and let the quantities of image data and the range of different data types arising from life science experiments explode further [Ouyang & Zimmer, 2017].

Imaging technologies have found their way into virtually all areas of the life sciences. Due to the highly dynamic evolution of molecular and cell biology in past decades, imaging at the microscopic scale of tissues, cells, and subcellular structures – and as of late even at the nanoscopic scale of single molecules and atoms [Ouyang & Zimmer, 2017] – probably forms the most important field of present bioimaging (cf., e.g., [Meijering et al., 2016]). However, the range of research questions in the life sciences being tackled with appropriate imaging methods is not restricted to these scales. Bioimaging also spans the macroscopic scale of complete organisms with manifold applications, e.g., in plant phenotyping or disease detection [Walter

et al., 2015; Li et al., 2020; Mohanty et al., 2016]. In a broader sense, even image data acquired as outcome of experiments that do not directly image biological or biomedical subjects, but encode experimental measurements in visual information, like in case of microarray or electrophoresis experiments [Dowsey et al., 2003; Rueda, 2014], are considered as part of this application domain.

The huge variety of bioimage data, and the even larger plurality of data analysis challenges emerging from these data, have rendered computational tools for data storage, processing, analysis and interpretation an indispensable cornerstone in modern image-based life science research [Levet et al., 2021]. The steadily growing amounts of data have in many cases long ago crossed the limits of what could be analyzed and interpreted manually in a reasonable time frame. Moreover, without appropriate algorithmic solutions, in some areas of life science progress might significantly be thwarted, and some scientific questions may never be answered at all [Cardona & Tomancak, 2012; Myers, 2012]. In addition, visual exploration or manual measurements add a considerable degree of subjectivity to analysis results. This conflicts with high quality and reproducibility of findings which are key to sustainable research [Murphy et al., 2005]. Thus, objective and automatic computational methods for data analysis are indispensable as a matter of principle. In consequence, a large amount of different algorithms and tools for manifold bioimage analysis problems have been proposed to date, and their number is still growing continuously [Eliceiri et al., 2012; Lucas et al., 2021].

This thesis presents an overview of the author's work in the field of image analysis for life science applications over the last years. Most of this work was carried out in close collaboration with cooperation partners from different areas of the life sciences. Our work mirrors the plethora of research questions, types of bioimage data, and image analysis challenges occurring in this field. From a methodical point of view, our solutions cover a large range of different algorithmic approaches, techniques and analysis strategies from the fields of image processing and analysis, as well as machine learning. We thereby focus on the detection and segmentation of objects and structures in images, and their measurement and quantification.

In the remainder of this thesis, we discuss our work in detail. The structure of presentation roughly follows the general hierarchy of abstraction levels as it is often found in classical image processing and analysis [Jähne, 2005, Chap. 1], proceeding from low to increasingly higher degrees of abstraction. We start our discussion in Section 1.1 and Chapter 2 with work on the detection of elementary image primitives allocated to the lowest abstraction level in digital images. In our case, we focus on point- or spot-like structures and small particles which constitute a frequently appearing pattern in bioimage data. We have worked on the extraction of such primitives in the context of object detection and image registration problems.

Subsequently, in Section 1.2 and Chapter 3 we present our contributions to the problem domain of object segmentation from bioimage data. We present work on the segmentation of various types of cells, e.g., leaf epidermis cells in plants, human cancer cells, or neurons. Depending on the appearance of the cells and their characteristics, we adapt and extend different methodical approaches for their segmentation. In addition, we tackle the problem of extracting phenotypical data of roots from images combining novel methods of deep learning with classical morphological image analysis. The segmentation of objects from bioimage data is usually followed by further analysis steps. We discuss solutions for measuring and quantifying specific properties of extracted objects in Section 1.3 and Chapter 4.

Besides devising algorithmic solutions optimally tailored to the concrete research question at hand and enabling automated and objective data extraction and analysis, a second major

objective in all of our research efforts has always been to evolve our methodical findings into intuitive and user-friendly software tools. Following the spirit of open-source software development in bioimage analysis [Cardona & Tomancak, 2012], we released numerous of our tools to the public under open licenses. Moreover, we have not just released a collection of separate tools, but also devised an integrated concept and elaborate library to support effective and focused development of new algorithmic solutions, which naturally forms the basis for most of our own tools. We discuss this concept, related tools, as well as supplemental software solutions emerging from our research in Section 1.4 and Chapter 5.

In Section 1.5 we finally summarize the main findings and contributions of this thesis.

1.1 Detection of Low-Level Features: Keypoints, Spots, Particles

Spots, keypoints, and other punctiform features are in image analysis primarily considered as image primitives on the lowest level of abstraction. Particularly keypoints often act as functional components in complex algorithms, e.g., in object recognition or image registration [Loncomilla et al., 2016; Ma et al., 2021]. Consequently, in such applications not each single entity of a certain type of structure is of relevance, but a more global view on the complete feature set is usually sufficient to solve a certain task. In contrast, in bioimages spots and point-like structures are often directly linked to biological or biomedical objects in experiments and are of central importance for the research questions. Here, usually each single object needs to be accurately detected and is of individual importance. Accordingly, there exist different views on point-like features in images, and depending on the underlying research question or application scenario different strategies to deal with these structures are implemented in practice.

Below, we discuss solutions for two bioimage analysis problems where spots and point-like structures play central roles. The first example in Subsec. 1.1.1 presents work from the field of proteomics where spot-like structures in images define the key outcome of proteomics experiments. Though these spots form the dominating structure in the images, we could show that tasks like image registration can well be solved without the need to solely rely on the spots.

The second example originates from the field of fluorescence microscopy, which is probably the most common area of application where point-like structures and especially particles have to be detected in images. Following the definition in [Chenouard et al., 2014], we summarize under the term 'particle' in this context all potential target objects ranging from a single molecule to a macromolecular complex, organelle, or virus. Subjects of this kind often appear as small to medium-sized spots in fluorescence microscope images. Many research questions in the life sciences are linked to the spatial distribution or dynamic behavior of such objects within cells, tissues, or organisms. Hence, many bioimage analysis pipelines subsume tasks like localizing, counting, or tracking particles in spatial or temporal contexts [Meijering et al., 2012]. We faced the challenge of extracting data about various types of particles from fluorescence microscope images in different research contexts. To solve these tasks, we devised a scale-adaptive detection approach which we will discuss together with sample applications in Subsec. 1.1.2.

1.1.1 Features for image registration in proteomics experiments

Proteins are essential building blocks of cells and involved in virtually all cell functions. They are synthesized within a cell by a two-stage process comprising *transcription* and *translation*. First, nucleotide sequences of the DNA are transcribed into RNA molecules, and second, these are further translated into linear chains of amino acids. These chains form the fundamental

structure of proteins and define their individual physio-chemical properties [Alberts et al., 2015, Chap. 3]. The process of protein synthesis is based on a complex regulatory system that dynamically adapts transcription of genes and subsequent translation into proteins, but also modifications of proteins and their degradation, to changes in the cellular environment [Vogel & Marcotte, 2012]. In consequence, the set of all proteins present in a cell, i.e., the *proteome*, dynamically changes over time. Monitoring these changes and tracking the composition of the proteome as a function of cell state allows to systematically discover functional processes within a cell, and to reveal the role of proteins as central players in these processes [Speicher, 2004].

Two-dimensional gel electrophoresis (2-DE) is an established procedure in proteomics for dissecting the composition of proteomes [Magdeldin et al., 2014]. 2-DE separates proteins spatially by exploiting their individual isoelectrical properties and molecular weights. In 2-DE a mixture of proteins is first exposed to an electric field on a gel with a pH gradient which causes the proteins to migrate to an equilibrium position in the carrier medium in accordance with their isoelectric points [Pomastowski & Buszewski, 2014]. Second, proteins are further split-up in a direction orthogonal to the first one by binding a specific detergent to the proteins which masks their intrinsic charges and induces a uniform load per weight unit. In an electric field they are then pushed through the field at a speed proportional to their loads, triggering a separation according to their molecular weights [Pomastowski & Buszewski, 2014]. This finally results in two-dimensional patterns of protein accumulations, which after staining with an appropriate dye, e.g., appear as dark spots on a brighter background (Fig. 1.1 and 1.2).

Protein mixtures subjected to 2-DE usually originate from varying cellular conditions. Differential comparisons of the resulting spot patterns form the fundament for correlating changes in the proteomes with these conditions. To enable the application of
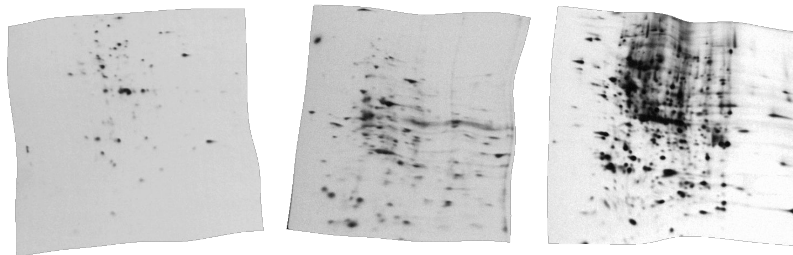
Figure 1.1: Sample 2D gel images from the LECB 2D PAGE Gel Images Data Sets used in [Möller & Posch, 2009] with low to high structural complexity (left to right), artificially deformed applying thin plate spline transformations.

computational methods at this stage, images of the 2-DE gels are acquired (Fig. 1.1). This results in several image analysis challenges, like quantifying the protein spots or identifying relevant differences in spot patterns of gels [Dowsey et al., 2003].

In addition, the non-rigid nature of the gels gives raise to local deformations during the 2-DE process which require compensation prior to any direct comparison of gels. Such a compensation of deformations defines a task of non-rigid image registration [Sorzano et al., 2008]. Generally the two basic concepts of featureless versus feature-based registration can be distinguished, both being used in the domain of 2-DE gel images [Aittokallio et al., 2005]. While methods of the first category exclusively rely on intensity information, methods of the second class build on matches of distinct image features. The predestined type of features in gel images are of course the protein spots themselves, and they commonly yield the dominating fundament for feature-based image registration in this application [Robotti et al., 2021]. Although a multitude of spot detection approaches have been devised [Martinotti & Ranzato, 2016], robust spot detection often marks a notable difficulty [Clark & Gutstein, 2008]. Also, relying exclusively on domain-specific spot features in the registration of 2-DE gel images and directly coupling the number of feature locations to a potentially low number of protein spots can have drawbacks.
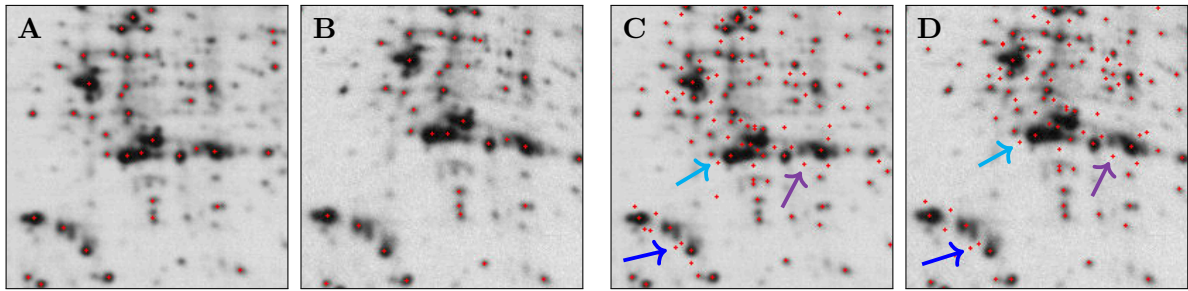
Figure 1.2: Features detected with the ring detector (A,B) and the keypoint detector SIFT (C,D) on original (A,C) and synthetically deformed image clips (B,D). Arrows with identical colors mark exemplary pairs of corresponding feature locations provided by SIFT which are not directly linked to specific protein spots.

We addressed the problem of robust feature detection for 2-DE gel images in [**Möller & Posch, 2009**] and could prove that domain-independent, general purpose keypoint features may help to overcome these issues. In an extensive comparative study on different types of gel images, we evaluated the performance of spot-specific and model-based detectors versus the general purpose keypoint detectors SIFT [Lowe, 2004] and SURF [Bay et al., 2006]. We found that the number of features extracted by SIFT and SURF and their robustness in terms of repeated detection in images synthetically deformed by common non-rigid transformations (Fig. 1.1) in tendency outperformed spot-specific detectors (cf. Fig. 1.2). The built-in scale-invariance of SIFT and SURF allows to detect features on different levels of abstraction, like constellations of spots supplemental to individual spots (see arrows in Fig. 1.2), and provides inherent advantages. This revealed new opportunities and flexibility in the design of robust algorithms for feature-based gel image registration by adopting domain-independent detectors.

1.1.2 Particle detection in fluorescence microscope images

In proteomics studies based on gel electrophoresis, functional processes within cells are monitored indirectly via changes in the proteome. This decouples the occurrence of proteins from the places where they localize and operate in cells, leaving the spatial dimension of protein activity and interaction unconsidered. The localization of proteins and other actors in translational processes and their spatio-temporal traits, however, form important determinants for understanding cell structure and function [Lundberg & Borner, 2019].

The spatio-temporal dimension of processes within cells can be approached by visual observation of subcellular structures and functional units directly in the cells. Potential target structures exhibit an enormous diversity, ranging from essential building blocks of cells like nuclei, over the universe of proteins and other biomolecules, to dynamic compartments emerging and disappearing situatively [Alberts et al., 2015]. To visualize structures, fluorescence microscopy has evolved as one of the key technologies [Thorn, 2016]. It relies on fluorophores to label target objects. These can either be attached to specific antibodies which bind to the objects in question, or they may be genetically introduced into cells which then express fluorescent proteins [Nienhaus & Nienhaus, 2017]. When fluorescently labeled samples are exposed to light of a suitable wavelength, fluorophores emit photons of a defined wavelength by which target objects become visible and can be localized (e.g., [Dobrucki & Kubitscheck, 2017]).

The automatic detection and quantification of structures in fluorescence microscope images brings manifold challenges to bioimage analysis. Due to countless influence factors present in experimental setups, like the technology of the microscope, the protocols for sample prepa-

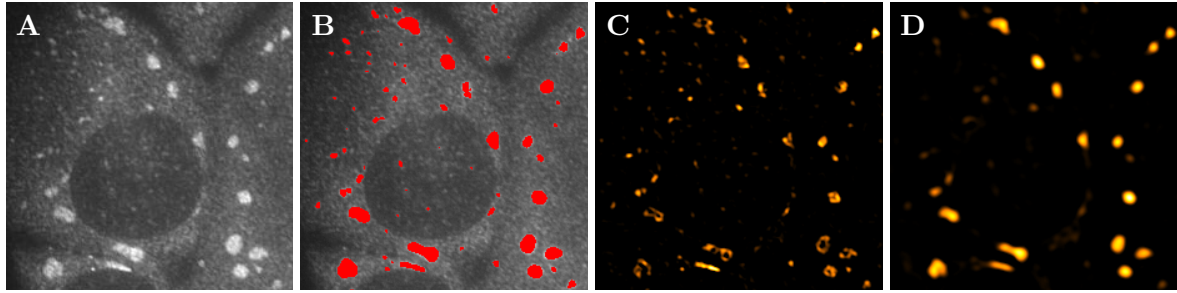**Figure 1.3:** Particle detection result for data used in [Greß et al., 2010]: input image (A), detection result (B), heatmaps of correlation images over scale ranges $[2, 3]$ (C) and $[3, 4]$ (D). Heatmaps show that smaller scales primarily focus on point-like structures (C), while in larger scales spot-like patterns dominate (D).

ration and imaging, intrinsic characteristics of fluorescent dyes, or physio-chemical properties of the targets [Dobrucki & Kubitscheck, 2017], there is no unique appearance of fluorescently labeled structures in images. Image quality and the level of difficulty of analysis tasks may vary significantly, e.g., due to differing amounts of distracting fluorescence signals or low contrast between fluorescently stained target structures and image background [Waters, 2009]. Also, the structural properties of the fluorescently labeled entities themselves are essential.

A frequent task in the context of fluorescence microscope images is the detection of particles. Ideally, particles stick out as bright points or small spots from an otherwise dark background in such images. In practice, however, issues common to fluorescence microscope images in general, and also the size of the particles, inhomogeneities in their intensity profiles, and large variations within their shapes, may limit their overall visibility (Fig. 1.3 A and 1.4). Hence, their detection and quantification in fluorescence microscope images is usually challenging.

In [**Greß et al., 2010**] we proposed an approach for detecting particles in fluorescence microscope images capable of dealing with many of the challenges outlined above. The motivation for devising a new algorithm for this task originated from a project with cooperation partners from molecular medicine who aimed to count stress granules (SGs) and processing bodies (PBs) in microscope images. SGs and PBs are cellular compartments localized in the cytosol of cells which assemble primarily in response to stress, e.g., induced by viral infection [Riggs et al., 2020]. They play important roles in post-transcriptional regulation and are considered important with regard to many severe human diseases [Ivanov et al., 2019]. To date a large amount of knowledge about their formation and molecular compounds is available, while their general functions still wait for further elucidation [Youn et al., 2019; Riggs et al., 2020]. One building block in experiments towards discovering their roles and functions are statistical investigations on their occurrence and spatial distributions under varying cellular conditions.

To detect SGs and PBs, we built on work of Olivo-Marin [2002] who exploits correlations between subband images of an à trous wavelet decomposition of a given image to detect bright spots. The detection is based on the idea that locally maximal wavelet coefficients associated with significant image structures tend to propagate across different scales, while locally maximal coefficients originating from noise do not. Thus, in a correlation image calculated as product over a user-selected subset of denoised wavelet images, i.e., over a selected range of scales, large values are assumed to refer to relevant spots. While the original approach showed a generally good detection performance for spots with moderate differences in size, it reached its limits in our target application where we had to deal with greatly varying sizes of SGs and PBs.

Hence, in [**Greß et al., 2010**] we refined the method towards improved scale-adaptivity. By correlating wavelet coefficients not only over one selected interval of scales but over various

intervals, we managed to detect particles of considerably different sizes in a single run. The key to achieve this result is a new scheme for particle hypothesis selection based on statistical tests. Given several scale intervals for which correlation images are calculated, competing particle hypotheses may result for a single image location. Since at every image location at most one particle of a certain size can exist (Fig. 1.3 C, D), a decision for the correct scale is imperative. We solved this issue by creating a hierarchical forest of particle hypotheses in which competing hypotheses between adjacent scales are compared and singled out based on the concept of meaningful events [Desolneux et al., 2003]. This finally provided us with a flexible detection algorithm capable of automatically generating particle hypotheses over different scales, and selecting the correct scale based on profound statistical measures (Fig. 1.3 B).

In [**Greß et al., 2010**] we could show that our extended algorithm provides reliable counts of SGs and PBs from images of fluorescently labeled U2OS osteosarcoma cells (see Fig. 1.3 for an example with SGs). Particles were robustly detected on a wide range of scales with an increased sensitivity and detection accuracy. These results were confirmed in our subsequent publication [**Möller et al., 2011b**], which we discuss more detailed in Subsec. 1.2.2. Here, cells from the human hepatoma line HUH7 were analyzed, and this time not only SGs and PBs were fluorescently labeled, but a DAPI-staining of the nuclei was also given. This formed the fundament to segment individual cells, using the nuclei as initializations for active contours [**Möller et al., 2010, 2011b**]. As result SGs and PBs could be assigned to individual cells. Such cellwise statistics are of interest if in larger cell populations not all cells are equally affected by stress factors, and global counts on complete images would bias quantitative investigations.

In our work [**Bley et al., 2015**] also the sizes of particles were of interest to investigate the role of SGs in situations of cellular stress, particularly with regard to cellular mechanisms of mRNA stabilization. SGs were detected using our tool *MiCA* (Subsec. 1.4.3) as user interface to the implementation of the algorithm in our software library and toolbox *MiToBo* (Subsec. 1.4.1). Given manual cell annotations, the number of cells where SGs were observed was determined and area fractions calculated as the percentage of cell area occupied by SGs. Extracting area fractions not only requires the decision if a particle is locally present or not, but also an accurate identification of all its pixels for estimating size. Our approach proved suitable also for this extended task, yielding reliable quantitative data. New insights into the function of SGs in post-transcriptional control could be gained, and the dynamics of protein exchange between SGs and their environment further uncovered.

The particle detection algorithm was initially developed for counting SGs and PBs. As no specific assumptions about the target objects are exploited except that they stick out from their local surroundings, we successfully applied the approach also to other types of particles. In [**Bürstenbinder et al., 2017**] we used the detector in studies on the distribution of proteins in the plasma membrane of *Nicotiana benthamiana* leaf cells. The proteins localize to plasma membrane subdomains which results in

Figure 1.4: Detection of labeled proteins localizing to subdomains in the plasma membrane: inverted input image with particles of varying size and contrast (A), and result (B).

non-uniform, punctuate distribution patterns in the images, which we accurately quantified with our detector (Fig. 1.4). In [Franke et al., 2015] and [Möller & Schattat, 2019] we adopted
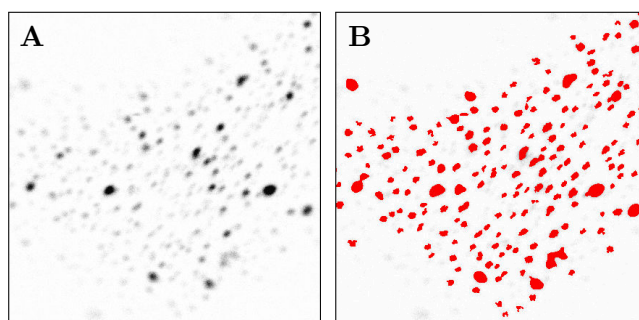
our particle detection approach as part of semi-automatic pipelines for quantification and categorization of plastids in leaf epidermis tissue. In images of *Arabidopsis thaliana* and *Nicotiana benthamiana* the large majority of plastids could automatically be extracted, reducing the overall processing time of the images dramatically compared to pure manual analysis. Also nuclei and peroxisomes in leaf cells of *N. benthamiana* could successfully be detected, underlining the flexibility of our algorithm and the wealth of potential areas of application.

1.2 Object Segmentation

The overall aim of object segmentation is to identify individual objects of interest in an image, separate them from each other and from the image background, and generate an explicit representation for each of them. Traditionally, segmentation tasks are tackled adopting one of two complementary perspectives, either focusing on regions or contours [Wu & Castleman, 2008].

A region is defined as a set of adjacent pixels which adhere to a common homogeneity criterion, e.g., share similar colors or textural properties [Sonka et al., 1999, Sec. 5.3]. Region-based segmentation can either be implemented in image space where similarities among pixels in spatial neighborhoods are the central guideline along which regions are formed, or in feature space which considers pixel properties independent of spatial relations. Though in optimal case each segmented region is finally associated with exactly one individual object, in practice over- or undersegmentation frequently occur, and sophisticated workflows for post-processing may be needed to obtain desired target objects from the results of an initial segmentation.

While the region-based view focuses on similarities of pixels, the dual concept of contours puts emphasis on the delimitation of objects. The contour of an object is defined as its outer boundary that encloses the interior and separates the object from its surroundings. Contour-based segmentation relies on the assumption that object boundaries become apparent as discontinuities in the intensity landscape of an image [Sonka et al., 1999, Sec. 5.2]. The direct extraction of contours completely enclosing each object is naturally the most desirable result, but, e.g., due to weak local contrast, often only disconnected contour fragments can be extracted. Thus, linking fragments to end up with closed contours is a common issue in contour-based segmentation.

One approach to bypass several of the typical problems of region- and contour-based methods are active contour models. They have gained wide popularity due to their great trade off between flexibility and efficiency [Delgado-Gonzalo et al., 2014; Xing & Yang, 2016]. The mathematical concepts underlying active contours inherently support the implementation of topological constraints, e.g., allow to enforce the extraction of a fixed number of closed contours from an image. The target criteria of a segmentation task are in this approach encoded in energy functionals. These functionals allow to consider and arbitrarily combine a multitude of different criteria, e.g., based on contour cues [Xie & Mirmehdi, 2008b], region properties [Michailovich et al., 2007; Wang et al., 2009], or texture measures and shape priors [Cremers et al., 2007], and form the basis for the large flexibility of these models.

In recent years, the field of image analysis has faced fundamental upheavals, triggered by the revival of neural networks and the emergence of advanced machine learning concepts. In particular deep learning based on novel network architectures has led to tremendous progress in solving image analysis and computer vision problems, e.g, regarding object recognition or semantic labeling [Eckersley et al., 2021]. For object segmentation, adopting methods of deep learning marks a fundamental abandonment from predefined image primitives and hand-crafted features, like keypoints, contours, or regions. In deep learning the segmentation task is formalized as the problem of estimating optimal parameters of a functional model which maps input

images onto desired segmentation results. Hence, the central assumption underlying deep learning is that the model by itself learns during training which clues contained in the input image data provide relevant information to successfully solve an object segmentation task at hand.

The general developments in object segmentation over the decades are also mirrored in the field of bioimage analysis. Here, manifold kinds of target objects need to be segmented, with cells being probably the most frequently occurring type [Meijering, 2012; Boquet-Pujadas et al., 2021]. Object segmentation may be seen as one of the most challenging, yet most crucial steps in many bioimage analysis pipelines [Meijering et al., 2016]. The outcomes of an initial segmentation often form the fundament for subsequent analysis steps (cf. Sec. 1.3), and their accuracy is paramount to the overall quality of the final results delivered by a complete pipeline.

Like in general object segmentation, also in bioimage analysis a large collection of segmentation approaches has been devised, covering the whole range of different views on the segmentation problem [Meijering, 2012, 2020]. In our own work, we solved the problem of cell segmentation in various contexts. In Subsec. 1.2.1 we present our approach for applying matched filters as parts of elaborate contour-based analysis pipelines to segment different kinds of plant cells, and in Subsec. 1.2.2 we discuss a study on general properties of active contour models and their successful adoption to segment different types of cells in microscope images.

A second area for which we devised segmentation methods are root images. An initial segmentation of roots is prerequisite for topological analysis and extraction of phenotypic data [Atkinson et al., 2019]. We adopted convolutional neural networks (CNNs) in combination with skeleton-based morphological analysis to segment roots from images and identify structural parts of the roots. In Subsec. 1.2.3 we present our approach with which we successfully participated in the *Arabidopsis thaliana* Root Segmentation Challenge at the ICCV Workshop on Computer Vision in Plant Phenotyping and Agriculture (CVPPA) in 2021.

1.2.1 Contour extraction based on matched filters

One of the first steps in contour-based object segmentation is to enhance discontinuities in the intensity landscape of images, which are commonly associated with object boundaries. Linear filters are a popular tool for this task (e.g., [Snyder & Qi, 2017, Chap. 5]). Assuming that objects are entirely brighter or darker than the surrounding background, contours are predominantly modeled as steep ascents or descents in local intensities along object boundaries. Thus, they can be localized adopting linear filters that approximate first or second order partial derivatives of an image function [Nixon & Aguado, 2020, Chap. 4].

Cells are often visualized by fluorescently labeling their cell membranes or walls. In microscope images these structures appear brighter than the background, but likewise brighter than the cell interior (Fig. 1.6). Hence, cell boundaries form ridge-like, thin and elongated structures in the intensity landscape. For detecting such special structures, matched filters have proven suitable [Fraz et al., 2012]. They are designed to match the target structures, which in case of cell boundaries calls for anisotropic properties of

Figure 1.5: Example of a matched filter kernel with Mexican Hat profile for detecting ridge-like structures in images.

the filter kernels (Fig. 1.5). Consequently, when convolving image functions with such kernels the filter responses depend on the orientation of the kernel and require the filter to be applied in different orientations to an image. While this induces a great computational effort, it allows to accurately localize target structures with high sensitivity.
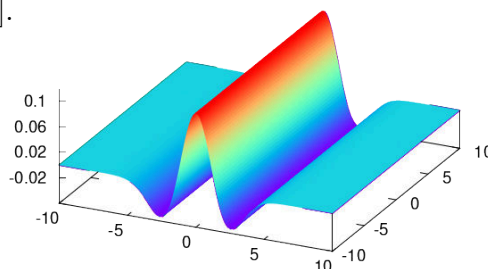
Solving the jigsaw puzzle of pavement cell shape. We successfully adopted matched filters as fundament of a fully automatic approach for the segmentation of leaf epidermis pavement cells from microscope images. The method is part of a larger pipeline integrating cell segmentation with quantification and analysis of cell shape – named *PaCeQuant* as an acronym for **pa**vement **ce**ll shape **quant**ification. *PaCeQuant* and its implementation in a user-friendly software tool were initially published in [**Möller et al., 2017**], while a detailed description of its usage dedicated to end users can be found in the invited book chapter [**Möller et al., 2019b**].

The epidermis of leaves and its various types of cells play an essential role in the regulation of leaf function, growth and development. The most frequently occurring cell type in the epidermis of many plants are pavement cells which attract attention due to their prominent shapes [Jacques et al., 2014], being highly interlocked and jigsaw puzzle-like (Fig. 1.6). To unravel the cellular processes underlying leaf development and expansion, properties of pavement cells and particularly the changes in shape which they undergo during growth processes or in mutant studies have proven a valuable source of information, e.g., to identify key regulators of leaf morphogenesis [Zhang et al., 2011; Armour et al., 2015]. Hence, the quantification of pavement cell shape with appropriate feature measures forms an important fundament of ongoing research in this field [Sapala et al., 2019]. To place such investigations on a solid data basis, the analysis of sufficiently large and representative data sets is indispensable. This in turn is hardly possible without techniques for an automatic segmentation of cells from images.

PaCeQuant is the first tool to provide an integrated, fully automatic segmentation of pavement cells and a subsequent quantification of their shape characteristics, well-suited for high-throughput screenings of large cell populations [**Möller et al., 2017**]. For shape analysis, staining protocols exist to fluorescently label the membranes or walls of pavement cells [**Möller et al., 2017, 2019b**]. By means of confocal microscopy single optical sections of the stained material can then be scanned, resulting in images as shown in Fig. 1.6, A and C. To extract cell contours from such images, we propose a pipeline in *PaCeQuant* which at its core applies a matched filter with a Mexican hat profile (Fig. 1.5, cf. [Sofka & Stewart, 2006]) to identify candidate locations for cell contours. This filtering step is complemented with preceding image enhancement and subsequent post-processing steps. These

Figure 1.6: Images of leaf epidermis pavement cells of the *A. thaliana* wildtype (A) and of a Katanin mutant (C) with stained cell walls acquired from cotyledons of 5-day-old seedlings, and pseudo-colored segmentation results of *PaCeQuant* (B,D). Images are inverted for improved visibility.

amongst others subsume local binarization of filter responses, and gap closing exploiting cues of a binary watershed segmentation. The complete pipeline generally allows to robustly extract cell contours from images as demonstrated in Fig. 1.6 B and D.
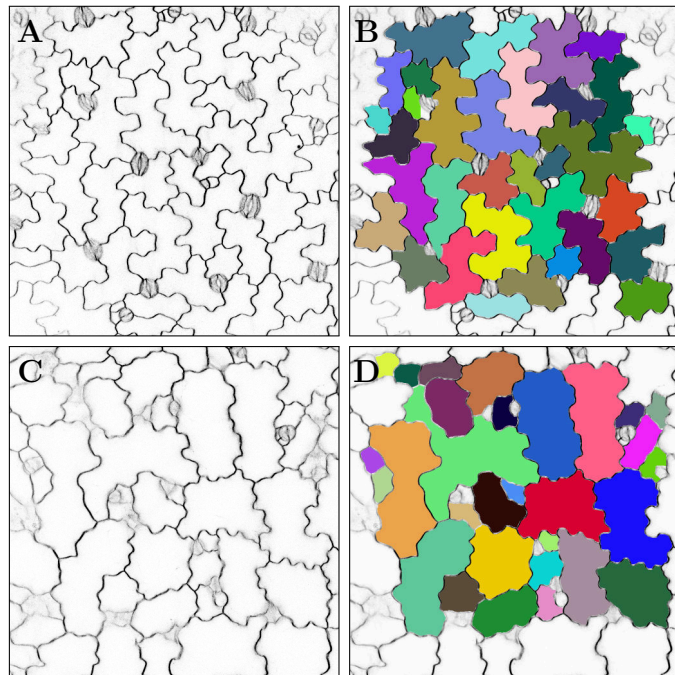
Before *PaCeQuant* emerged, shape analysis of pavement cells was in many cases severely limited. Often only few cells were considered as the extraction of cell boundaries from images relied on time-consuming manual procedures [Vanhaeren et al., 2015]. With *PaCeQuant* this situation changed and highly accurate cell contours can now be automatically extracted [**Möller et al., 2017**]. For a set of 15 cells for which ground truth annotations were given we achieved median recall and precision values on the cell pixels of around 0.98. Moreover, an extended study on a set of 373 cells, for which the segmentation results were manually reviewed, revealed that less than 10% of the cells suffered from relevant local segmentation inaccuracies. Notably, the statistical distributions of shape features extracted from the cells were only slightly altered by these issues (more details about our shape features are discussed in Subsec. 1.3.1).

In summary, this proves our novel segmentation as integral part of the *PaCeQuant* pipeline to be suitable for high-throughput screenings of large data sets, yielding a solid fundament for investigations on cell shape formation. Over time the core application was extended to allow for larger flexibility in configuring single stages of the pipeline, and it was supplemented with user-friendly tools to ease the practical handling of *PaCeQuant* itself as well as of extracted contour and shape data (see Subsec. 1.3.1 and 1.4.3 for details). *PaCeQuant* is still under active development investigating, e.g., faster segmentation strategies [Schwede, 2020].

Matched filters for cell segmentation in cytoskeleton studies. Another area of application requiring an accurate delineation of cell boundaries is the analysis of subcellular organization on the single cell level. As mentioned in Subsec. 1.1.2, regulatory processes within cells are deeply linked to the spatial organization of intracellular structures. One key structure of interest within cells is the cytoskeleton for which we devised quantitative approaches to characterize its structural properties. We will present more details on this work in Subsec. 1.3.2. Below, we first discuss our solution for segmenting cells from corresponding image data as an indispensable prerequisite for later characterization of the cytoskeleton in individual cells.

Compared to cell segmentation for shape analysis, segmentation in the context of cytoskeleton analysis poses additional challenges to automatic approaches. Here, it is not sufficient to stain cell membranes, but simultaneously relevant structures of the cytoskeleton need to be visualized. In addition, to properly cover the three-dimensional structure of the cytoskeleton, z-stacks are acquired. Cell membranes and cytoskeletons can be imaged using fluorophores of differing wavelengths which allows to separate them in different image channels. Yet cell segmentation remains demanding as the overall image quality is often low, e.g., due to increased amounts of distracting signals, strongly varying local contrast, or phantom structures (Fig. 1.7).

In [**Möller & Bürstenbinder, 2019**] we present an approach to tackle the segmentation of cell contours from such image data based on maximum projections of the cell membrane channel along the z-axis (Fig. 1.7 A, C). Methodically, we adopt the idea of contour extraction from *PaCeQuant* [**Möller et al., 2017**], i.e., make use of matched filters to enhance cell boundaries. Due to larger gaps remaining between fragments after binarization and initial post-processing, however, advanced gap closing methods are required. We transform the gap closing problem into a shortest path problem on the graph defined by the set of pixels and their pairwise neighborhood relations. Edge weights are derived from the matched filter responses at pixels linked by an edge, and the Dijkstra algorithm is applied for determining paths between selected pixels with minimal costs to close gaps in optimal accordance with local image information.

Experiments on test data proved that our approach is capable of automatically extracting on average more than 90% of the contour of each cell, missing mostly only short boundary sections (Fig. 1.7 B, D). These high recalls are, however, accompanied by significantly smaller
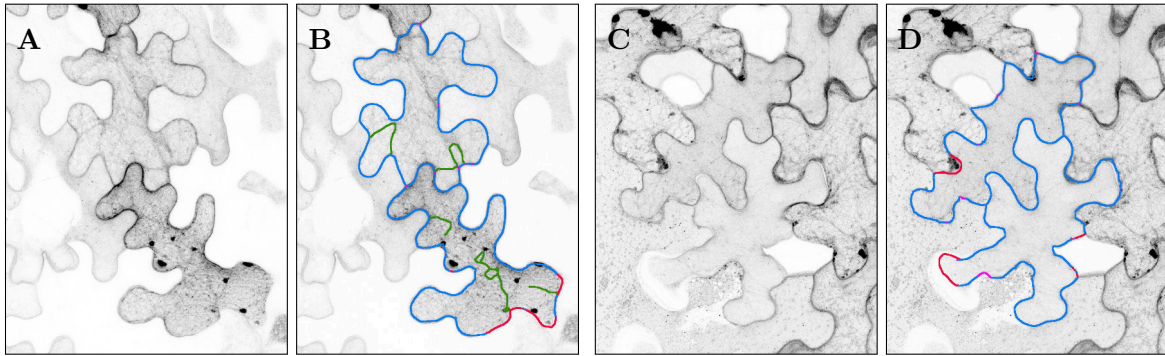
Figure 1.7: Semi-automatic contour extraction for cytoskeleton analysis: maximum projections along the z-axis of z-stacks of stained cell membranes (A, C), and final segmentation results (B, D) illustrating automatically extracted contour sections (blue), automatically filled gaps (magenta), manually closed gaps (red), and surplus contours in the cells (green) to be removed manually. Images are inverted, and surplus contours outside of the target cells have been omitted for clearer presentation.

precisions. They are mainly caused by spurious structures sharing a significant similarity with real contours and hardly erasable automatically (Fig. 1.7 B). Thus, manual post-processing is inevitable to remove surplus fragments, and also to close the remaining gaps. Anyway, it is usually much easier to remove structures than to manually fill in missing segments (e.g., using functionality available in our own tool for editing segmentation results, cf. Subsec. 1.4.3). In conclusion, our approach still yields a user-friendly, though semi-automatic solution for cell segmentation in such challenging scenarios, keeping necessary user intervention at low level.

1.2.2 Active contour models: properties, energies and optimization schemes

Object segmentation with active contour models builds on the idea to translate the extraction of object contours in an image into an optimization problem. Given appropriate mathematical models of contours which allow to precisely describe their shapes and positions within an image, optimization methods are applied to optimally align given contour models with object boundaries present in an image [Delgado-Gonzalo et al., 2014]. The central cornerstone in this approach are energy functionals defined over contour models and image data, which encode the target criteria of the segmentation task. They allow to assess the quality of a segmentation, i.e., quantify the agreement of a segmentation as currently represented by a contour model with expectations encoded in the energy functionals. Given that an optimal positioning and shaping of the models with regard to intended object contours yields minimal energy values, the process of segmentation is turned into an energy minimization problem. In practice, active contours can often be adopted to new application areas straightaway by customizing their energy functionals and without need for any substantial changes to the underlying contour representations or optimization strategies.

Snakes versus level sets. The two most popular variants of mathematical representations for object boundaries in the context of active contour models are snakes [Kass et al., 1988] and level sets [Sethian, 1999; Osher & Fedkiw, 2003]. Snakes represent object contours explicitly in terms of parametric curves in 2D (Fig. 1.8), while level sets adopt an embedding function Φ to implicitly represent a contour through the values of Φ (Fig. 1.9). Each of the two approaches has its own individual specifities which need to be considered before selecting one of the models for a given segmentation task.

To provide guidance in this selection process, in [**Möller & Posch, 2012**] we investigated both representations in a comparative study, carving out their individual characteristics based on sample segmentation tasks. As a specialty of our work and in contrast to others, we exploited the fact that energy functionals tailored for either of the two models can in general be transferred directly to the other representation. Hence, we applied for both concepts the same energy functionals to focus the comparison on individual properties of the representations rather than on differences related to segmentation criteria.

As sample applications we considered the task of tissue segmentation for scratch assay analysis, originally solved with level sets in [Glaß et al., 2011, 2012], and a task of cell segmentation from fluorescence microscope images where we adopted snakes for a similar problem in [**Möller et al., 2011b**] (see below). In both applications we used region-based segmentation criteria formulated on the fundament of the Chan-Vese energy functional [Chan & Vese, 2001]. Snakes were optimized

Figure 1.8: Snakes in 2D make use of parametric curves, e.g., represented by points and directed line segments in image space.

by iterative gradient descent which requires to choose a suitable step size, while for level sets we adopted a parameterless sequential optimization scheme [Song & Chan, 2002].

The study at first confirmed evident differences inherent in both contour representations. E.g., snakes require explicit strategies to keep a valid topology, like regular corrections for self-overlaps, while for level sets additional efforts are necessary to avoid topological changes, e.g., splits or merges of objects during the segmentation process [Han et al., 2003; Glaß et al., 2011]. Apart from this, our study primarily revealed that identical energy functionals not necessarily result in the same segmentations. While minor differences in overall segmentation quality

Figure 1.9: Level set function Φ in 2D with positive (red) and negative (blue) values encoding the object interior and exterior, separated by the zero level $\Phi(x, y) = 0$ (black line) representing the contour.

were observed with regard to the scratch assay task, in cell segmentation an additional regularization term had to be added to the energy functional for the level sets to achieve comparable performance. This indicates a larger sensitivity of level sets to noise, and in particular points to an implicit regularization inherent in the parametric snake representation. Hence, besides demonstrating key properties of both models, our study especially outlined implicit differences not obvious at first sight, but without doubt of high relevance when selecting representations and energy functionals for specific applications in practice.

Cascaded optimization schemes for snakes. One widespread strategy to adopt active contour models for new target applications is to combine and fine-tune existing energy functionals, or extend them with novel and apt criteria. Though a large collection of distinct criteria has been developed into ready-to-use energy functionals over time, in certain situations it still remains challenging to solve a segmentation problem by only focusing on energy functionals.

We faced such a problem in research on the role of stress granules (SGs) and processing bodies (PBs) in post-transcriptional regulation which we already discussed in Subsec. 1.1.2. In the project, among other things, appearance and quantities of SGs and PBs in individual cells of a population should be analyzed, requiring an accurate cell segmentation. Established staining protocols allow to acquire multichannel fluorescence microscope images with various target
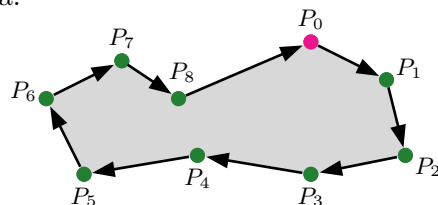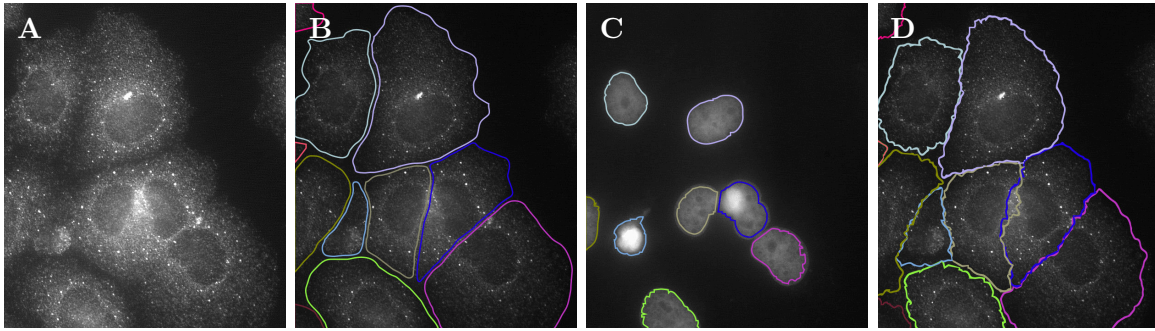
Figure 1.10: Image clip from the training data set used in [Möller et al., 2010] with results: PB channel (A, contrast enhanced), ground truth annotation (B), initial nuclei contours (C), and final segmentation result (D).

objects being simultaneously labeled, like SGs, PBs, or cell nuclei. However, reserving a separate channel for an explicit staining of cell membranes or cytoplasm to ease cell segmentation is often not desirable or not at all possible, and may impair the flexibility in experimental design. Hence, if cell regions can be extracted solely based on indirect cues from other available channels this yields a distinct advantage for the automatic analysis of such image data in practice.

We worked out a solution for this in [**Möller et al., 2011b**] and [**Möller et al., 2010**], applying snakes to segment cells from channels with fluorescently labeled PBs (Fig. 1.10). In these channels the cells exhibit a considerable inhomogeneity in their intensities, i.e., they appear grained with a gradual decrease in brightness from the centers outwards, and without obvious discontinuities along their boundaries. To deal with these issues, we proposed a novel cascaded optimization scheme for snakes by which the contours of all cells in an image are simultaneously and incrementally expanded from nuclei (Fig. 1.10 C) to complete cell areas (Fig. 1.10 D). The underlying idea is to gradually adjust the contours to the changing intensity characteristics of the cell tissue over several iterative runs and to get stepwise closer to the cell boundaries. The segmentation in each run is initialized from dilated results of the previous stage. As energy functional we adopted the region-based Chan-Vese energy [Chan & Vese, 2001] supplemented with a coupling term to penalize snake overlap [Zimmer & Olivo-Marin, 2005].

In our initial publication on this approach [**Möller et al., 2011b**] segmentation was performed applying a cascaded optimization over three levels. This hard-coded scheme was replaced in [**Möller et al., 2010**] by a data-driven optimization schedule which for each cell automatically determines the appropriate number of levels, exploiting changes in intensity distributions and sizes of cell regions between subsequent levels as termination criteria. In addition, parts of a cell segmented in former levels were explicitly masked to drive snake expansion.

Based on test data sets with several dozens of cells and given ground truth annotations, in [**Möller et al., 2011b**] and [**Möller et al., 2010**] we could demonstrate that our approach properly handles grained cell tissue and local intensity variations. In both works, we achieved median recall and precision values on the cell pixels of about 0.9 – noting that the definition of ground truth contours is often a challenge in itself even for experts from the field (cf. Fig. 1.10 B). Altogether, our results indicate that for complex segmentation tasks where the design of custom-fit energy functionals is difficult, extended optimization schemes may still yield a suitable starting point to adopt active contour models for solving such tasks.

Snake-based neuron segmentation. Neurons define a special class of cells in higher organisms [Levitan & Kaczmarek, 2015]. The most striking difference compared to other cell types is their outstanding morphology which is linked to the ability to develop long elongated protrusions,

termed neurites (Fig. 1.11). Neurons establish contacts with each other via these neurites to form complex and spacious networks [Yuste, 2015; Fornito et al., 2016].

Neurons have been in the focus of research for a long time, and data about their morphology, like numbers and lengths of neurites or spines, are of high relevance to discover their functions and behaviors [Meijering, 2010]. To collect such data, it is usually sufficient to reduce neurons to a skeleton representation which can be extracted from images without need for an accurate localization of the complete neuron region.

Complementary to approaching neurons by observing their morphology, the subcellular level of proteins and their distributions within neurons – and particularly along neurites – provides another rich source of information. Correlating neurite morphology with spatial distributions of specific proteins adds to the understanding of regulatory processes underlying neuron formation and remodeling. To extract such distribution data from images, however, neurite centerlines are not sufficient and segmentation of the complete cell areas is imperative.
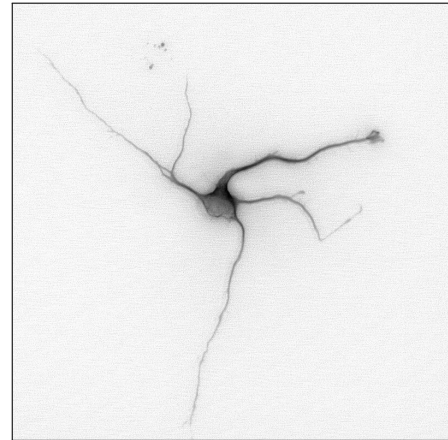
Figure 1.11: Image of a neuron with several neurites (inverted, contrast enhanced).

In our publication [**Misiak et al., 2014**] we were among the first to automatically quantify distributions of proteins along neurites. We approached this task with a fully automatic analysis pipeline using snakes as one of its central components. Neurons are localized making use of two different binarization algorithms. A global Niblack thresholding [Niblack, 1985, pp. 115] yields a rather rough, but mostly connected representation of a neuron, while an Otsu thresholding [Otsu, 1979] provides a much more accurate localization of the neuron boundaries, however, at the price of stronger fragmentation. Accordingly, we use the coarse binarization as initialization and optimize the snakes applying a distance-based energy derived from the Otsu binarization result. Supplemented by additional heuristics to automatically identify different structural parts of a neuron, i.e., the soma and the growth cones of each neurite, we automatically extract profiles of certain proteins as basis for analyzing their spatial localization.

In summary, our approach showed superior performance in segmenting complete neuron areas compared to two other tools which are among the few that also provide such a functionality, proving that active contour models are able to extract even thin and filigree structures in images. The high quality of our segmentation results was also underlined by extracted profiles of proteins along the neurites (Fig. 1.12). We successfully confirmed distributions from literature for filamentous actin (F-actin), which is known to contribute to protrusions and mainly localized in the growth cones, and α-tubulin, which is enriched along neurite

Figure 1.12: 3D heatmap visualization of the spatial distribution of a specific protein within the segmented region of the neuron shown in Fig. 1.11.

shafts. In addition and most important, however, we could for the first time obtain a distinct distribution for the Zipcode binding protein (ZBP1/IGF2BP1) which is assumed to fulfill important functions in the spatio-temporal fine-tuning of gene expression [**Misiak et al., 2014**].

1.2.3 Deep learning and morphological analysis to unravel root topology

Image segmentation with techniques of deep learning is based on the idea to learn parameters of a mathematical model which is able to map given input images onto desired segmentation results. One of the keys to the outstanding performance of these approaches is given by the complexity of the models used, which basically rely on concatenations of a huge number of non-linear data transformations [Goodfellow et al., 2016]. This causes them to be defined over extremely high-dimensional parameter spaces. In consequence, estimating optimal parameters for such models requires large amounts of training data which may pose obstacles to their use in practice. If sufficient data are available, however, especially convolutional neural networks (CNNs) have often superseded established approaches and pushed the performance in solving image analysis problems to new levels being out of reach before [Gu et al., 2018].

Adopting CNNs to solve a certain task requires to make several elementary decisions, e.g., regarding appropriate model architectures, loss functions, or optimization strategies [Breuel, 2015; Mishkin et al., 2017]. Due to a lack in general thumb rules, these decisions often rely on experience or time-consuming trial and error, where only a small subset of the virtually unlimited universe of possible configurations is explored. Henc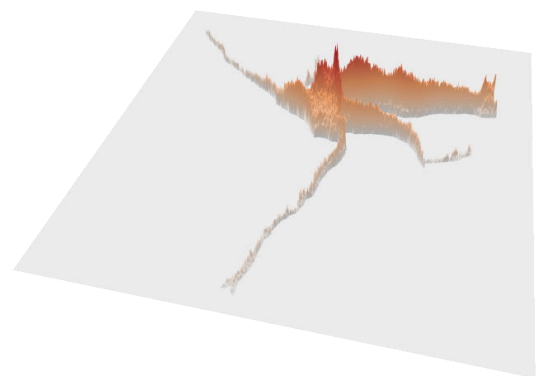e, comparative studies on the characteristics of different network architectures and hyperparameters in a certain application domain may provide valuable information and assistance to make the right decisions.

We carried out such studies with regard to the task of segmenting plant roots in images for root phenotyping. In this area of application, methods of deep learning have gained increasing importance in the last years. Roots as one of the essential organs of plants exhibit manifold functions with regard to the plants themselves, but also in the context of the surrounding ecosystem [de Kroon et al., 2003]. Thus, they have been in the focus of research for a long time. One key approach in studying root functions are phenotypical analyses of root systems under varying environmental conditions. These analyses nowadays often rely on image data, e.g., acquired with the help of minirhizotrons directly in the soil (cf. Subsec. 1.4.4) or from plates with culture medium [Atkinson et al., 2019]. Though the manual extraction of phenotypical data from such images is still common, over the years various mainly semi-automatic software tools for root segmentation have been released (e.g., [Lobet et al., 2011; Shahzad et al., 2018; Narisetti et al., 2019]). Recently, these are increasingly supplemented with automatic methods based on deep learning (e.g., [Wang et al., 2019; Yasrab et al., 2019; Gaggion et al., 2021]).

We addressed the problem of segmenting roots of *A. thaliana* from images of square plates in the context of the Root Segmentation Challenge at CVPPA[1] 2021. Given video sequences of plants covering complete growth periods, and ground truth annotations only for selected frames [Gaggion et al., 2021], the task was to segment the roots from the images and to identify the main root of each plant and separate it from the lateral roots (Fig. 1.13).

We solved the task as described in [**Möller et al., 2021**] by first separating the roots from the background adopting a CNN with U-Net architecture [Ronneberger et al., 2015] and VGG16 backbone [Simonyan & Zisserman, 2015]. Then segmented root regions are skeletonized, gaps closed and disconnected branches reconnected. Finally, a graph search modified from the one proposed in [**Möller & Bürstenbinder, 2019**] (Subsec. 1.2.1) for closing gaps in cell contours is applied to localize the main root of each plant. It is extracted as the path between the lowest pixel in each plant component and the hypocotyl. This procedure allowed us to achieve completeness and correctness values on the main root pixels of the challenge test data set of over 0.91 and 0.95, respectively, and likewise respectable values on the complete root

---

[1] CVPPA 2021 workshop website, `https://cvppa2021.github.io/` (accessed: September 10, 2021)

systems of 0.89 and 0.95. Since we were – among the 16 teams that registered for the challenge – the only team submitting complete results on the test data set, we were declared the winner of the challenge.

Our U-Net model was trained in a cascaded fashion first using a loss function combining cross entropy and dice loss adopted from [Smith et al., 2020], and subsequently fine-tuning the network applying focal loss [Lin et al., 2017]. The choice of this strategy resulted from studies on different alternative network architectures, like Hi-Fi [Zhao et al., 2018] or SegRoot [Wang et al., 2019], on additional loss functions, like the Weighted Hausdorff Distance in combination with patch-based point loss [Xu et al., 2019], and on a range of learning rates. The studies were conducted prior to the challenge deadline and

Figure 1.13: Sample data from the CVPPA Root Segmentation Challenge test data set [Gaggion et al., 2021]: input image (left) and our segmentation result with main and lateral roots in red and gray, respectively (right).

rounded off afterwards, and are detailed in our workshop paper [**Möller et al., 2021**]. Amongst others, our studies revealed that a thorough mutual adjustment of loss function, normalization strategy and learning rate is key to successfully solve a segmentation task with CNNs. Moreover, the findings point out potential pitfalls and promising starting points for further research to improve the overall performance of CNNs in the segmentation of roots in images.

1.3 Analyzing Shape and Texture of Objects

The detection of low-level image features like points or the segmentation of objects, which we discussed in previous sections, are most of the time only initial steps in more complex pipelines. Image primitives and objects usually form the fundament for the extraction of further quantitative data from bioimages. In the simplest case, specific entities are just counted or their sizes are measured, like in our work on the appearance of stress granules and processing bodies in cells (Subsec. 1.1.2). Other research questions may require to assess spatial intensity distributions or texture properties (cf. Subsec. 1.2.2), characterize shape, or monitor object dynamics [Rittscher, 2010]. As there exists an enormous variety of object properties potentially of interest, the large universe of different feature measures proposed in this and other contexts is hardly to be overviewed (e.g., [Merchant et al., 2008; Li et al., 2015]). Hence, careful consideration is required before existing measures are chosen or new ones are devised to address an analysis task. In this section, we will first present our solutions to quantify cell shape in the context of *PaCeQuant* which we introduced in Subsec. 1.2.1, and second we will discuss our approaches for carrying out comparative studies on internal cell organization with texture measures.

1.3.1 Shape quantification and analysis

Leaf epidermis pavement cells pose a special challenge to characterize and quantify their shapes due to their striking morphology which is dominated by their prominent lobes and necks (Fig. 1.14, cf. Subsec. 1.2.1). Beyond that, the task to adequately quantify these cell shapes gets even more challenging as the shapes of the cells and the degree of anisotropic cell growth

may vary significantly and unpredictably in different experimental situations. Yet, often only few and rather generic shape features are adopted to quantify the shape of pavement cells [Staff et al., 2012], only to a limited extend capable to capture the whole complexity of the shapes [Ivakov & Persson, 2013; Wu et al., 2016]. To still consider the specific lobe and neck structures in quantification, thus, generic shape measures are often supplemented by data resulting from manual counts of lobes per cell area [Xu et al., 2010; Gao et al., 2015].

However, to adequately and objectively characterize the shapes of pavement cells, each experiment may rather require an individual and specifically selected set of automatically extracted features. These need to account for the specific characteristics of pavement cell shapes as well as for the large range of potential variation.

Accordingly, to support a wide range of possible application scenarios out-of-the-box, we offer a large variety of different shape measures in *PaCeQuant*. They allow for a straightforward interpretation and correlation with phenotypic specifities, ranging from general and established global shape measures, over contour- and skeleton-based fea-

Figure 1.14: Examples for lobe (red) and neck (green) regions along the boundary of a pavement cell, blue circles mark three-cell contact points.

tures, to a collection of novel measures explicitly designed for the specific characteristics of pavement cells [**Möller et al., 2017**]. For the last ones we were the first to introduce formal definitions of commonly considered, but never properly defined structural measures, like the neck width or the cellular core region [Li et al., 2003; Bannigan & Baskin, 2005]. Providing such definitions paves the way towards automatic quantification of these structural properties from images, resulting in larger flexibility to characterize cell shape.

Initially *PaCeQuant* offered a set of 27 shape features for each single cell, later complemented by the largest empty circle measure [**Poeschl et al., 2020**]. Another eight measures are provided to characterize each individual lobe [**Möller et al., 2017**]. Noteworthy among these lobe-specific features is the lobe type which principally allows to classify each lobe in one of two possible categories (Fig. 1.14): lobes of type I refer to two-cell contact points, i.e., are flanked by only one neighboring cell, while lobes of type II are associated with three-cell contact points being adjacent to two other cells [Wu et al., 2016]. *PaCeQuant* has been the first tool which allowed to categorize lobes in that way, based on its unique ability to automatically segment complete cell conglomerates (Subsec. 1.2.1). This opens perspectives towards more precise and target-oriented studies on lobe type specific-effects in cell morphogenesis as shown in our own work [**Mitra et al., 2019**], and as taken up also by others (e.g., [Nowak et al., 2021]).

The set of shape features offered by *PaCeQuant* in conjunction with the high accuracy of segmented cell contours (Subsec. 1.2.1) yields a profound fundament to precisely quantify shape characteristics. This was shown on sample data evaluated in [**Möller et al., 2017**] where we could successfully confirm earlier findings from literature, e.g., with regard to leaf development of *A. thaliana* cotyledons, or concerning shape alterations in *A. thaliana* mutants impaired in Katanin function, which is a protein known to hamper lobe growth (cf. Fig. 1.6). Finally, we could approve different growth properties of type I and type II lobes in cell expansion.

In our work [**Mitra et al., 2019**] *PaCeQuant* served as basis for novel discoveries regarding the role of microtubule-associated proteins (MAPs) in cell morphogenesis. Specifically the member IQD5 of the IQ67 DOMAIN protein family was investigated. The family was previ-
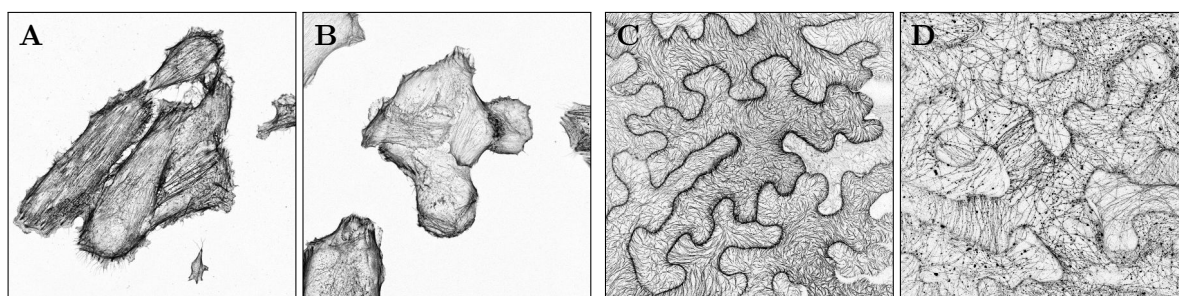
Figure 1.15: (Inverted and contrast enhanced) images with stained cytoskeleton structures: cells from the control (A) and knock-down (B) populations analyzed in [Möller et al., 2014], and samples with patterns of microtubule localization for the IQD13 (C) and IQD5 (D) family members from [Bürstenbinder et al., 2017].

ously identified as one of the largest classes of known MAPs in plants [Abel et al., 2005] (see also Subsec. 1.3.2), and IQD5 was identified to potentially play a special role in this family (Fig. 1.16) [**Bürstenbinder et al., 2017**]. Different studies were conducted to characterize IQD5 in terms of expression domains, subcellular localization, and biological roles. *PaCeQuant* was successfully adopted in these studies to quantify differences in cell shape between wildtype and mutants defective in IQD5. Besides, also the cell contours themselves provided by *PaCeQuant* contributed to novel findings with regard to different organic compounds being essential ingredients for cell wall expansion. *PaCeQuant's* contours formed the basis for automatically quantifying these compounds along the cell walls from local intensities.

1.3.2 Comparative analysis of texture

The cytoskeleton is the central backbone of a cell. It is formed by a complex system of different types of protein filaments in the cytoplasm and defines essential properties of the cell, like shape, mechanical properties, and its overall internal organization [Alberts et al., 2015, Chap. 16]. The cytoskeleton is highly dynamic. Based on dynamic self-organization and steady interactions with a multitude of proteins, which are members of various signaling pathways in the cell, parts of the cytoskeleton can rapidly be disassembled or newly generated, or the formation and localization of segments can be modified. Thus, the filaments of the cytoskeleton are capable of exhibiting various higher-order structures and network configurations, e.g., depending on the current state of the cellular environment [Nédélec et al., 2003; Letort et al., 2015]. Analyzing structural properties of the cytoskeleton and its conformational changes under varying conditions allows to uncover functional roles of proteins and signaling molecules in the regulation of the cytoskeleton and, hence, in its diverse cellular functions.

The filaments of a cytoskeleton can be distinguished into different types, amongst others into actin filaments and microtubules [Alberts et al., 2015]. Depending on the research question, different fluorescence staining protocols exist to label and visualize the target filaments in microscope images. In Fig. 1.15 sample images of stained cytoskeleton structures are shown. Based on such images, it is possible to extract data about the specific characteristics of filaments, e.g., for quantifying the specifities of a cytoskeleton or for structural comparisons.

The analysis of filament structures in images is often focused on an accurate measurement of fibers, e.g., to quantify their lengths, widths, or orientations [Matschegewski et al., 2012; Kimori et al., 2016]. This naturally requires to identify individual fibers in images which is often rather challenging. Particularly in cases of densely packed fibers with many crossing overs, it is even hard for human experts to solve this task. An automatic analysis is often biased towards

filaments which are easy to segment, but not necessarily representative. Moreover, in some experiments cytoskeleton structures are even completely destroyed or severely fragmented, rendering filaments in general no longer the dominating structure representative for the status of the cytoskeleton, and neither a suitable basis for quantifying skeleton characteristics.

In many applications the underlying research questions do not necessarily require to measure single filaments, though this seems to be the evident solution. E.g., for a global comparison of cytoskeleton properties and the search for structural similarities and differences, it is sufficient to quantify the overall appearance of the cytoskeleton rather than of individual substructures. We have elaborated this idea in [**Möller et al., 2014**] and [**Bürstenbinder et al., 2017**]. In addition, our approach has been invited for presentation in the book chapter [**Möller et al., 2019c**]. This chapter is dedicated to end users describing the usage of our tool collection which subsumes the implementation of our approach in the *CytoskeletonAnalyzer2D* as well as the *CellBoundaryExtractor2D* which implements our semi-automatic approach for cell boundary extraction from [**Möller & Bürstenbinder, 2019**] (cf. Subsec. 1.2.1).

The key idea of our method is to adopt texture measures [Xie & Mirmehdi, 2008a] to characterize the structural properties of stained cytoskeleton components. Initially, the region of each cell is subdivided into a collection of non-overlapping tiles, and for each tile a feature vector with texture measures is extracted. This results for each cell of a data set in a collection of feature vectors which locally characterize the structural appearance of the cytoskeleton within the cell. Subsequently, clustering is applied to all feature vectors over all cells to identify a set of basic and representative types of structural patterns commonly appearing in the data set. Since each feature vector is uniquely assigned to a cluster, it is associated with one of these types. Hence, for each cell a distribution of the basic structural patterns can be extracted to which its feature vectors are linked to. This essentially forms a structural fingerprint for the cytoskeleton of the cell, yielding a suitable basis for distinct comparative analyses.

In [**Möller et al., 2014**] we applied our method to the analysis of filamentous actin (F-actin) which is one form of appearance of the actin cytoskeleton in cells [Alberts et al., 2015]. Here, we adopted Haralick measures [Haralick et al., 1973] to quantify the textural appearance of the cytoskeleton. We analyzed a set of 92 human melanoma-derived HT144 cells subsuming a control population of 35 cells and a knock-down population of 57 cells where a protein was depleted known to control F-actin organization and causing degenerative changes in the cytoskeleton. Comparative analysis was performed by hierarchical clustering of the structural fingerprint vectors of all cells. In doing so, we could demonstrate the suitability of our approach to separate the two categories of cells to a satisfying degree. On the top level of the cluster hierarchy two clusters emerged, one containing almost 90% of the knock-down cells, while approximately two thirds of the control cells were gathered in the second cluster. Given a certain degree of natural variation among the cells, differences in the responses to stimuli, and also considering that cell division events which strongly affect actin organization occur spontaneously in 10 to 20% of the cells, this result is conclusive.

In [**Bürstenbinder et al., 2017**] we applied our approach to comprehensively characterize members of the IQ67 DOMAIN (IQD) family of microtubule-associated proteins (MAPs). The Arabidopsis IQD family consists of 33 members which can be allocated to four different phylogenetic groups differing in their organization of conserved motifs [Abel et al., 2005]. The microtubule cytoskeleton is central for cell growth, and MAPs function in regulating microtubule dynamics, stability and organization.

23 members of the IQD family localize to the microtubule skeleton and strongly label microtubules when expressed in leaves of *N. benthamiana*. Notably, the patterns of localization (and possibly also the microtubule cytoskeletons themselves) differ between the individual family members which may, e.g., provide cues for their individual functional roles. Accordingly, we analyzed these 23 members most strongly associated with the microtubule cytoskeleton, and compared their structural patterns of microtubule localization using our approach. This time we adopted local binary patterns [Ojala et al., 2002] as structural measures and used the structural fingerprint vectors of the cells to evaluate group-wise similarities between the different members

Figure 1.16: Network of the 23 IQD67 DOMAIN protein family members most strongly associated with the microtubule cytoskeleton, based on structural similarity of localization patterns (adopted from Fig. 5 in [Bürstenbinder et al., 2017]). The node colors refer to the four different phylogenetic groups which can be identified within the IQD family.

of the IQD family. Amongst others we could extract a similarity network showing large congruence with the phylogenetic groups defined for the IQD family (Fig. 1.16, node colors). This gives raise to the assumption that closely related family members share similar functions in regulating the microtubule cytoskeleton. As remarkable outliers IQD11 and IQD16 were identified which suggests unique, but similar functions of IQD11 and IQD16 compared to other IQD members. This could be confirmed in supplemental experiments. Interestingly, IQD5 also differed from other members of the corresponding phylogenetic group, which gave rise to a more detailed functional characterization of IQD5 in [**Mitra et al., 2019**] (see above).

Altogether, our approach for cytoskeleton analysis contributed to the first comprehensive characterization of the complete IQ67 DOMAIN protein family. Our structure characterization based on local texture measures turned out to be well-suited for distinguishing different types of patterns, revealing flexible alternative directions for the analysis of subcellular structures, particularly in cases where filament structures are either difficult to segment or absent at all.

1.4 Software Development and Tools for Bioimage Analysis

Computational solutions for bioimage analysis and their development process are expected to meet a variety of requirements. First of all, there is the need for custom-fit algorithms tuned to optimally exploit given research data. Beyond that, to make new approaches available to end users these need to be integrated into user-friendly tools, but without loosing the flexibility to promptly react on shifts in research foci or novel imaging technologies with timely adaptations. Carpenter et al. [2012] outline a comprehensive collection of criteria which should be aimed at in developing bioimage analysis software with a prospective high value for developers, users, and the whole bioimaging community. These criteria subsume user-friendliness by easy accessibility, documentation and support, developer-friendliness via established routines for source
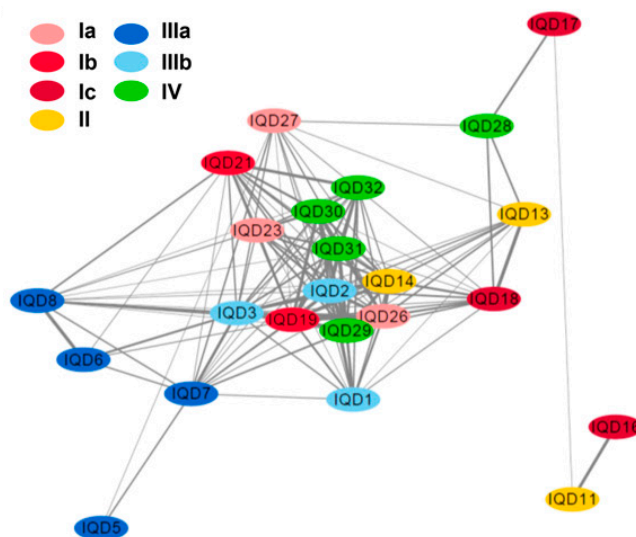
code organization and maintenance, modularity, interoperability, straight extensibility, and easy adaptability to future needs. One key paradigm generally considered important for bioimage analysis software is its public availability in an open-source fashion [Swedlow & Eliceiri, 2009; Cardona & Tomancak, 2012]. Open software tools, which can intuitively be applied also by non-computer scientist users, inherently add to transparent and reproducible research. As everyone can (re)use the software and contribute extensions, it encourages scientific collaboration [Schindelin et al., 2015].

As already became apparent in previous sections, we developed a considerable number of basic to elaborate algorithms and workflows in the course of our research over the years. We followed open-source principles from the beginning and released the majority of our approaches to the public applying open-source licenses. In this section, we present several of our software projects and tools in detail, starting in Subsec. 1.4.1 with *MiToBo*. This is our central library and toolbox yielding the basis for most of our developments, and collecting the majority of our tools for microscope image analysis. The core of *MiToBo* is formed by *Alida* which we introduce in Subsec. 1.4.2. *Alida* not only provides a powerful library, but constitutes an integrated concept for software development in data analysis. It meets many of the desirable criteria for bioimage analysis software outlined above and can be seen as the guide line along which we developed our algorithms and tools. The rich feature collection offered by *Alida* and *MiToBo* in daily research is illustrated in Subsec. 1.4.3 based on the sample applications *MiCA* [**Möller & Posch, 2011**] and *PaCeQuant* with its supplemental tool collection [**Poeschl et al., 2020**].

While we developed *Alida* and *MiToBo* from scratch, *rhizoTrak*, which is our tool for annotating and analyzing images of plant roots (Subsec. 1.4.4), is derived from the existing open-source software TrakEM2 [Cardona et al., 2012] originally dedicated to neuroscience research.

1.4.1 MiToBo – a library and toolbox for bioimage analysis

When our group entered the field of bioimage analysis more than ten years ago, one of the dominating software tools in this area was ImageJ[2]. Its roots date back to the second half of the 1990s where the first versions of ImageJ were released based on the fundament of its predecessor NIH Image [Schneider et al., 2012]. ImageJ has always been oriented towards maximum benefit for end users and the community, following a strict open source and public domain paradigm as stated among the basic concepts on its website[3]. It aims at easy extensibility which is particularly reflected in a flexible plugin mechanism for integration of new functionality.

Over the decades a huge and very active community of users and developers emerged around ImageJ and its main developer Wayne S. Rasband [Schneider et al., 2012]. This led to a continuous expansion of ImageJ's core functionality [Levet et al., 2021, Tab. 1], the release of countless extensions for manifold (bio)image analysis tasks provided by the community, and also the emergence of special and extended variants of ImageJ like Fiji [Schindelin et al., 2012].

While this substantiates ImageJ's status as one of the most prominent and widely used software tools for bioimage analysis today, limitations and design issues in its software architecture became more and more obvious over the years [Rueden et al., 2017]. The user-centered development process with a clear focus on the use of ImageJ via its graphical user interface (GUI) and the need to implement a separate graphical front end for each plugin, encouraged tight interlocks of functionality and GUI even in ImageJ's core. This significantly hampers the reuse of ImageJ's huge repertoire of methods in other contexts, e.g., headless or as library.

---

[2]ImageJ website, `https://imagej.nih.gov/ij/` (accessed: November 28, 2021)

[3]Basic concepts of ImageJ: `https://imagej.nih.gov/ij/docs/concepts.html` (accessed: November 28, 2021)

ImageJ's enormous popularity and its extensibility via plugins might render the software a suitable foundation for our own developments. However, the fundamental deficiencies in its design collide seriously with several of our primary ideas for developing bioimage analysis software, like a high modularity and easy extensibility, and particularly a strict separation of functionality from user and developer interfaces. Accordingly, we decided to develop our own algorithms and workflows as separate library – but to preserve a maximum of compatibility with ImageJ to avoid loosing the benefits of ImageJ's extraordinary rich functionality and the eminent advantages that an established software offers by its prevalence and community.

As result *MiToBo*, our **Microscope Image Analysis ToolBox**, emerged. *MiToBo* [**Möller et al., 2016**] is a Java library comprising basic to advanced methods for image processing and analysis. As a specialty, *MiToBo* offers sophisticated mechanisms to access this functionality in different ways, i.e., via a unified API on the source code level, from the command line or in scripts, and via a generic plugin in ImageJ and Fiji. New functions and tools are automatically registered in *MiToBo* without additional programming effort and are immediately ready for execution via the different generically generated user interfaces.

The core functionality in *MiToBo* which is responsible for this flexibility in accessing the library internals is not specific to algorithm development for bioimage analysis. The underlying concepts and technologies can directly be transferred to almost every area of application where functions are used to manipulate data. Consequently, we implemented the basic concepts of these mechanisms in a separate library termed *Alida* which is independent of any concrete application domain. The concepts of *Alida* are further elaborated in Subsec. 1.4.2.

MiToBo's and *Alida's* developments started in parallel with similar efforts for overcoming the design shortcomings of ImageJ in its closer periphery. Especially the ImageJ2 project [Rueden et al., 2017] in parts pursued similar goals of decoupling functionality and interfaces. Although ImageJ2 envisaged a full rewrite of ImageJ's core, focusing on modularity and interoperability, ImageJ2 and *MiToBo* and *Alida*, respectively, share many fundamental conceptual ideas, e.g., regarding the integration of extensible user and developer interfaces for accessing the functions within libraries, or concerning the generic execution of tools and plugins. Notably, the ImageJ2 project, from which later on the SciJava cooperation project[4] emerged, released its functional core in an image-independent fashion as library called SciJava Common in 2013[5].

Naturally, the developments of *MiToBo* and *Alida* from the beginning and still today mutually support and complement each other. While *Alida's* functionality steadily grows not least because of new requirements within *MiToBo*, vice versa *MiToBo* constitutes a perfect test case for *Alida's* concepts with regard to the specific area of (bio)image analysis. *MiToBo* has been and still is the fundament of most of our algorithmic developments for bioimage analysis. It formed the basis for various bachelor, master and PhD projects from which also new functions and extensions emerged. Over the years, *MiToBo* has grown to a comprehensive collection of basic to advanced image analysis methods and elaborate pipelines. *MiToBo*[6] is open source and publicly available on GitHub under GPL license[7]. It features its own update site[8] for easy installation in Fiji, resulting in a tight integration in the ImageJ ecosystem [Schindelin et al., 2015]. Since 2019 *MiToBo* is community partner in the Scientific Community Image Forum [Rueden et al., 2019]. Most of the tools that we implemented and released in the course of

---

[4]SciJava website, `https://scijava.org/` (accessed: November 23, 2021)
[5]scijava-common on GitHub, https://github.com/scijava/scijava-common/releases/tag/scijava-common-1.0.0
[6]MiToBo homepage, `https://mitobo.informatik.uni-halle.de` (accessed October 26, 2021)
[7]GNU General Public License 3, `http://www.gnu.org/licenses/gpl-3.0.de.html` (September 20, 2021)
[8]Update sites in Fiji/ImageJ2, `https://imagej.net/update-sites/` (accessed: November 23, 2021)

our research are part of *MiToBo*, e.g., *MiCA* [**Möller & Posch, 2011**], *NeuronAnalyzer2D* [**Misiak et al., 2014**], *PaCeQuant* [**Möller et al., 2017**], *CellBoundaryExtractor2D* and *CytoskeletonAnalyzer2D* [**Möller & Bürstenbinder, 2019; Möller et al., 2019c**] which are all described in this thesis, or the *ScratchAssayAnalyzer* [Glaß et al., 2011, 2012].

Finally, *MiToBo* has proven suitable also as software basis for lectures. We employ *MiToBo* in conjunction with ImageJ regularly in our introductory lecture on image processing where students learn the theoretic foundations of basic image processing methods, and implement and apply them in practice. *MiToBo* allows to keep the focus on the algorithmic side and practical experiments rather than enforcing additional workload on infrastructure or user interfaces.

1.4.2 The Alida concept for data analysis

The motivation for devising a novel concept on design and development of data analysis algorithms and tools was triggered by our aim to build modular and extensible software with a strict separation of functionality from interfaces. Another important intention was to facilitate an automatic documentation of data analysis procedures. Given an integrated logging of all actions applied to data during an analysis process, development and testing are simplified as algorithmic configurations are automatically recorded. Once a workflow is established, built-in logging capabilities allow for a gapless documentation of analysis and data generation procedures. This is an essential prerequisite, e.g., for later interpretation, validation and reproduction of results, or for long-term archival of scientific data [Goldberg et al., 2005].

Our new concept, which we denoted *Alida* as an acronym for **A**utomatic **L**ogging of Process **I**nformation in **D**ata **A**nalysis, was initially published in [**Möller et al., 2011a**]. It builds on the assumption that each data analysis process is formed by a series of analysis or manipulation actions applied to data items. The complete sequence of all these actions together with the data items processed or generated during analysis are summarized in the processing graph. It is implicitly defined by every data analysis pipeline. The nodes of the graph represent the actions on the data, and the edges represent the flow of data and control between them.

To account for a high modularity in designing data analysis pipelines based on these ideas and to enable automatic process documentation, every action on data needs to be conceived as self-contained step, controllable via a unified interface for configuration and execution. In *Alida* this is implemented using 'operators' as fundamental building blocks where all actions take place. Operators are enforced to provide all relevant data, i.e., input and output data items as well as configuration parameters and their current settings, through such a unified interface. In addition, *Alida* defines a unified execution procedure to generically invoke operators. In combination, this yields the basis for *Alida's* core functions, i.e., for automatic process documentation, but likewise for generic user interface generation and a visual programming editor, which we will outline with more details below.

The key idea in *Alida* to automatically document analysis procedures as presented in [**Möller et al., 2011a**] is to transform the implicit processing graph of a data analysis pipeline into an explicit representation. This requires to register all calls to operators on execution and store affected data items, settings of configuration parameters, and meta information, e.g., software versions of the program code used for data analysis. In *Alida's* concept with its unified interfaces and execution procedures, all these data are inherently available for recording in an internal database once an operator is invoked. By exporting the processing graph with all these data from the database to an explicit representation a full-featured documentation of the whole analysis pipeline is readily available.
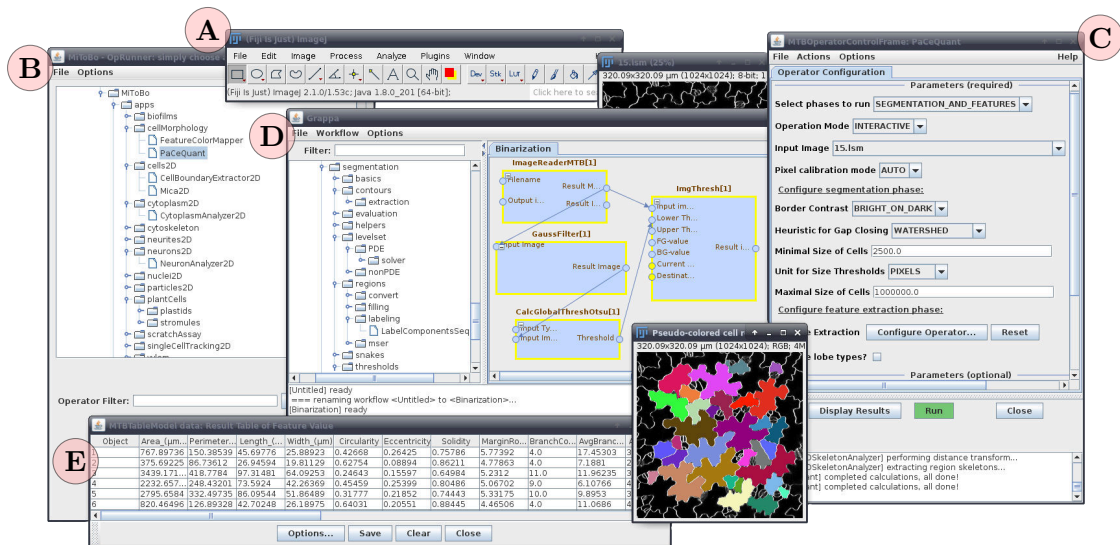
Figure 1.17: Screenshot of *MiToBo's* GUI in Fiji showing the main window of Fiji (A), the operator chooser window of *MiToBo's* graphical operator runner (B), *PaCeQuant's* operator configuration and control window (C), the main window of the workflow editor *Grappa* (D), image windows, and a table with result data (E).

Despite their unquestionable relevance, user interfaces (UIs) are often not in the primary focus of developers. Efforts to release suitable UIs are often not compensated by the scientific merits to gain, e.g., compared to tackling a challenging problem with new algorithmic solutions [Levet et al., 2021]. One option to overcome this issue is to relieve developers from explicit UI programming. *Alida's* conceptual basis is a natural starting point for automatic generation of UIs from source code in a generic fashion as we demonstrated in [**Möller & Posch, 2013**].

User interfaces basically need to bidirectionally transform input and output (I/O) data of an operator, i.e., data items to be analyzed, configuration parameter settings, and analysis results, between their internal and an explicit external representation, interpretable and editable by the user. To facilitate this in a generic fashion in *Alida*, we introduced a flexible and dynamically extendable mechanism of I/O providers which allows to transform manifold kinds of data between different representations. Besides primitive and corresponding wrapper data types, enumerations, arrays, and collections, also operators by themselves are out-of-the-box supported as input parameters of other operators. Given the specification of input and output parameters of an operator, this provider mechanism forms the basis for generic execution of the operator via a command line interface (CLI) or a graphical user interface (GUI). Accordingly, *Alida* ships with a command line runner and a graphical operator runner. The first one offers sophisticated functionality for parsing data items and parameters from command line [Möller & Posch, 2021, Sec. 2.5] and yields the basis for running *Alida* operators from scripts in a high-throughput fashion. The latter one lists available operators in a graphical overview from where they can be selected, configured and executed via automatically generated GUI components. This runner also forms the basis for *MiToBo's* ImageJ plugin that constitutes the entry point for running *MiToBo* operators in a generic fashion in ImageJ and Fiji (Fig. 1.17).

Image analysis problems are usually not solved by applying a single operator, but require the combination of different analysis operations into sophisticated pipelines. The concept of pipelines and workflows is central to many data analysis tools, like for the data analytics platform KNIME [Berthold et al., 2009], Galaxy [Afgan et al., 2018], or for tools like ICY [De Chaumont et al., 2012] or CellProfiler [McQuin et al., 2018] specialized on bioimage data

analysis. In *Alida* such pipelines correspond to a special type of operators that define their inputs, outputs and parameters like usual operators, but setup their functionality on other operators. Hence, the extension of *Alida's* concept towards support for workflows in [**Möller & Posch, 2013**] is straightforward. Moreover, by interpreting the process of solving an image analysis problem as the search for a workflow which combines operators into a suitable graph structure, the design process can be translated into a graph editing task. This led to the development of the graphical workflow editor *Grappa* which allows to interactively arrange operators in workflows. An early version of *Grappa* was initially developed by S. Kirchner in his bachelor project supervised by the author of this thesis [Kirchner, 2012]. This was later extended and released to the public in [Kirchner et al., 2012] and [**Möller & Posch, 2013**].

On the technical side, the *Alida* concept is implemented as library in Java following object-oriented design principles [**Möller et al., 2011a; Posch & Möller, 2017**]. For dynamic handling of available operators and I/O providers, and for specifying configuration parameters of operators, we adopt Java's annotation mechanism. All GUI components rely on Java Swing, and *Grappa* is based on the JGraphX library for visual graph editing[9]. The processing graphs are exported to XML representations where we extend the schema descriptions of *graphML* [Brandes et al., 2002] for our needs and provide the tool *Chipory* for visual exploration of the graphs which extends the free graph editing and layout tool Chisio [Kucukkececi et al., 2017].

Today *Alida*[10], which in the meantime changed its acronym to **A**dvanced **L**ibrary for **I**ntegrated **D**evelopment of Data Analysis **A**pplications, reached a mature status. New features are integrated diligently, mainly triggered by needs arising from *Alida's* role as integral core of *MiToBo*. Over the years, the special type of controllable operators was introduced which support interactions during execution, i.e., can be controlled via 'Run', 'Pause' and 'Stop' buttons if executed from a GUI. In addition, a batch mode for operators was added, and an online help which automatically extracts its contents from annotations in operator classes. With regard to generic GUI generation, the possibility to add and remove operator parameters dynamically during run time and define dependencies between parameters via callback functions opened new perspectives for slim GUIs showing only the parameters relevant in the current context. This mechanism is for example used to keep *PaCeQuant's* GUI clear despite its various configuration options (Subsec. 1.4.3). Finally, an infrastructure around collections of *MiToBo* operators was established that simplifies their reuse in external programs in a unified manner and from which, e.g., the *MTBCellCounter* [Franke et al., 2015] or *rhizoTrak* (Subsec. 1.4.4) benefit.

1.4.3 Alida and MiToBo in practice: MiCA and PaCeQuant

Alida's central aim to enable easy reuse of functionality and to allow developers to concentrate on algorithm design rather than infrastructure, while still offering intuitive interfaces to end users, significantly reduces developer workload. Prominent examples to illustrate this in practice are *MiCA*, the **Mi**ToBo **C**ell **I**mage **A**nalyzer [**Möller & Posch, 2011**], and *PaCeQuant* [**Möller et al., 2017; Poeschl et al., 2020**].

We released *MiCA* as one of the first more complex operators in *MiToBo*. It offers an integrated analysis of fluorescently labeled subcellular structures on the single cell level. For the detection of subcellular structures, *MiCA* integrates the scale-adaptive particle detection approach introduced in Subsec. 1.1.2, while active contour models are adopted for cell boundary extraction (cf. Subsec. 1.2.2). To support users in combining different energy functionals to solve

---

[9]JGraphX project repository, `https://github.com/jgraph/jgraphx` (accessed: September 21, 2021)

[10]Alida homepage, `https://alida.informatik.uni-halle.de` (accessed: October 17, 2021)

a segmentation task, *MiCA* implements an intuitive normalization scheme for active contour energies [**Möller & Posch, 2011**]. The contours can, e.g., be initialized automatically from DAPI-stained nuclei applying thresholding and morphological operations. If input images lack a DAPI channel, subcellular structures can still be quantified for complete images and later on be combined with manual cell annotations like in [**Bley et al., 2015**].

The algorithms for cell segmentation with active contours as well as for the nuclei and particle detection are implemented as individual operators in *MiCA*. They are transparently integrated in *MiCA's* GUI based on *Alida's* support for using operators as parameters. Each operator can directly be configured via its own configuration window (Fig. 1.18). For the snake-based cell segmentation, operators with different optimization strategies are provided, and flexible combinations of energy functionals can be used and individually

Figure 1.18: *MiCA's* GUI in *MiToBo*, generated fully automatically, with main window (A) and nested configuration windows for the integrated cell segmentation via snakes (B, C), and for the snake energy functional (D) which allows to define a weighted sum of different energy components.

configured (Fig. 1.18 D). The sets of available active contour operators and energy functionals are dynamically managed in an automatic fashion based on *Alida's* mechanism to index all currently available classes derived from a common super class.

MiCA illustrates the large flexibility of *Alida* to handle complex and nested types of configuration parameters in GUI generation, but the overall number of parameters is moderate. In contrast, *PaCeQuant* offers a significantly larger number of parameters [**Möller et al., 2019b**]. To keep *PaCeQuant's* GUI clear, we make use of *Alida's* functionality to dynamically change the interface of an operator and adapt the GUI to current configuration settings. Depending on the selected phases of segmentation and/or feature extraction which are to be executed (Fig. 1.17 C, top parameter), we add or remove configuration parameters associated with the different phases via callbacks, resulting in dynamic changes of the GUI. Fig. 1.17 C shows *PaCeQuant's* configuration and control window including parameters for both phases of segmentation and feature extraction. If the selection of phases is toggled to run only one of them, the parameter section of the phase which is no longer selected is automatically hidden.

The explicit separation of the two phases of cell segmentation and feature extraction in the core operator of *PaCeQuant* enables intermediate user intervention, e.g., if cell segmentation locally fails or cell contours are externally provided and only shape features are to be extracted. To support users in correcting inaccurate segmentations, and also to ease the evaluation of extracted shape features, the core operator was over time supplemented with additional tools forming an integrated toolset for cell shape analysis[11]. Beyond an R package for statistical
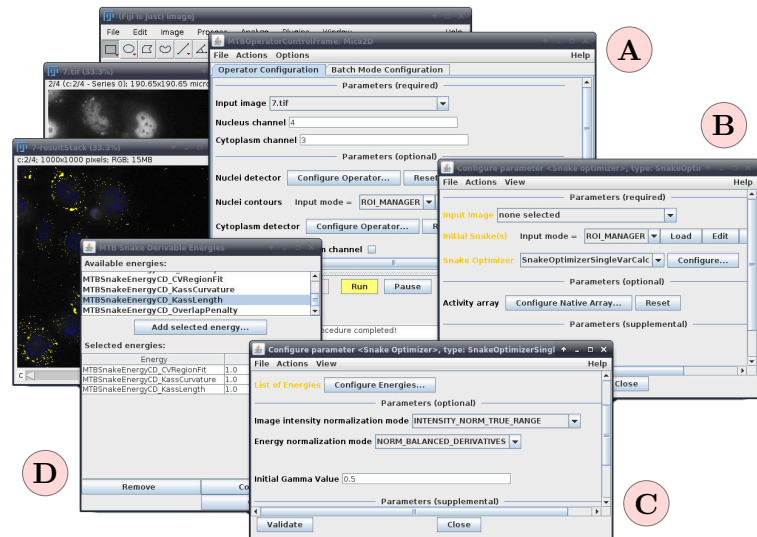
---

[11]Website of PaCeQuant and related tools, accessed: October 26, 2021,
  https://mitobo.informatik.uni-halle.de/index.php/Applications/PaCeQuantToolset

investigations on shape features, we offer two additional tools in *MiToBo* [**Poeschl et al., 2020**]. The *LabelImageEditor* is an interactive tool for manual post-processing of segmentation results. It allows, e.g., to remove surplus contours by merging adjacent regions via simple point-and-click actions or to remove intrusions via a constraint expansion of single regions which stops as soon as other regions are touched. The *FeatureColorMapper* allows to color cell regions according to the values of extracted shape features. The implementation of both tools as *MiToBo* operators enables their full integration into the *MiToBo* ImageJ/Fiji plugin and allows to directly release them to end users as part of the *MiToBo* distribution rolled out via its ImageJ update site.

1.4.4 Annotation and analysis of root images with rhizoTrak

The main area of application of our tool *rhizoTrak*[12] which we published in [**Möller et al., 2019a**] are studies on root growth and development in plant research. Roots are among the most important, yet least accessible organs of plants. To monitor roots *in situ* in long-term developmental studies, acquiring time-series image data with minirhizotrons is an established technique [Smit et al., 2000]. Minirhizotrons are transparent tubes inserted in the soil next to a plant which enable repeated scans of their immediate surroundings with scanners or cameras. From the resulting images (see Fig. 1.19, left, for an example) the number and distribution of root fragments as well as data about root mass and status can be extracted.

An automatic segmentation of roots from minirhizotron images is challenging due to the inhomogeneous image background, e.g., formed by soil, air bubbles, water drops, or earthworm corridors. Several approaches for semi- or fully automatic segmentation have been proposed, recently also adopting techniques of deep learning [Wang et al., 2019; Smith et al., 2020] (cf. Subsec. 1.2.3). However, the reliability of automatic approaches
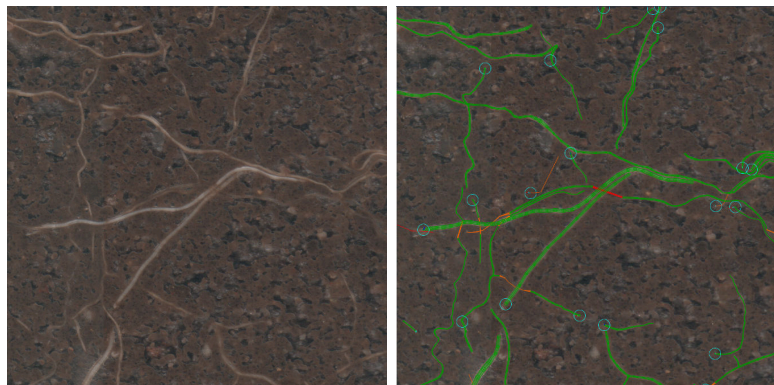
Figure 1.19: Crop of a minirhizotron image (left) with annotations exported from *rhizoTrak* (right), colors refer to root status, e.g., green to living and red to dead segments (image courtesy of A. Weigelt).

is often not yet sufficient for unsupervised segmentation without human intervention. Hence, data extraction commonly still relies on time-consuming manual annotation, which triggers the need for efficient and user-friendly annotation tools. Though various tools are available for this purpose (e.g., [Le Bot et al., 2010; Lobet et al., 2011]), each lacks important features, e.g., with regard to root modeling, editing of annotations, time-series support, or data export [**Möller et al., 2019a**]. With *rhizoTrak* we provide a single open-source tool integrating most of the relevant functionality and providing large flexibility, particularly for interactive annotation of time-series image data.

In the development of *rhizoTrak* we did not start from scratch, but largely benefit from the open-source Fiji plugin TrakEM2 [Cardona et al., 2012] which forms the fundament of *rhizoTrak*. TrakEM2 targets at morphological data mining and three-dimensional modeling,

---

[12]rhizoTrak website, `https://prbio-hub.github.io/rhizoTrak/` (accessed: November 24, 2021)

subsuming manifold functionality, e.g., for image visualization and navigation, registration, annotation, and measuring. As one of its intended areas of application is neural circuit reconstruction, and since networks of neuronal cells exhibit significant parallels with root systems, many of TrakEM2's built-in functionalities and data types as well as its elaborate GUI are perfectly matching our needs for annotating roots.

In *rhizoTrak* we extend TrakEM2's functionality and data structures to account for the differences between 3D volumetric image data of neural systems and time-series data of roots. Besides improved image import and visualization options, this mainly subsumes a larger flexibility in defining status labels attached to treeline segments for tagging root status. In addition, TrakEM2's connectors, by which annotation objects can be logically linked together, e.g., to properly model an object over different slices in a 3D stack, are defined more fail-safe in *rhizoTrak*. On the programming level unique identifiers were introduced for assigning connectors to objects whereas in TrakEM2 only geometric coincidence is used. These advanced connectors are also the fundament for assisting the user with inherent consistency checks to prevent the creation of topologically inconsistent annotations. *rhizoTrak* supports the export of annotation data in RSML format [Lobet et al., 2015] enabling interaction with other tools. In addition, statistical measurements of annotated roots can be exported for processing with external tools, e.g., with our own R package *rhizoTrakAddons* supporting quick statistical investigations.

The public version of *rhizoTrak* does not yet include functionality for (semi-)automatic root segmentation, though internally corresponding methods are already under development. For easy integration of additional image analysis functionality, *rhizoTrak* features a plugin interface which allows to dynamically load *MiToBo* operators into *rhizoTrak* at run time. These are required to implement an interface enabling data exchange between *MiToBo* and *rhizoTrak*. In particular approaches based on deep learning may presumably lead to further boosts in segmentation performance in near future. As such approaches require significant amounts of properly annotated training data, tools like *rhizoTrak* for manual annotation – or inspection and correction of automatically generated segmentation results – form a valuable fundament in these processes. Hence, ongoing work aims to evolve *rhizoTrak* towards an integrated platform for manual annotation as well as automatic segmentation and analysis of time-series root images.

1.5 Contributions and Conclusions

This thesis provides a comprehensive overview of the author's research on methods and tools for image analysis in the life sciences over more than one decade. The large diversity of approaches presented here mirrors the huge variety in analysis tasks emerging from bioimaging and related areas, and illustrates the importance of objective computational methods.

In our work, we extended existing and devised novel and advanced algorithms for various problems of low-level feature detection and object segmentation. Regarding elastic registration of 2-DE gel images, we pointed out domain-independent keypoint features as a larger and more robust fundament for feature-based registration approaches compared to the commonly used protein spot features [**Möller & Posch, 2009**]. For the task of particle detection from microscope images, we extended an existing algorithm towards improved scale-adaptivity [**Greß et al., 2010**], resulting in a larger flexibility and robustness in detection. Besides providing reliable counts of stress granules in different research contexts for which the algorithm was originally designed [**Bley et al., 2015**; Mensch et al., 2018], it underlined its broad applicability also by robustly detecting plastids, nuclei and peroxisomes [Franke et al., 2015; Erickson et al., 2018], or protein localizations in the plasma membrane [**Bürstenbinder et al., 2017**].

Active contour models have gained widespread use for the segmentation of objects in bioimages. For a task at hand, the selection of the best suited type of active contours is not straightforward. To ease this selection process, we carved out individual characteristics of snakes and level sets in a comparative study in [**Möller & Posch, 2012**]. In [**Misiak et al., 2014**] we adopted snakes to segment the complete areas of neurons. With this approach we were among the first to extract distributions of biomolecules inside the neurites of a neuron and, hence, to provide localization profiles in a fully automated fashion. To adopt snakes for the segmentation of cells with inhomogeneous intensity characteristics and vanishing boundaries, we proposed a novel cascaded optimization scheme in [**Möller et al., 2010, 2011b**]. This demonstrated how challenging segmentation problems in bioimages can be tackled with active contours beyond solely adapting energy functionals.

Deep learning has added a novel conceptual level to image analysis in general, and specifically to the detection and segmentation of objects in images. Adopting CNNs for these tasks has boosted performance with regard to diverse applications, provided that sufficiently large amounts of representative training data are available. In bioimage analysis research questions and related image analysis problems are often highly specific, rendering the acquisition of appropriate training data cumbersome. For the CVPPA *Arabidopsis thaliana* Root Segmentation Challenge 2021 such data were provided. With our approach combining root segmentation with CNNs and skeleton-based morphological analysis we were the only team among the 16 teams registered for the challenge that submitted complete segmentation and main root extraction results. Hence, we were declared as the winners of the challenge. Beyond that, our work provided deeper insights into the capabilities and behaviors of various CNN network architectures and loss functions, forming a valuable foundation for future work to improve CNN performance in this application domain [**Möller et al., 2021**].

With *PaCeQuant* [**Möller et al., 2017**] and its supplemental tool collection [**Poeschl et al., 2020**] we established a novel, pioneering pipeline for the analysis of leaf epidermis pavement cells in a high-throughput fashion. *PaCeQuant* is the first tool to integrate cell segmentation with subsequent quantification of manifold shape characteristics in a fully automated manner. It paves the way towards an analysis of larger and more representative data sets, and in particular overcomes the restriction of analyzing isolated cells towards scrutinizing cell conglomerates. This allows to consider additional morphological aspects of cells, like different types of lobes, in research studies. *PaCeQuant* has contributed to new findings in our own research, e.g., regarding microtubule-associated proteins [**Mitra et al., 2019**], and has also been employed by other labs [Erguvan et al., 2019; Li et al., 2019; Wong et al., 2019a,b]. Besides, we have been invited to present *PaCeQuant* with its tools and the R package *PaCeQuantAna* in two book chapters [**Möller et al., 2019b; Poeschl et al., 2020**].

For the comparative analysis of cytoskeleton structures, we devised a novel approach to quantify and compare cytoskeleton characteristics based on texture measures and methods of unsupervised machine learning [**Möller et al., 2014, 2019c**]. Our workflow overcomes the need of many other approaches to explicitly segment filamentous structures, which results in larger flexibility with regard to potential fields of application. It proved suitable to group cells into different genotypical categories according to their actin cytoskeletons [**Möller et al., 2014**], and formed the fundament for the first comprehensive characterization of the IQ67 DOMAIN protein family based on their microtubule localization patterns [**Bürstenbinder et al., 2017**]. Other research groups employed our approach for the analysis of actin filaments, e.g., in studies on fibroblasts, cardiomyoblasts, or functions of specific proteins [Hilbig et al., 2018; Huerta-García et al., 2019; Leinhos et al., 2019].

Not least due to a very close collaboration in research with many partners from different areas of the life sciences, we have always focused our work on evolving our algorithms and methods into software tools and applications easy to handle by end users. With *Alida* and *MiToBo* we established a development platform and toolbox that strongly supports these goals [**Möller et al., 2016; Posch & Möller, 2017**]. The general concepts of *Alida* and *MiToBo* provide a flexible and feature-rich development and test infrastructure. It has proven to form a suitable fundament for practical experiments in our lectures, and also served as basis for several student projects and degree theses, from which results were also reintegrated into *MiToBo*. In a comparative review of free software tools for microscope image analysis, an early version of *MiToBo* was rated among the top five out of 15 tools regarding combined usability and functionality [Wiesmann et al., 2015, Fig. 7].

MiToBo's deep anchoring in the ImageJ and Fiji ecosystem renders it easy to release our solutions to the public, and we substantially benefit from the very active bioimage analysis community. Our ImageJ update site has been among the top 20% of the sites with largest absolute access counts in the period from July 2020 to July 2021[13]. Many of the tools presented in this thesis are part of *MiToBo*, e.g., *MiCA* [**Möller & Posch, 2011**], *NeuronAnalyzer2D* [**Misiak et al., 2014**], *CytoskeletonAnalyzer2D* [**Möller et al., 2019c**], or *PaCeQuant* [**Möller et al., 2017**]. Also other tools like the successful *ScratchAssayAnalyzer* [Glaß et al., 2011, 2012] or a plugin for biofilm analysis, both implemented as *MiToBo* operators by Markus Glaß, found their way into the public as part of the *MiToBo* toolbox (e.g., [VanOudenhove et al., 2016; Berens et al., 2017; Gierl et al., 2020; Pavez Lorie et al., 2020]).

Our tool *rhizoTrak* [**Möller et al., 2019a**] likewise benefits from the ImageJ and Fiji ecosystem and is distributed via our ImageJ update site. With *rhizoTrak* we established a feature-rich open-source alternative to commercial solutions like WinRHIZO™ Tron[14] for the manual annotation of roots in minirhizotron time-series data. *rhizoTrak* provides full access to all annotation data and, hence, yields a flexible platform not only for extracting statistical data of roots, but also for the acquisition of training data for root segmentation based on deep learning. Moreover, a direct integration of such techniques in the future towards devising fully automatic workflows for root segmentation from minirhizotron images is straightforward.

In summary, the work presented in this thesis illustrates the broad range of image analysis challenges we have worked on over the years. We developed custom-fit algorithmic solutions and established advanced workflows for diverse problems of object detection, segmentation, and quantitative analysis. Not least by releasing these approaches as user-friendly tools to the bioimage analysis community and to end users, our computational methods successfully contribute to novel findings and advances in various areas of the life sciences.

---

[13]Data derived from counts as published on ImageJ Update Site Statistics website, `https://imagej.net/update-sites/stats` (accessed: October 19, 2021)

[14]WinRHIZO™ Tron website, Regent Instruments Inc., `https://regentinstruments.com/assets/winrhizotron_about2.html` (accessed: October 17, 2021)

Bibliography

Abel, S., Savchenko, T., & Levy, M. (2005). Genome-wide comparative analysis of the IQD gene families in *Arabidopsis thaliana* and *Oryza sativa*. *BMC Evolutionary Biology*, *5*(72). DOI: 10.1186/1471-2148-5-72.

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B. A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., & Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, *46*(W1), W537–W544. DOI: 10.1093/nar/gky379.

Aittokallio, T., Salmi, J., Nyman, T. A., & Nevalainen, O. S. (2005). Geometrical distortions in two-dimensional gels: applicable correction methods. *Journal of Chromatography B*, *815*(1-2), 25–37. DOI: 10.1016/j.jchromb.2004.07.037.

Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., & Walter, P. (2015). *Molecular Biology of the Cell*. Garland Science, Taylor and Francis Group, 6th ed.

Armour, W. J., Barton, D. A., Law, A. M., & Overall, R. L. (2015). Differential Growth in Periclinal and Anticlinal Walls during Lobe Formation in Arabidopsis Cotyledon Pavement Cells. *The Plant Cell*, *27*(9), 2484–2500. DOI: 10.1105/tpc.114.126664.

Atkinson, J. A., Pound, M. P., Bennett, M. J., & Wells, D. M. (2019). Uncovering the hidden half of plants using new advances in root phenotyping. *Current Opinion in Biotechnology*, *55*, 1–8. DOI: 10.1016/j.copbio.2018.06.002.

Bannigan, A., & Baskin, T. I. (2005). Directional cell expansion - turning toward actin. *Current Opinion in Plant Biology*, *8*(6), 619–624. DOI: 10.1016/j.pbi.2005.09.002.

Bay, H., Tuytelaars, T., & Van Gool, L. (2006). SURF: Speeded Up Robust Features. In *European Conference on Computer Vision*, (pp. 404–417). Springer.

Berens, E., Sharif, G., Schmidt, M., Yan, G., Shuptrine, C., Weiner, L., Glasgow, E., Riegel, A., & Wellstein, A. (2017). Keratin-associated protein 5-5 controls cytoskeletal function and cancer cell vascular invasion. *Oncogene*, *36*(5), 593–605. Doi: 10.1038/onc.2016.234.

Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Ohl, P., Thiel, K., & Wiswedel, B. (2009). KNIME – the Konstanz information miner: version 2.0 and beyond. *ACM SIGKDD Explorations Newsletter*, *11*(1), 26–31. DOI: 10.1145/1656274.1656280.

Bley, N., Lederer, M., Pfalz, B., Reinke, C., Fuchs, T., Glaß, M., Möller, B., & Hüttelmaier, S. (2015). Stress granules are dispensable for mRNA stabilization during cellular stress. *Nucleic Acids Research*, *43*(4), e26. DOI: 10.1093/nar/gku1275.

Boquet-Pujadas, A., Olivo-Marin, J.-C., & Guillén, N. (2021). Bioimage Analysis and Cell Motility. *Patterns*, *2*(1), 100170. DOI: 10.1016/j.patter.2020.100170.

Brandes, U., Eiglsperger, M., Herman, I., Himsolt, M., & Marshall, M. S. (2002). GraphML Progress Report Structural Layer Proposal. In P. Mutzel, M. Jünger, & S. Leipert (Eds.) *Proc. of Intern. Symp. on Graph Drawing*, (pp. 501–512). Berlin, Heidelberg: Springer.

Breuel, T. M. (2015). The Effects of Hyperparameters on SGD Training of Neural Networks. ArXiv:1508.02788, https://arxiv.org/abs/1508.02788.

Bürstenbinder, K., Möller, B., Plötner, R., Stamm, G., Hause, G., Mitra, D., & Abel, S. (2017). The IQD Family of Calmodulin-Binding Proteins Links Calcium Signaling to Microtubules, Membrane Subdomains, and the Nucleus. *Plant Physiology*, *173*(3), 1692–1708. DOI: 10.1104/pp.16.01743.

Cardona, A., Saalfeld, S., Schindelin, J., Arganda-Carreras, I., Preibisch, S., Longair, M., Tomancak, P., Hartenstein, V., & Douglas, R. J. (2012). TrakEM2 Software for Neural Circuit Reconstruction. *PLOS One*, *7*(6), e38011. DOI: 10.1371/journal.pone.0038011.

Cardona, A., & Tomancak, P. (2012). Current challenges in open-source bioimage informatics. *Nature Methods*, *9*(7), 661–665. DOI: 10.1038/nmeth.2082.

Carpenter, A. E., Kamentsky, L., & Eliceiri, K. W. (2012). A Call for Bioimaging Software Usability. *Nature Methods*, *9*(7), 666–670. DOI: 10.1038/nmeth.2073.

Chan, T. F., & Vese, L. A. (2001). Active contours without edges. *IEEE Transactions on Image Processing*, *10*(2), 266–277. DOI: 10.1109/83.902291.

Chenouard, N., Smal, I., De Chaumont, F., Maška, M., Sbalzarini, I. F., Gong, Y., Cardinale, J., Carthel, C., Coraluppi, S., Winter, M., et al. (2014). Objective comparison of particle tracking methods. *Nature Methods*, *11*, 281–289. DOI: 10.1038/nmeth.2808.

Clark, B. N., & Gutstein, H. B. (2008). The myth of automated, high-throughput two-dimensional gel analysis. *Proteomics*, *8*(6), 1197–1203. DOI: 10.1002/pmic.200700709.

Cremers, D., Rousson, M., & Deriche, R. (2007). A Review of Statistical Approaches to Level Set Segmentation: Integrating Color, Texture, Motion and Shape. *International Journal of Computer Vision*, *72*(2), 195–215. DOI: 10.1007/s11263-006-8711-1.

Croft, W. J. (2006). *Under the microscope: a brief history of microscopy*, vol. 5 of *Series in popular science*. World Scientific.

Daston, L., & Lunbeck, E. (2011). *Histories of scientific observation*. Univ. of Chicago Press.

De Chaumont, F., Dallongeville, S., Chenouard, N., Hervé, N., Pop, S., Provoost, T., Meas-Yedid, V., Pankajakshan, P., Lecomte, T., et al. (2012). Icy: an open bioimage informatics platform for extended reproducible research. *Nature Methods*, *9*(7), 690–696. DOI: 10.1038/nmeth.2075.

de Kroon, H., Mommer, L., & Nishiwaki, A. (2003). Root Competition: Towards a Mechanistic Understanding. In *Root Ecology*, (pp. 215–234). Springer.

Delgado-Gonzalo, R., Uhlmann, V., Schmitter, D., & Unser, M. (2014). Snakes on a Plane: A perfect snap for bioimage analysis. *IEEE Signal Processing Magazine*, *32*(1), 41–48. DOI: 10.1109/MSP.2014.2344552.

Desolneux, A., Moisan, L., & More, J.-M. (2003). A grouping principle and four applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*(4), 508–513. DOI: 10.1109/TPAMI.2003.1190576.

Dobrucki, J. W., & Kubitscheck, U. (2017). Fluorescence Microscopy. In U. Kubitscheck (Ed.) *Fluorescence Microscopy: From Principles to Biological Applications*, (pp. 85–132). Wiley-VCH, 2 ed.

Dowsey, A. W., Dunn, M. J., & Yang, G.-Z. (2003). The role of bioinformatics in two-dimensional gel electrophoresis. *Proteomics*, *3*(8), 1567–1596. DOI: 10.1002/pmic.200300459.

Dunst, S., & Tomancak, P. (2019). Imaging Flies by Fluorescence Microscopy: Principles, Technologies, and Applications. *Genetics*, *211*(1), 15–34. DOI: 10.1534/genetics.118.300227.

Eckersley, P., Nasser, Y., et al. (2021). EFF AI Progress Measurement Project (2017-), Sec. 3.B: Vision and Image Modelling. `https://eff.org/ai/metrics#Vision`, access: Sep. 11, 2021.

Eliceiri, K. W., Berthold, M. R., Goldberg, I. G., Ibáñez, L., Manjunath, B. S., Martone, M. E., Murphy, R. F., Peng, H., Plant, A. L., Roysam, B., et al. (2012). Biological imaging software tools. *Nature Methods*, *9*(7), 697–710. DOI: 10.1038/nmeth.2084.

Erguvan, Ö., Louveaux, M., Hamant, O., & Verger, S. (2019). ImageJ SurfCut: a user-friendly pipeline for high-throughput extraction of cell contours from 3D image stacks. *BMC Biology*, *17*(1), 38. DOI: 10.1186/s12915-019-0657-1.

Erickson, J., Adlung, N., Lampe, C., Bonas, U., & Schattat, M. (2018). The Xanthomonas effector XopL uncovers the role of microtubules in stromule extension and dynamics in *Nicotiana benthamiana*. *The Plant Journal*, *93*(5), 856–870. Doi: 10.1111/tpj.13813.

Fornito, A., Zalesky, A., & Bullmore, E. (2016). *Fundamentals of brain network analysis*. Academic Press.

Franke, L., Storbeck, B., Erickson, J. L., Rödel, D., Schröter, D., Möller, B., & Schattat, M. H. (2015). The 'MTB Cell Counter' a versatile tool for the semi-automated quantification of sub-cellular phenotypes in fluorescence microscopy images. A case study on plastids, nuclei and peroxisomes. *Journal of Endocytobiosis and Cell Research*, *26*, 31–42.

Fraz, M. M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A. R., Owen, C. G., & Barman, S. A. (2012). Blood vessel segmentation methodologies in retinal images – A survey. *Computer Methods and Programs in Biomedicine*, *108*(1), 407–433. DOI: 10.1016/j.cmpb.2012.03.009.

Gaggion, N., Ariel, F., Daric, V., Lambert, É., Legendre, S., Roulé, T., Camoirano, A., Milone, D. H., Crespi, M., Blein, T., & Ferrante, E. (2021). ChronoRoot: High-throughput phenotyping by deep segmentation networks reveals novel temporal parameters of plant root system architecture. *GigaScience*, *10*(7). DOI: 10.1093/gigascience/giab052.

Gao, Y., Zhang, Y., Zhang, D., Dai, X., Estelle, M., & Zhao, Y. (2015). Auxin binding protein 1 (ABP1) is not required for either auxin signaling or *Arabidopsis* development. *Proceedings of the National Academy of Sciences*, *112*(7), 2275–2280. DOI: 10.1073/pnas.1500365112.

Gierl, L., Stoy, K., Faíña, A., Horn, H., & Wagner, M. (2020). An open-source robotic platform that enables automated monitoring of replicate biofilm cultivations using optical coherence tomography. *NPJ Biofilms and Microbiomes*, *6*(18). DOI: 10.1038/s41522-020-0129-y.

Giving software its due (2019). Editorial. *Nature Methods*, *16*(3), 207.

Glaß, M., Möller, B., Zirkel, A., Wächter, K., Hüttelmaier, S., & Posch, S. (2011). Scratch Assay Analysis with Topology-Preserving Level Sets and Texture Measures. In J. Vitrià, J. M. Sanches, & M. Hernández (Eds.) *Proc. of 5th Iberian Conf. on Pattern Recognition and Image Analysis*, no. 6669 in LNCS, (pp. 100–108). Springer, Berlin, Heidelberg. DOI: 10.1007/978-3-642-21257-4_13.

Glaß, M., Möller, B., Zirkel, A., Wächter, K., Hüttelmaier, S., & Posch, S. (2012). Cell migration analysis: Segmenting scratch assay images with level sets and support vector machines. *Pattern Recognition*, *45*(9), 3154–3165. DOI: 10.1016/j.patcog.2012.03.001.

Goldberg, I. G., Allan, C., Burel, J.-M., Creager, D., Falconi, A., Hochheiser, H., Johnston, J., Mellen, J., Sorger, P. K., & Swedlow, J. R. (2005). The Open Microscopy Environment (OME) Data Model and XML file: open tools for informatics and quantitative analysis in biological imaging. *Genome Biology*, *6*(5), R47. DOI: 10.1186/gb-2005-6-5-r47.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Greß, O., Möller, B., Stöhr, N., Hüttelmaier, S., & Posch, S. (2010). Scale-adaptive Wavelet-based Particle Detection in Microscopy Images. In H.-P. Meinzer, T. M. Deserno, H. Handels, & T. Tolxdorff (Eds.) *Bildverarbeitung für die Medizin*, Informatik Aktuell, (pp. 266–270). Berlin: Springer. ISBN 978-3-642-11967-5.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., & Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, *77*, 354–377. Doi: 10.1016/j.patcog.2017.10.013.

Han, X., Xu, C., & Prince, J. L. (2003). A topology preserving level set method for geometric deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*(6), 755–768. DOI: 10.1109/TPAMI.2003.1201824.

Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, (6), 610–621.

Hazelwood, K. L., Olenych, S. G., Griffin, J. D., Cathcart, J. A., & Davidson, M. W. (2007). Entering the Portal: Understanding the Digital Image Recorded Through a Microscope. In S. L. Shorte, & F. Frischknecht (Eds.) *Imaging Cellular and Molecular Biological Functions*, (pp. 3–43). Berlin, Heidelberg: Springer.

Hilbig, D., Sittig, D., Hoffmann, F., Rothemund, S., Warmt, E., Quaas, M., Stürmer, J., Seiler, L., Liebscher, I., Hoang, N. A., Käs, J., Banks, L., & Aust, G. (2018). Mechano-Dependent Phosphorylation of the PDZ-Binding Motif of CD97/ADGRE5 Modulates Cellular Detachment. *Cell Reports*, *24*(8), 1986–1995. DOI: 10.1016/j.celrep.2018.07.071.

Huerta-García, E., del Pilar Ramos-Godinez, M., López-Saavedra, A., Alfaro-Moreno, E., Gómez-Crisóstomo, N., Colín-Val, Z., Sánchez-Barrera, H., & López-Marure, R. (2019). Internalization of Titanium Dioxide Nanoparticles is Mediated by Actin-Dependent Reorganization and Clathrin- and Dynamin-Mediated Endocytosis in H9c2 Rat Cardiomyoblasts. *Chemical Research in Toxicology*, *32*(4), 578–588. DOI: 10.1021/acs.chemrestox.8b00284.

Ivakov, A., & Persson, S. (2013). Plant cell shape: modulators and measurements. *Frontiers in Plant Science*, *4*, 439. DOI: 10.3389/fpls.2013.00439.

Ivanov, P., Kedersha, N., & Anderson, P. (2019). Stress Granules and Processing Bodies in Translational Control. *Cold Spring Harbor Perspectives in Biology*, *11*(5), a032813.

Jacques, E., Verbelen, J.-P., & Vissenberg, K. (2014). Review on shape formation in epidermal pavement cells of the *Arabidopsis* leaf. *Functional Plant Biology*, *41*(9), 914–921.

Jähne, B. (2005). *Digital Image Processing*. Springer, 6th ed.

Kass, M., Witkin, A., & Terzopoulos, D. (1988). Snakes: Active Contour Models. *International Journal of Computer Vision*, *1*(4), 321–331.

Kimori, Y., Hikino, K., Nishimura, M., & Mano, S. (2016). Quantifying morphological features of actin cytoskeletal filaments in plant cells based on mathematical morphology. *Journal of Theoretical Biology*, *389*, 123–131. DOI: 10.1016/j.jtbi.2015.10.031.

Kirchner, S. (2012). Konzeption und Implementierung einer graphischen Programmieroberfläche für Alida. Bachelor Thesis, Martin Luther University Halle-Wittenberg, Germany.

Kirchner, S., Posch, S., & Möller, B. (2012). Graphical Programming in Alida and ImageJ 2.0 with Grappa. In *Proc. of ImageJ User & Developer Conference*, (pp. 138–143). Mondorf-les-Bains, Luxembourg.

Kucukkececi, C., Dogrusoz, U., Belviranli, E., & Dilek, A. (2017). Chisio: A Compound Graph Editing and Layout Framework. ArXiv:1708.07762, https://arxiv.org/abs/1708.07762.

Le Bot, J., Serra, V., Fabre, J., Draye, X., Adamowicz, S., & Pagès, L. (2010). DART: a software to analyse root system architecture and development from captured images. *Plant and Soil*, *326*(1), 261–273. DOI: 10.1007/s11104-009-0005-2.

Leinhos, L., Peters, J., Krull, S., Helbig, L., Vogler, M., Levay, M., van Belle, G., Ridley, A., Lutz, S., Katschinski, D., & Zieseniss, A. (2019). Hypoxia suppresses myofibroblast differentiation by changing RhoA activity. *Journal of Cell Science*, *132*(5). DOI: 10.1242/jcs.223230.

Letort, G., Ennomani, H., Gressin, L., Théry, M., & Blanchoin, L. (2015). Dynamic reorganization of the actin cytoskeleton. *F1000Research*, *4*. DOI: 10.12688/f1000research.6374.1.

Levet, F., Carpenter, A. E., Eliceiri, K. W., Kreshuk, A., Bankhead, P., & Haase, R. (2021). Developing open-source software for bioimage analysis: opportunities and challenges. *F1000Research*, *10:302*. DOI: 10.12688/f1000research.52531.1.

Levitan, I. B., & Kaczmarek, L. K. (2015). *The Neuron: Cell and Molecular Biology*. Oxford University Press, 4 ed.

Li, S., Blanchoin, L., Yang, Z., & Lord, E. M. (2003). The Putative Arabidopsis Arp2/3 Complex Controls Leaf Cell Morphogenesis. *Plant Physiology*, *132*(4), 2034–2044.

Li, Y., Liu, Y., & Zolman, B. (2019). Metabolic Alterations in the Enoyl-CoA Hydratase 2 Mutant Disrupt Peroxisomal Pathways in Seedlings. *Plant Physiology*, *180*(4), 1860–1876.

Li, Y., Wang, S., Tian, Q., & Ding, X. (2015). Feature representation for statistical-learning-based object detection: A review. *Pattern Recognition*, *48*(11), 3542–3559.

Li, Z., Guo, R., Li, M., Chen, Y., & Li, G. (2020). A review of computer vision technologies for plant phenotyping. *Computers and Electronics in Agriculture*, *176*, 105672.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal Loss for Dense Object Detection. In *Proc. of the IEEE Intern. Conf. on Computer Vision*, (pp. 2980–2988).

Lobet, G., Pagès, L., & Draye, X. (2011). A novel image-analysis toolbox enabling quantitative analysis of root system architecture. *Plant Physiology*, *157*(1), 29–39.

Lobet, G., Pound, M. P., Diener, J., Pradal, C., Draye, X., Godin, C., Javaux, M., Leitner, D., Meunier, F., Nacry, P., et al. (2015). Root system markup language: toward a unified root architecture description language. *Plant Physiology*, *167*(3), 617–627.

Loncomilla, P., del Solar, J. R., & Martínez, L. (2016). Object recognition using local invariant features for robotic applications: A survey. *Pattern Recognition*, *60*, 499–514.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110. DOI: 10.1023/B:VISI.0000029664.99615.94.

Lucas, A. M., Ryder, P. V., Li, B., Cimini, B. A., Eliceiri, K. W., & Carpenter, A. E. (2021). Open-source deep-learning software for bioimage segmentation. *Molecular Biology of the Cell*, *32*(9), 823–829. DOI: 10.1091/mbc.E20-10-0660.

Lundberg, E., & Borner, G. H. (2019). Spatial proteomics: a powerful discovery tool for cell biology. *Nature Reviews Molecular Cell Biology*, *20*(5), 285–302. DOI: 10.1038/s41580-018-0094-y.

Ma, J., Jiang, X., Fan, A., Jiang, J., & Yan, J. (2021). Image Matching from Handcrafted to Deep Features: A Survey. *International Journal of Computer Vision*, *129*(1), 23–79.

Magdeldin, S., Enany, S., Yoshida, Y., Xu, B., Zhang, Y., Zureena, Z., Lokamani, I., Yaoita, E., & Yamamoto, T. (2014). Basics and recent advances of two dimensional-polyacrylamide gel electrophoresis. *Clinical Proteomics*, *11*, 16. DOI: 10.1186/1559-0275-11-16.

Martinotti, S., & Ranzato, E. (2016). 2-DE Gel Analysis: The Spot Detection. In E. Marengo, & E. Robotti (Eds.) *2-D PAGE Map Analysis: Methods and Protocols*, (pp. 155–164). Springer.

Matschegewski, C., Staehlke, S., Birkholz, H., Lange, R., Beck, U., Engel, K., & Nebe, J. B. (2012). Automatic Actin Filament Quantification of Osteoblasts and Their Morphometric Analysis on Microtextured Silicon-Titanium Arrays. *Materials*, *5*(7), 1176–1195.

Mazzarello, P. (1999). A unifying concept: the history of cell theory. *Nature Cell Biology*, *1*(1), E13–E15. DOI: 10.1038/8964.

McQuin, C., Goodman, A., Chernyshev, V., Kamentsky, L., Cimini, B. A., Karhohs, K. W., Doan, M., Ding, L., Rafelski, S. M., et al. (2018). CellProfiler 3.0: Next-generation image processing for biology. *PLOS Biology*, *16*(7), e2005970. DOI: 10.1371/journal.pbio.2005970.

Meijering, E. (2010). Neuron Tracing in Perspective. *Cytometry Part A*, *77A*(7), 693–704.

Meijering, E. (2012). Cell Segmentation: 50 Years Down the Road. *IEEE Signal Processing Magazine*, *29*(5), 140–145. DOI: 10.1109/MSP.2012.2204190.

Meijering, E. (2020). A bird's-eye view of deep learning in bioimage analysis. *Computational and Structural Biotechnology Journal*, *18*, 2312–2325. DOI: 10.1016/j.csbj.2020.08.003.

Meijering, E., Carpenter, A. E., Peng, H., Hamprecht, F. A., & Olivo-Marin, J.-C. (2016). Imagining the future of bioimage analysis. *Nature Biotechnology*, *34*(12), 1250–1255.

Meijering, E., Dzyubachyk, O., & Smal, I. (2012). Chapter nine - Methods for Cell and Particle Tracking. In P. M. Conn (Ed.) *Imaging and Spectroscopic Analysis of Living Cells*, vol. 504 of *Methods in Enzymology*, (pp. 183–200). Academic Press.

Mensch, A., Meinhardt, B., Bley, N., Hüttelmaier, S., Schneider, I., Stoltenburg-Didinger, G., Kraya, T., Müller, T., & Zierz, S. (2018). The p.S85C-mutation in MATR3 impairs stress granule formation in Matrin-3 myopathy. *Experimental Neurology*, *306*, 222–231.

Merchant, F., Shah, S., & Castleman, K. (2008). Object Measurement. In *Microscope Image Processing*, (pp. 195–219). Academic Press, Elsevier.

Michailovich, O., Rathi, Y., & Tannenbaum, A. (2007). Image Segmentation Using Active Contours Driven by the Bhattacharyya Gradient Flow. *IEEE Transactions on Image Processing*, *16*(11), 2787–2801. DOI: 10.1109/TIP.2007.908073.

Mishkin, D., Sergievskiy, N., & Matas, J. (2017). Systematic evaluation of convolution neural network advances on the Imagenet. *Computer Vision and Image Understanding*, *161*, 11–19.

Misiak, D., Posch, S., Lederer, M., Reinke, C., Hüttelmaier, S., & Möller, B. (2014). Extraction of protein profiles from primary neurons using active contour models and wavelets. *Journal of Neuroscience Methods*, *225*, 1–12. DOI: 10.1016/j.jneumeth.2013.12.009.

Mitra, D., Klemm, S., Kumari, P., Quegwer, J., Möller, B., Poeschl, Y., Pflug, P., Stamm, G., Abel, S., & Bürstenbinder, K. (2019). Microtubule-associated protein IQ67 DOMAIN5 regulates morphogenesis of leaf pavement cells in *Arabidopsis thaliana. Journal of Experimental Botany*, *70*(2), 529–543. DOI: 10.1093/jxb/ery395.

Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using Deep Learning for Image-Based Plant Disease Detection. *Frontiers in Plant Science*, *7*, 1419. DOI: 10.3389/fpls.2016.01419.

Möller, B., & Bürstenbinder, K. (2019). Semi-automatic Cell Segmentation from Noisy Image Data for Quantification of Microtubule Organization on Single Cell Level. In *Proc. of IEEE 16th Intern. Symp. on Biomedical Imaging (ISBI)*, (pp. 199–203). Venice, Italy.

Möller, B., Chen, H., Schmidt, T., Zieschank, A., Patzak, R., Türke, M., Weigelt, A., & Posch, S. (2019a). rhizoTrak: a flexible open source Fiji plugin for user-friendly manual annotation of time-series images from minirhizotrons. *Plant and Soil*, *444*, 519–534.

Möller, B., Glaß, M., Misiak, D., & Posch, S. (2016). MiToBo - a toolbox for image processing and analysis. *Journal of Open Research Software*, *4*(1), e17. DOI: 10.5334/jors.103.

Möller, B., Greß, O., & Posch, S. (2011a). Knowing what happened - automatic documentation of image analysis processes. In J. Crowley, B. Draper, & M. Thonnat (Eds.) *Proc. of 8th Intern. Conf. on Computer Vision Systems*, vol. 6962 of *LNCS*, (pp. 1–10). Sophia Antipolis, France: Springer. DOI: 10.1007/978-3-642-23968-7_1.

Möller, B., Greß, O., Stöhr, N., Hüttelmaier, S., & Posch, S. (2011b). *VISIGRAPP 2010, Revised Selected Papers of Int. Joint Conf. on Computer Vision, Imaging and Computer Graphics. Theory and Applications*, vol. 229 of *Communications in Computer and Information Science*, chap. Adaptive Segmentation of Particles and Cells for Fluorescent Microscope Imaging, (pp. 154–169). Springer. DOI: 10.1007/978-3-642-25382-9.

Möller, B., Piltz, E., & Bley, N. (2014). Quantification of Actin Structures using Unsupervised Pattern Analysis Techniques. In *Proceedings of 22nd Int. Conf. on Pattern Recognition (ICPR)*, (pp. 3251–3256). Stockholm, Sweden: IEEE. DOI: 10.1109/ICPR.2014.560.

Möller, B., Poeschl, Y., Klemm, S., & Bürstenbinder, K. (2019b). Morphological Analysis of Leaf Epidermis Pavement Cells with PaCeQuant. In F. Cvrčková, & V. Žárský (Eds.) *Plant Cell Morphogenesis: Methods and Protocols*, chap. 22, (pp. 329–349). Springer.

Möller, B., Poeschl, Y., Plötner, R., & Bürstenbinder, K. (2017). PaCeQuant: A Tool for High-Throughput Quantification of Pavement Cell Shape Characteristics. *Plant Physiology*, *175*(3), 998–1017. DOI: 10.1104/pp.17.00961.

Möller, B., & Posch, S. (2009). Robust Features for 2-D Electrophoresis Gel Image Registration. *Electrophoresis*, *30*, 4137–4148. DOI: 10.1002/elps.200900293.

Möller, B., & Posch, S. (2011). MiCA - easy cell image analysis with normalized snakes. In *Proc. of 6th Workshop on Microscopic Image Analysis with Applications in Biology (MIAAB), held in conjunction with Intern. Conf. on Systems Biology (ICSB)*. Heidelberg, Germany.

Möller, B., & Posch, S. (2012). Comparing active contours for the segmentation of biomedical images. In *Proc. of 9th IEEE Intern. Symp. on Biomedical Imaging (ISBI)*, (pp. 736–739). Barcelona, Spain. DOI: 10.1109/ISBI.2012.6235653.

Möller, B., & Posch, S. (2013). A framework unifying the development of image analysis algorithms and associated user interfaces. In *Proc. of 13th IAPR International Conference on Machine Vision Applications (MVA)*, (pp. 447–450). Kyoto, Japan.

Möller, B., & Posch, S. (2021). *Alida — Advanced Library for Integrated Development of Data Analysis Applications, User and Programmer Guide*. http://alida.informatik.uni-halle.de/downloads/manual/AlidaManual.pdf, accessed July 31, 2021.

Möller, B., & Schattat, M. (2019). Quantification of Stromule Frequencies in Microscope Images of Plastids combining Ridge Detection and Geometric Criteria. In *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 2: BIOIMAGING*, (pp. 38–48). INSTICC, SciTePress. DOI: 10.5220/0007390300380048.

Möller, B., Schreck, B., & Posch, S. (2021). Analysis of Arabidopsis Root Images - Studies on CNNs and Skeleton-Based Root Topology. In *Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, 7th Workshop on Computer Vision in Plant Phenotyping and Agriculture (CVPPA)*, (pp. 1294–1302).

Möller, B., Stöhr, N., Hüttelmaier, S., & Posch, S. (2010). Cascaded Segmentation of Grained Cell Tissue with Active Contour Models. In *Proc. of 10th IEEE Intern. Conf. on Pattern Recognition (ICPR)*, (pp. 1481–1484). Istanbul, Turkey. DOI: 10.1109/ICPR.2010.366.

Möller, B., Zergiebel, L., & Bürstenbinder, K. (2019c). Quantitative and Comparative Analysis of Global Patterns of (Microtubule) Cytoskeleton Organization with CytoskeletonAnalyzer2D. In F. Cvrčková, & V. Žárský (Eds.) *Plant Cell Morphogenesis: Methods and Protocols*, chap. 10, (pp. 151–171). New York, NY: Springer.

Murphy, R., Meijering, E., & Danuser, G. (2005). Guest Editorial, Special Issue on Molecular and Cellular Bioimaging. *IEEE Trans. Image Process.*, *14*(9), 1233–1236.

Myers, G. (2012). Why bioimage informatics matters. *Nature Methods*, *9*(7), 659–660.

Narisetti, N., Henke, M., Seiler, C., Shi, R., Junker, A., Altmann, T., & Gladilin, E. (2019). Semi-automated Root Image Analysis (saRIA). *Scientific Reports*, *9*(1), 1–10.

Nédélec, F., Surrey, T., & Karsenti, E. (2003). Self-organisation and forces in the microtubule cytoskeleton. *Current Opinion in Cell Biology*, *15*(1), 118–124.

Niblack, W. (1985). *An Introduction to Digital Image Processing*. Strandberg.

Nienhaus, G. U., & Nienhaus, K. (2017). Fluorescence Labeling. In U. Kubitscheck (Ed.) *Fluorescence Microscopy: From Principles to Biological Applications*, (pp. 143–173). Wiley-VCH, 2 ed.

Nixon, M., & Aguado, A. (2020). *Feature Extraction and Image Processing for Computer Vision*. Academic Press, 4th ed.

Nowak, J., Eng, R., 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. DOI: 10.1038/s41467-020-20730-y.

Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(7), 971–987. DOI: 10.1109/TPAMI.2002.1017623.

Olivo-Marin, J.-C. (2002). Extraction of spots in biological images using multiscale products. *Pattern Recognition*, *35*(9), 1989–1996. DOI: 10.1016/S0031-3203(01)00127-3.

Osher, S., & Fedkiw, R. (2003). *Level Set Methods and Dynamic Implicit Surfaces*, vol. 153 of *Applied Mathematical Sciences*. Springer.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 62–66.

Ouyang, W., & Zimmer, C. (2017). The imaging tsunami: Computational opportunities and challenges. *Current Opinion in Systems Biology*, *4*, 105–113.

Pavez Lorie, E., Stricker, N., Plitta-Michalak, B., Chen, I.-P., Volkmer, B., Greinert, R., Jauch, A., et al. (2020). Characterisation of the novel spontaneously immortalized and invasively growing human skin keratinocyte line HaSKpw. *Scientific Reports*, *10*(1), 1–20.

Poeschl, Y., Möller, B., Müller, L., & Bürstenbinder, K. (2020). *User-friendly assessment of pavement cell shape features with PaCeQuant: Novel functions and tools*, vol. 160 of *Methods in Cell Biology*, (pp. 349–363). Academic Press. DOI: 10.1016/bs.mcb.2020.04.010.

Pomastowski, P., & Buszewski, B. (2014). Two-dimensional gel electrophoresis in the light of new developments. *Trends in Analytical Chemistry*, *53*, 167–177.

Posch, S., & Möller, B. (2017). Alida - Advanced Library for Integrated Development of Data Analysis Applications. *Journal of Open Research Software*, *5*(1), 7. DOI: 10.5334/jors.124.

Riggs, C. L., Kedersha, N., Ivanov, P., & Anderson, P. (2020). Mammalian stress granules and P bodies at a glance. *Journal of Cell Science*, *133*(16), jcs242487. DOI: 10.1242/jcs.242487.

Rittscher, J. (2010). Characterization of Biological Processes through Automated Image Analysis. *Annual Review of Biomedical Engineering*, *12*(1), 315–344.

Robotti, E., Calà, E., & Marengo, E. (2021). Two-Dimensional Gel Electrophoresis Image Analysis. In D. Cecconi (Ed.) *Proteomics Data Analysis*, (pp. 3–13). Springer.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Intern. Conf. on Med. Image Comp. and Computer-Assisted Intervention*, (pp. 234–241). Springer Intern. Publ. DOI: 10.1007/978-3-319-24574-4_28.

Rueda, L. (2014). *Microarray Image and Data Analysis: Theory and Practice*. CRC Press.

Rueden, C. T., Ackerman, J., Arena, E. T., Eglinger, J., Cimini, B. A., Goodman, A., Carpenter, A. E., & Eliceiri, K. W. (2019). Scientific Community Image Forum: A discussion forum for scientific image software. *PLOS Biology*, *17*(6), e3000340.

Rueden, C. T., Schindelin, J., Hiner, M. C., DeZonia, B. E., Walter, A. E., Arena, E. T., & Eliceiri, K. W. (2017). ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics*, *18*, 529. DOI: 10.1186/s12859-017-1934-z.

Sapala, A., Runions, A., & Smith, R. S. (2019). Mechanics, geometry and genetics of epidermal cell shape regulation: different pieces of the same puzzle. *Current Opinion in Plant Biology*, *47*, 1–8. DOI: 10.1016/j.pbi.2018.07.017.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, *9*(7), 676–682. DOI: 10.1038/nmeth.2019.

Schindelin, J., Rueden, C. T., Hiner, M. C., & Eliceiri, K. W. (2015). The ImageJ ecosystem: An open platform for biomedical image analysis. *Molecular Reproduction and Development*, *82*(7-8), 518–529. DOI: 10.1002/mrd.22489.

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675. DOI: 10.1038/nmeth.2089.

Schwede, B. (2020). Effiziente Segmentierung und Analyse von Pavement-Zellen. Bachelor Thesis, Martin Luther University Halle-Wittenberg, Germany.

Sethian, J. A. (1999). *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*, vol. 3 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press.

Shahzad, Z., Kellermeier, F., Armstrong, E. M., Rogers, S., Lobet, G., Amtmann, A., & Hills, A. (2018). EZ-Root-VIS: A Software Pipeline for the Rapid Analysis and Visual Reconstruction of Root System Architecture. *Plant Physiology*, *177*(4), 1368–1381. DOI: 10.1104/pp.18.00217.

Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. In Y. Bengio, & Y. LeCun (Eds.) *Proc. of 3rd Intern. Conf. on Learning Representations (ICLR)*. San Diego, CA, USA.

Smit, A. L., George, E., & Groenwold, J. (2000). Root Observations and Measurements at (Transparent) Interfaces with Soil. In A. L. Smit, A. G. Bengough, C. Engels, M. van Noordwijk, et al. (Eds.) *Root Methods: A Handbook*, (pp. 235–271). Springer.

Smith, A. G., Petersen, J., Selvan, R., & Rasmussen, C. R. (2020). Segmentation of roots in soil with U-Net. *Plant Methods*, *16*, 13. DOI: 10.1186/s13007-020-0563-0.

Snyder, W. E., & Qi, H. (2017). *Fundamentals of Computer Vision*. Cambridge Univ. Press.

Sofka, M., & Stewart, C. V. (2006). Retinal vessel centerline extraction using multiscale matched filters, confidence and edge measures. *Trans. on Med. Imag.*, *25*(12), 1531–1546.

Song, B., & Chan, T. (2002). A fast algorithm for level set based optimization. *UCLA Cam Report*, *2*(68).

Sonka, M., Hlavac, V., & Boyle, R. (1999). *Image Processing, Analysis, and Machine Vision*. Brooks/Cole Publishing Company, International Thomson Publishing, 2nd ed.

Sorzano, C. O. S., Arganda-Carreras, I., Thévenaz, P., Beloso, A., Morales, G., Valdés, I., Pérez-García, C., Castillo, C., Garrido, E., & Unser, M. (2008). Elastic image registration of 2-D gels for differential and repeatability studies. *Proteomics*, *8*(1), 62–65.

Speicher, D. W. (2004). Overview of proteome analysis. In D. W. Speicher (Ed.) *Proteome Analysis*, (pp. 1–18). Elsevier.

Staff, L., Hurd, P., Reale, L., Seoighe, C., Rockwood, A., & Gehring, C. (2012). The Hidden Geometries of the *Arabidopsis thaliana* Epidermis. Doi: 10.1371/journal.pone.0043546.

Swedlow, J. R., & Eliceiri, K. W. (2009). Open source bioimage informatics for cell biology. *Trends in Cell Biology, Special Issue - Imaging Cell Biology*, *19*(11), 656–660.

Thorn, K. (2016). A quick guide to light microscopy in cell biology. *Molecular Biology of the Cell*, *27*(2), 219–222. DOI: 10.1091/mbc.e15-02-0088.

Vanhaeren, H., Gonzalez, N., & Inzé, D. (2015). A Journey Through a Leaf: Phenomics Analysis of Leaf Growth in *Arabidopsis thaliana. The Arabidopsis Book*, *13*, e0181.

VanOudenhove, J., Medina, R., Ghule, P., Lian, J., Stein, J., Zaidi, S., & Stein, G. (2016). Transient RUNX1 Expression during Early Mesendodermal Differentiation of hESCs Promotes Epithelial to Mesenchymal Transition through TGFB2 Signaling. *Stem Cell Reports*, *7*(5), 884–896. DOI: 10.1016/j.stemcr.2016.09.006.

Vogel, C., & Marcotte, E. M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics*, *13*(4), 227–232.

Walter, A., Liebisch, F., & Hund, A. (2015). Plant phenotyping: from bean weighing to image analysis. *Plant Methods*, *11*, 14. DOI: 10.1186/s13007-015-0056-8.

Wang, L., He, L., Mishra, A., & Li, C. (2009). Active contours driven by local Gaussian distribution fitting energy. *Signal Processing*, *89*(12), 2435–2447. DOI: 10.1016/j.sigpro.2009.03.014.

Wang, T., Rostamza, M., Song, Z., Wang, L., McNickle, G., Iyer-Pascuzzi, A. S., Qiu, Z., & Jin, J. (2019). SegRoot: A high throughput segmentation method for root image analysis. *Computers and Electronics in Agriculture*, *162*, 845–854. DOI: 10.1016/j.compag.2019.05.017.

Waters, J. C. (2009). Accuracy and precision in quantitative fluorescence microscopy. *Journal of Cell Biology*, *185*(7), 1135–1148. DOI: 10.1083/jcb.200903097.

Wiesmann, V., Franz, D., Held, C., Münzenmayer, C., Palmisano, R., & Wittenberg, T. (2015). Review of free software tools for image analysis of fluorescence cell micrographs. *Journal of Microscopy*, *257*(1), 39–53. DOI: 10.1111/jmi.12184.

Wilson, R. (2017). *Astronomy through the Ages*. Princeton University Press.

Wong, J., Kato, T., Belteton, S., Shimizu, R., Kinoshita, N., Higaki, T., et al. (2019a). Basic Proline-Rich Protein-Mediated Microtubules are Essential for Lobe Growth and Flattened Cell Geometry. *Plant Physiology*, *181*(4), 1535–1551. DOI: 10.1104/pp.19.00811.

Wong, J. H., Spartz, A. K., Park, M. Y., Du, M., & Gray, W. M. (2019b). Mutation of a Conserved Motif of PP2C.D Phosphatases Confers SAUR Immunity and Constitutive Activity. *Plant Physiology*, *181*(1), 353–366. DOI: 10.1104/pp.19.00496.

Wu, Q., & Castleman, K. (2008). Image Segmentation. In Q. Wu, F. Merchant, & K. Castleman (Eds.) *Microscope Image Processing*, (pp. 159–194). Academic Press, Elsevier.

Wu, T.-C., Belteton, S. A., Pack, J., Szymanski, D. B., & Umulis, D. M. (2016). LobeFinder: a Convex Hull-Based Method for Quantitative Boundary Analyses of Lobed Plant Cells. *Plant Physiology*, *171*(4), 2331–2342. DOI: 10.1104/pp.15.00972.

Xie, X., & Mirmehdi, M. (2008a). A Galaxy of Texture Features. In M. Mirmehdi, X. Xie, & J. Suri (Eds.) *Handbook of Texture Analysis*, (pp. 375–406). Imperial College Press.

Xie, X., & Mirmehdi, M. (2008b). MAC: Magnetostatic Active Contour Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(4), 632–646.

Xing, F., & Yang, L. (2016). Robust Nucleus/Cell Detection and Segmentation in Digital Pathology and Microscopy Images: A Comprehensive Review. *IEEE Reviews in Biomedical Engineering*, *9*, 234–263. DOI: 10.1109/RBME.2016.2515127.

Xu, T., Wen, M., Nagawa, S., Fu, Y., Chen, J.-G., Wu, M.-J., Perrot-Rechenmann, C., Friml, J., Jones, A. M., & Yang, Z. (2010). Cell surface- and rho GTPase-based auxin signaling controls cellular interdigitation in Arabidopsis. *Cell*, *143*(1), 99–110.

Xu, W., Parmar, G., & Tu, Z. (2019). Geometry-Aware End-to-End Skeleton Detection. In *Proc. of British Machine Vision Conference*, vol. 2.

Yasrab, R., Atkinson, J. A., Wells, D. M., French, A. P., Pridmore, T. P., & Pound, M. P. (2019). RootNav 2.0: Deep learning for automatic navigation of complex plant root architectures. *GigaScience*, *8*(11), giz123. DOI: 10.1093/gigascience/giz123.

Youn, J.-Y., Dyakov, B. J., Zhang, J., Knight, J. D., Vernon, R. M., Forman-Kay, J. D., & Gingras, A.-C. (2019). Properties of Stress Granule and P-Body Proteomes. *Molecular Cell*, *76*(2), 286–294. DOI: 10.1016/j.molcel.2019.09.014.

Yuste, R. (2015). From the neuron doctrine to neural networks. *Nature Reviews Neuroscience*, *16*(8), 487–497. DOI: 10.1038/nrn3962.

Zhang, C., Halsey, L. E., & Szymanski, D. B. (2011). The development and geometry of shape change in *Arabidopsis thaliana* cotyledon pavement cells. *BMC Plant Biology*, *11*(27), 1–13.

Zhao, K., Shen, W., Gao, S., Li, D., & Cheng, M.-M. (2018). Hi-Fi: Hierarchical Feature Integration for Skeleton Detection. In *Proc. of Int. Joint Conference on Artificial Intelligence*, (pp. 1191–1197). DOI: 10.24963/ijcai.2018/166.

Zimmer, C., & Olivo-Marin, J.-C. (2005). Coupled parametric active contours. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(11), 1838–1842.

Chapter 2

Keypoints, Spots, and Particles

Birgit Möller
Stefan Posch

Institute of Computer Science, Martin-Luther-University Halle-Wittenberg, Halle, Germany

Research Article

Robust features for 2-DE gel image registration

Proteomics is a rapidly growing field of modern biology. Since quantitative data of proteins involved in dynamic processes of living organisms are essential for understanding the basics of life, techniques like 2-DE and related procedures for automatic data interpretation are at the heart of this research field. They are strongly required to enable analysis and interpretation of the emerging amount of available data. Analyzing and interpreting gel image data usually requires the comparison of gels from different experiments and, thus, a prior registration of gels. This can be accomplished using featureless, feature-based or hybrid registration approaches combining both techniques. Recently, the latter ones have shown high performance, and it is undoubtful that in general robust and reliable features are an essential ingredient and valuable source of information for high-quality image registration. In this paper we provide a thorough overview and elaborate analysis of the capabilities of available feature detectors for gel image registration. Particularly, a detailed and extensive comparative study is presented where common spot-specific detectors are included as well as image-content independent detectors that were not applied to the task of gel image registration until now. The study incorporates tests on several thousand synthetically deformed images from different experimental conditions. As a result it provides valuable quantitative data allowing for direct objective comparisons of various detectors, and is well suited to guide the design of new registration algorithms.

1 Introduction

1.1 General

Proteomics extends the analysis of the static genome toward understanding the far more complex and dynamic proteome of cells, tissues and entire organisms, and for the first time enables a systems view of biology [1]. To separate proteins from cell or tissue samples 2-DE has been established over the last two decades as the *de facto* experimental technique. To assess and compare protein quantification and differences from varying experimental conditions and technical or biological replicates, it is essential to account for variations

Correspondence: Dr. Birgit Möller, Institute of Computer Science, Martin-Luther-University Halle-Wittenberg Von-Seckendorff-Platz 1, 06120 Halle/Saale, Germany
E-mail: birgit.moeller@informatik.uni-halle.de
Fax: +49-345-5527039

Abbreviations: CPT, centerpull transformation; **PBLT,** piecewise bilinear transformation; **SIFT,** scale invariant feature transform; **SURF,** speeded-up robust feature; **TPS,** thin plate spline

and distortions between gels and resulting gel images. To assist this analysis, the automatic processing of gel images is of large interest, especially as the amount of gel data available increases substantially [2]. For differential analysis, gel images have to be compared and, thus, usually pairs of gel images are to be registered in advance. As deviations between gel images are as well global as also local in nature, non-rigid transformations have to be applied.

Techniques for registration are classified into feature-less and feature-based approaches, where combinations were also proposed [3]. For feature-less methods, the intensity information of the images is exploited directly (see, *e.g.*, [4]). On the contrary, features in both images are detected for feature-based registration. These are subsequently matched between images, and the resulting matches are employed to determine a suitable transformation for registration. Results and quality of these approaches obviously depend on the amount, spatial distribution, and localization accuracy of features used for matching. For registration of gel images, protein spots have typically been used as features (*e.g.* [5–7]) as they are detected anyway to identify proteins. However, for the registration process there is no need to restrict potential types of features to spots, especially as even commercial programs show large variance

in spot detection [8]. On the other hand, in computer vision literature several feature detectors have been proposed, which are independent of specific image structures or contents and, hence, potentially offer larger flexibility for the task of gel image registration.

In this work we aim at assessing the appropriateness of different feature detectors as basis for feature-based registration of gel images, particularly focusing on a thorough comparison of common gel-specific spot detectors and image-content independent keypoint detectors. The work extends a previous study [9] and includes four spot detectors, namely the Laplace, Ring, Watershed, and Meaningful Boundaries detector and two general-purpose keypoint detectors, Scale Invariant Feature Transform (SIFT) and Speeded-Up Robust Features (SURF), which are widely used for various image analysis tasks. As transformation classes for simulating gel image distortions we consider piecewise bilinear transformations (PBLT), thin plate splines (TPS), and the centerpull transformation (CPT), as these have previously been used to register gel images. We investigate the characteristics of the various feature detectors as well as the reproducibility of features detected, taking number, localization, and robustness into account. The results of this comparative study are expected to yield valuable guidance to choose suitable feature types and detectors for robust and accurate registration of gel images.

1.2 Related work

Automatic registration of gel images has been an area of active research for a long time. A large amount of different approaches, either feature-less or feature-based, and also hybrid combinations have been proposed to solve the problem of gel image alignment [7, 10, 11]. Nevertheless, the task still poses challenges, which even up-to-date commercial products often do not meet in a satisfactory way [8, 12].

Robust features are obviously required for both feature-based and combined techniques [7, 13]. Particularly, the undesirable task of manual landmark selection and warp improvement, which still takes a considerable amount of total processing time for a single pair of gel images [12], is to be eliminated in near future.

For gel image analysis, protein spots are domain-inherent striking image patterns, *i.e.* features that appear to yield a perfect base for gel registration. There is a large variety of common techniques for detecting spots in literature, like intensity level and flow analysis [14, 15], Laplacians [6], watersheds [5, 7], morphological operators [16], or parametric spot models like 2-D Gaussians [17]. Also, more complex algorithms have been proposed, *e.g.*, based on Markov Random Fields [18]. Although these spot-like structures appear self-evident for the registration task, there are also some serious drawbacks associated with them. Their detection often seems to be tricky, not robust even in cases of only small image variation [8] and still requires a lot

of manual intervention and parameter adjustment (*e.g.*, [14]). Additionally, using spots as features puts unnecessary restrictions on the number and characteristics of available features *per* image as well as to their spatial distribution.

In general, robust and reliable feature detection for registration tasks calls for a more or less uniform feature distribution over the entire image as well as a small amount of mismatched correspondences. Obviously, the quality of feature matches directly correlates with the quality of the overall registration result. Moreover, already the robustness of the features themselves sets the course for all later processing steps.

The challenge of identifying robust features for image registration is not specific to the field of gel images, but also appears in other computer vision applications, such as camera motion recovery [19], mosaicing [20], or robot navigation [21]. However, here often no assumptions about specific scene contents can be made, which raises the need for robust and flexible features that are independent of a concrete scenario. Flexible *keypoint* detectors are subject to a detailed study in [22]. The study shows that among detectors based on image derivatives (Hessian or moment matrices), Harris corners [23], and SUSAN [24], the more recently published scale invariant detectors, the SIFT [25] and SURF [26], are superior in a general context. Compared to explicit spot detectors they show larger flexibility as they are not confined to single image entities, but also consider larger compositions of significant structure. Particularly, their scale invariance allows for the extraction of characteristic intensity configurations on larger scales, *e.g.*, striking intensity distributions in images. With regard to gel image registration this opens perspectives to increase the total number of features as well as toward a more uniform distribution. Thereby, image-content independent detectors may contribute valuable improvements regarding the automatic detection of stable features in feature-based gel registration.

2 Materials and methods

2.1 Feature detectors

In this work we compare four traditional spot detectors and two recently published general-purpose keypoint detectors. To this end, quantitative evaluations were conducted using synthetically deformed real gel images as outlined below.

2.1.1 Spot detectors

In literature, a large amount of different spot-specific feature detectors can be found. The most frequently used techniques exploit typical spot properties like convexity (Laplace, Watershed) or elliptical shape (Ring Detector, Watershed). The Meaningful Boundaries approach relies on statistics over image contour information.

2.1.1.1 Laplace detector

The probably most striking property of protein spots is their convexity. Accordingly, in various works the shape of spots is explicitly modelled by Gaussians [27]. The Laplace Detector (*e.g.*, [6, 28]) is also tuned for convex spot detection; however, it relies on a less restrictive spot model. It solely assumes that spot centers are given by local image locations with significant local curvature. Accordingly, the Laplace Detector locates possible spot positions ($S(x, y) = 1$) by searching for positive values in second order derivatives, approximated in terms of the Laplace images $\partial_x^2 I$ and $\partial_y^2 I$ in x and y direction:

$$S(x, y) = \begin{cases} 1, & \text{if } (\partial_x^2 I(x, y) > 0) \wedge \partial_y^2 I(x, y) > 0) \\ 0, & \text{otherwise} \end{cases}$$

Convex spot positions are usually not isolated, hence connected components are extracted from the binary image S. Finally, the centroids of components exceeding a certain minimum size of several pixels are returned as final spot locations. To detect a specified number of n spots, the intensities of all centroid pixels are ranked by increasing values, and at least the first n entries of the list are returned as final detection result.

2.1.1.2 Ring and ellipse operators

Although there are types of gel images where protein spots do not show the typical circular or elliptical shape, this is a reasonable and justified assumption for a large amount of experimental data. The Ring operator proposed in [29] is based on this assumption and detects image locations showing an elliptical shape, with the inner parts of the ellipse being darker than the outer ones. Toward this goal, the detector first thresholds image intensity values $I(x, y)$ and gradient magnitudes $|\nabla I(x, y)|$,

$$f(x, y) = \begin{cases} 1, & \text{if } I(x, y) < t_I \\ 0, & \text{otherwise} \end{cases}$$

and

$$g(x, y) = \begin{cases} 1, & \text{if } |\nabla I(x, y)| < t_G \\ 0, & \text{otherwise} \end{cases}$$

and defines a set of pixels in homogeneous image regions with low intensity $P = \{(x, y) | f(x, y) \cdot g(x, y) = 1\}$, as potential spot center candidates. The thresholds t_I and t_G are determined applying Otsu's method [30]. To increase robustness with respect to the amount of spots present in the gels, contrary to [29] we use only pixels located near object boundaries to determine these thresholds. These pixels are selected by thresholding local Laplacians.

In a next step, for each $(x, y) \in P$ an elliptical target region centered at (x, y) is constructed as a potential spot region. It is formed by the elliptical region $C_{x,y}$ and its outer ring $R_{x,y}$:

$$C_{x,y} = \{(u, v) | (u - x)^2 + (v - y)^2 / \alpha^2 \le r_M^2\}$$
$$R_{x,y} = \{(u, v) | r_m^2 \le (u - x)^2 + (v - y)^2 / \alpha^2 \le r_M^2\}$$

If the outer ring region $R_{x,y}$ contains larger intensities than the inner part of $C_{x,y}$, its minimum value will be larger than the minimum of $C_{x,y}$. Consequently

$$S(x, y) = \min_{(u,v) \in R_{x,y}} I(u, v) - \min_{(u,v) \in C_{x,y}} I(u, v),$$

will be positive and a spot region is hypothesized. All resulting candidate positions are finally included in a component labeling step and their centroids again yield the final spot positions. Since the approach works fully automatic, there is no obvious way to specify the number of spots to be detected.

2.1.1.3 Watershed detector

A more sophisticated and widely used approach for spot detection is based on the watershed transformation [5, 7]. Its main idea is borrowed from geography and defines an algorithm for segmenting a landscape into valleys and ridges to analyze its overall topology [31]. It may be implemented by flooding the landscape, which results in catchment basins separated by ridges, the so-called watersheds. In our experiments we adopt the approach of [5] where the watershed transform is applied to the gradient magnitude image of a gel image. Thus, spots and homogeneous background regions are associated with the catchment basins while their borders form the watersheds. However, the watershed transformation tends to over-segmentation and therefore some post-processing steps are required to extract the desired subsets of meaningful watershed regions representing protein spots.

The initial catchment basins are subjected to thresholding – one threshold θ_I for spot intensities and one for relative differences θ_D, as spot regions are assumed to be either of low intensity or darker than all their neighbors. Subsequently, the convexity $C(r)$ of each resulting region r is required to be larger than a threshold θ_c. $C(r)$ is calculated summing second order partial derivatives of all pixels of r,

$$C(r) = \sum_{(x,y) \in r} \partial_x^2 I(x, y) + \partial_y^2 I(x, y)$$

The final post-processing step merges partial regions belonging to one single spot. It requires the local curvature along the common border of a pair of partial regions to be convex, and each resulting spot to be of approximately elliptical shape.

To verify the elliptical shape of a region r, an ellipse E_r is fit to the bounding rectangle of r and the difference between the contour cont(r) of r and E_r is calculated. Each contour pixel b of the region is projected onto E_r along a ray passing through the center of the ellipse, yielding $p^E(b)$. The ellipticity of r is associated with the overall distance between contour points and their projections, *i.e.* with a small approximation error

$$\varepsilon_{\text{mse}} = \frac{1}{|\text{cont}(r)|} \sum_{b \in \text{cont}(r)} (b - p^E(b))^2$$

A merge of adjacent partial regions is only accomplished if the ellipticity of the merged region is larger than the ellipticities of all partial regions.

The final result of the whole spot detection approach is given by the set of all spot regions, which either result from the merging process or are directly detected, as they fulfill all criteria and have no neighbor. The position of every spot region is defined by its center-of-mass as calculated on smoothed gel images. To specify the total number of spots to be returned, the thresholds θ_I and θ_D can be adjusted. As θ_I showed larger impact in our experiments, detector adaptation was done by properly adjusting this threshold.

2.1.1.4 Level lines and meaningful boundaries

The approach of Meaningful Boundaries [32] builds on the assumption that spot regions are constituted by low intensity values and exhibit a considerable contrast at their boundaries. Conceptually, a suitable local threshold is applied to the image to detect spots as regions exhibiting a large contrast along their boundaries. Additional constraints on these boundaries are quite relaxed requiring a certain compactness and minimal size.

To implement this idea, all potential threshold levels λ are applied to the image yielding the lower level sets

$$\Omega_\lambda = \{(x, y) | I(x, y) \leq \lambda\}$$

In each level set, the connected components are detected as potential spot regions and their boundaries computed, called level lines. Level lines monotonically grow with increasing levels and eventually merge to form larger regions. Thus, the set of all level lines constitutes a hierarchy. To select valid spots from all these hypotheses, the Helmholtz principle [33] is employed, which judges the significance of a hypothesis with respect to randomly generated contours. Specifically, the probability of a random contour with at least as large contrast as observed for the level line in question is computed. It can be considered as the *p*-value under the null hypothesis of random contours. For a level line *c* with length l_c and minimal contrast $|\nabla_{\min} I(x, y)|$ this probability is defined as

$$P(|\nabla I(x, y)| \geq |\nabla_{\min} I(x, y)|)^{\frac{l_c}{2}}$$

A level line is assumed significant if this *p*-value is below a threshold ε. The distribution of gradient magnitudes is estimated as the empirical distribution observed in the image. As a weak constraint on the shape of the spot its compactness l_c/a_c and in addition to [32] the area a_c above specified thresholds is checked. If we consider the ideal boundary of a spot, which fulfills these criteria, usually level lines of the adjacent levels will also be significant, yielding multiple boundaries for one spot. This is resolved choosing the level line with minimal *p*-value within each set of adjacent significant level lines. To avoid boundaries corresponding to several spots, in contrast to [32] we consider only level lines in the hierarchy below the first merge of

other level lines, *i.e.* from the leaves of the tree of level sets to the first inner node. Finally, the overall number of detected spot contours mainly depends on the choice of ε, thus, in our experiments adjusting the number of spots detected was done varying the values of ε.

2.1.2 Keypoint detectors

For applications in varying problem domains or in situations, where no assumption about scene contents or structures can be made, highly flexible and invariant feature detectors are required. Two such scale invariant keypoint detectors are SIFT [25] and SURF [26]. They have proven their general applicability in various scenarios. However, until now they have only rarely been applied to the task of non-rigid registration, and in particular, to the registration of images with specific structural patterns like gel images. Accordingly, our study yields new quantitative data about their robustness and efficiency with regard to the domain of 2-D gel images, and more generally regarding non-rigid transformations where no systematic surveys have been done until now.

2.1.2.1 SIFT

The basic concept of SIFT [25] is a thorough analysis of image characteristics in scale space. The scale of an image is associated with its resolution, and it has been proven that difference images between different scale space levels yield valuable information for robust and reliable re-detectable features. Different scales are acquired convolving the input image with Gaussian kernels $G_\sigma(x, y)$,

$$I_\sigma(x, y) = G_\sigma(x, y) * I(x, y)$$

and subsequent down-sampling. The standard deviation σ of neighboring scales differs by the same constant factor *k* for all scales. Keypoints are then given by local extrema in difference images $D_\sigma(x, y)$ between two scales:

$$D_\sigma(x, y) = I_{k\sigma}(x, y) - I_\sigma(x, y)$$

For extrema detection, the difference value $D_\sigma(x, y)$ of each point (x, y, σ) in scale space is compared to all neighbors in a $3 \times 3 \times 3$ neighborhood. By fitting a 3-D quadratic function to the local point, the extremum is localized with sub-pixel accuracy. In addition, the interpolated difference value D_σ at the extremum allows for ranking of the keypoints detected. With increase of D_σ also the local contrast at the keypoint position increases. Thus, ranking all detected keypoints according to D_σ allows to select a specified number of keypoints with highest contrast. In our experiments we use a threshold for D_σ to adapt the SIFT detector's feature numbers.

2.1.2.2 SURF

As the SIFT approach is computationally expensive SURF aims at high efficiency combined with comparable stability

of features [26]. SURF is based on an analysis of local Hessian matrices $H(x, y, \sigma)$ over various scales:

$$H(x, y, \sigma) = \begin{bmatrix} L_{xx}(x, y, \sigma) & L_{xy}(x, y, \sigma) \\ L_{yx}(x, y, \sigma) & L_{yy}(x, y, \sigma) \end{bmatrix}$$

where L are the results of convolving the input image with second order Gaussian derivatives. For efficiency reasons the entries of the matrix are calculated only approximately applying discrete box filters as approximations to the Gaussian derivative kernels. In addition, the images are not re-sampled within a pyramid, but detection results for various scales are produced by simply applying differently sized filters to the input image.

SURF keypoints are finally given by maximal determinant values of local Hessian matrices, detected by performing a non-maximum suppression in a $3 \times 3 \times 3$ neighborhood of each point and interpolating in scale and space, like in the SIFT approach. Also, the ranking of keypoints is similar to SIFT, *i.e.* to detect a specified number of feature points, the determinant values can be thresholded properly.

For SIFT and SURF we used publically available software packages, *i.e.* the free C++ implementation of SIFT by A. Vedaldi (http://vision.ucla.edu/~vedaldi/code/siftpp/siftpp.html) and the original SURF library provided by its authors (http://www.vision.ee.ethz.ch/~surf/). All other detectors were re-implemented by us.

2.2 Image datasets

The image characteristics of 2-DE gel images show a large variation, mainly depending on the applied staining or fluorescence technique and the overall experimental conditions. Accordingly, within a comparative study of feature detectors of course not all existing types of gel images can be included, but rather a representative sample of gel images has to be selected for meaningful quantitative results.

2.2.1 Original gels

To achieve this goal we have used three different image datasets during our experiments resembling a wide range of different degrees of image quality as well as technical setups. The first set denoted 'LECB' results from a selection of gels from the LECB 2D PAGE Gel Images Data Sets [34], freely available for public use (http://www.lecb.ncifcrf.gov/2DgelDataSets/). The two other sets, termed 'UCD-DIGE' and 'UCD-Stain,' were sampled from the UCD test data used in [35] and are also publically available.

The LECB data is more than 20 years old. However, as it has been used in several publications as freely available reference (*e.g.*, [32]), we included a subset of the data into our experiments. We used the Human leukemias data set as base for our tests. It contains 170 images in total, each

image sized 512×512 pixels in 8-bit GIF format, from which we selected 36 representative sample images. Since the amount of structure, *i.e.* the number of spots in the images, varies significantly, and since the performance of at least some of the detectors is to be expected to significantly depend on the amount of available structure, this selection was done based on a quantification of image structure. Each image was partitioned into non-overlapping blocks of 32×32 pixels, for each block the entropy was computed and averaged for each image.

Finally, the 12 images with lowest entropy (class 'LECB-C_0'), 12 images with medium entropy (class 'LECB-C_1') and the 12 images with largest entropy (class 'LECB-C_2') were chosen for the experiments. All images were converted to PGM format and automatically cropped given the annotated valid spot areas within the gels as specified in the complementary description files. Since not all area specifications were accurate and sometimes artifacts remained at the border of images, some images required manual post-processing (cropping, filling of spurious white regions with local background color).

The original UCD DIGE data [35] consists of 13 DIGE gel images resulting from analysis of brain proteins of female mice at three growth stages where each image includes three-color channels. For our experiments we selected the 13 single channel images fluorescently labeled with Cy3, which form the data set 'UCD-Dige.' Finally, our third set 'UCD-Stain' resembles a collection of 14 images selected from the UCD Silver Stain data [35] originating from a study of human brain proteins in mental disorder compared to a normal mental state control set. We selected all images except the master gel from the control set.

Both UCD image collections include gel images of high image resolution (approx. 2000×2000) with 16-bit sampling. To reduce the computational burden during the experiments, the images were down-sampled to half of their original size and converted to 8-bit PGM images as pre-study gave only marginal differences in performance.

2.2.2 Synthetic image deformations

To compare the reproducibility of various feature detectors, pairs of images with known correct transformations are essential as ground truth. These are required to derive exact positions of corresponding features necessary for quantitative evaluation. As these transformations are hard if at all to acquire for real image pairs we chose to synthetically deform the original images in our dataset applying randomly sampled transformations.

2.2.2.1 Transformations

The type of transformation class potentially affects the robustness of feature detection and subsequent matching. Thus, we use three different classes, which previously have been used to model deformations of gel images, namely PBLT [36, 37], TPS [38], and CPT [7].

A PBLT is defined on a rectangular lattice of squares yielding tie points $\vec{x}_{ij}$. The transformation $T(\vec{x})$ is specified by displacement vectors $\vec{u}_{ij}$ describing the distorted positions of the tie points. The induced transformation acting on a point $\vec{x} = (x, y)$ is given by the surrounding tie points and their displacements:

$$T(\vec{x}) = (1 - x)(1 - y)\,\vec{u}_{ij} + x(1 - y)\,\vec{u}_{i+1,j}$$
$$+ (1 - x)y\,\vec{u}_{i,j+1} + xy\,\vec{u}_{i+1,j+1}$$

where the coordinates $\vec{x}$ are given with respect to a local coordinate system for each square of the lattice with unit length and origin $\vec{x}_{ij}$.

For all our experiments we use a lattice with squares of size 50 pixels where each tie point is displaced with probability 0.5 by a randomly drawn displacement vector (see below).

The TPS transformation [39] employs linear combinations of radial basis functions $U(\vec{x}) = |\vec{x}|^2 \log |\vec{x}|^2$ to specify pixel displacements. The basis functions are centered at irregularly located centers $\vec{x}_n$. With each center weights $\vec{w}_n$ are associated. The resulting transformation is given as

$$T(\vec{x}) = \vec{a} + \vec{a}_x\,x + \vec{a}_y\,y + \sum_{i=0}^{n} \vec{w}_n\,U(|\,\vec{x}_n - \vec{x}\,|)$$

where the first three terms describe a common global affine transformation with translation $\vec{a}$ and a non-singular matrix $(\vec{a}_x\ \vec{a}_y)$. In the following we consider affine transforms without translational component and parameterize with rotation θ and shearing axis ϕ, and scaling factors λ_x, λ_y (see [19]). Note that a PBLT allows to model a global affine transformation implicitly. The number of centers to synthetically deform images is chosen as $1.5 \times 10^{-4} \cdot N$, where N is the number of pixels in the given image. On average this gives the same number of displaced centers as displaced tie points when applying the PBLT. In the following, a specific TPS is specified by a displacement for each center, and subsequently, the weights are determined minimizing the thin plate bending energy [39].

The CPT is inspired by a hypothetical force acting vertically on the gel at a single point, which in reality might result from non-uniform gel pouring [7]. The location of this point is specified by a horizontal offset h_{off}, and the strength of the deformation force by a vertical offset v_{off} and the parabola width pw. The final transformation is defined as

$$T(\vec{x}) = \begin{pmatrix} x \\ y + ax^2 + bx + c \end{pmatrix}$$

where the parameters a, b and c of the parabola equation are derived from $h_{\mathrm{off}}, v_{\mathrm{off}}$, and pw.

2.2.2.2 Random deformations

In the following we construct five classes of image deformations $\sigma_D = 1, 2, 3, 4, 5$ with increasing amount of distortion for each of the three transformation models.

Displacement vectors for tie points (PBLT) and centers (TPS) are drawn from a Gaussian distribution with zero mean and standard deviation σ_D as given by the class of deformation. The shearing angle ϕ is sampled uniformly from the interval $[0, \frac{\pi}{4}]$. The remaining parameters defining a transformation (see Table 1) were sampled according to the following schema. Consider a parameter α with neutral value Z (*i.e.* no distortion) and a maximum allowed deviation ΔZ from this neutral value. To sample α for deformation class σ_D we draw from a Gaussian with mean $Z + \frac{\sigma_D}{5}\Delta Z$ and standard deviation $\frac{1}{3}\frac{\Delta Z}{5}$. Thus, the admissible interval of parameter values is covered by the five deformation classes with small overlap (see Fig. 1). For a parameter with an admissible interval symmetric around Z and allowed deviation $\pm \Delta Z$ a mixture of two Gaussians with mean $Z \pm \frac{\sigma_D}{5}\Delta Z$ is used.

To simulate variations in the gray value structure, white noise was added to the interpolated intensities, which was sampled independently from a Gaussian distribution with standard deviation 5 (class 'LECB-*x*') resp. 1 (class 'UCD-*x*').

For each image we generated 10 randomly distorted images for each transformation class and each $\sigma_D \in \{1, 2, 3, 4, 5\}$. This results in a total of $63 \times 10 \times 3 \times 5 = 9450$ distorted images. For examples of distorted gels see Fig. 2, where TPS transformed images from the LECB data sets are shown.

2.3 Experimental evaluation

Comparing the robustness and efficiency of feature detectors is an important task in computer vision in general. To evaluate the reproducibility of features, the *repeatability score* Rs_r of a feature detector was introduced in [22]. It quantifies the probability of a feature in an undistorted

Table 1. Transformation parameters with neutral value (Z) and interval width ΔZ, where H and W give the height and width of the image

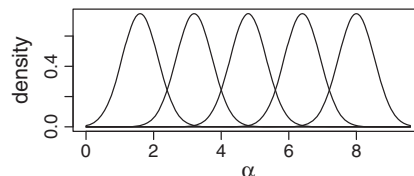|	θ	λ_x	λ_y	h_{off}	v_{off}	pw
Z	$0°$	1	1	0	0	W
ΔZ	$5°$	± 0.05	± 0.05	$\pm \frac{1}{4} W$	$\pm \frac{1}{8} H$	$\pm 0.3W$

Figure 1. The five Gaussian distributions used to sample parameter α with $Z = 0$ and $\Delta Z = 8$, one for each of the five deformation classes.

image I to be re-localized in a deformed version I_T of the same image with accuracy r:

$$\text{Rs}_r(I, I_T) = \frac{|\tilde{P}_I|}{|P_I|} \qquad (1)$$

with $\tilde{P}_I = \{\vec{x}_i \in P_I | \exists! \, \vec{x}_t \in P_T : |\vec{x}_t - T(\vec{x}_i)| \leq r\}$

P_I is the set of features $\vec{x}_i$ detected in the undistorted original gel image, P_T is the set of features $\vec{x}_t$ detected in the transformed image. $\tilde{P}_I$ is the subset of features in the original image that have exactly one counterpart in the transformed image with distance less or equal to r. It is safe to assume that this unique counterpart is indeed the correct corresponding feature for $\vec{x}_i$. Thus, for matching, exactly the features in $\tilde{P}_I$ may be matched to the correct transformed position with accuracy r or better. Otherwise, the corresponding feature has been detected with larger localization error, has not been detected at all, or multiple features have been detected within a circle of radius r around the correct transformed position. All these cases impede accurate registration of the image pair, as only few and/or inaccurate matches are available for registration. For the evaluation in this work we use $r = 1.5$ pixels. This value has already proven its suitability for the evaluation of feature detectors for rigid registration (*cf.* [22]). Summarizing the repeatability score is a good estimate for the fraction of features in P_I, which may be correctly matched with high location accuracy by an optimal matching process. In this sense it gives an upper limit for the performance of accurate matching of gel images, when matching is based on the features detected.

3 Results and discussion

Assessing and comparing the suitability of different feature detectors for a certain kind of registration problem is a difficult task. The performance of each detector heavily depends on the quality and characteristics of the image data as well as on parameter settings used during feature detection. Accordingly, these aspects have to be considered thoroughly for objective evaluation.

Below we discuss the results of a comparative study. Initially, we describe the specific characteristics of each feature detector, particularly, the total numbers of features detected on different kinds of images, which allows for a general assessment of detector performance. Subsequently, we
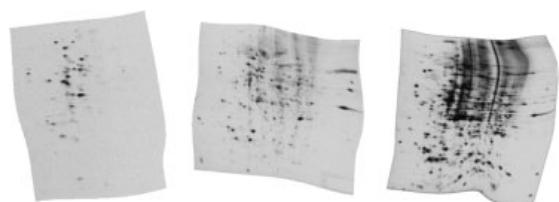
present a direct comparison of detectors based on repeatability scores as objective quantitative measures for overall detector quality. All discussions are based on representative example plots and figures for different detectors and image classes. Additional plots are available on the Supporting Information web page (http://www2.informatik.uni-halle.de/agprbio/AG/Publication/OnlineMaterial/xyz4711).

3.1 Detector characteristics and feature numbers

In general, feature detection results vary significantly with the characteristics of the detectors and also with different types of image sets. To illustrate the influence of image data characteristics we consider the ring detector (referred to as 'RING' below) as example detector that does not allow for easy parameter adjustment, but provides a reasonable set of default parameters. In Table 2 the average numbers of spots detected by the RING detector on all available sets of transformed images using the standard parameters as proposed in [29] are shown. For the LECB data set detected, spot numbers increase from set C_0 with low structure to C_2 with larger amounts of structure, which confirms the entropy-based categorization of the LECB data (see also Fig. 2). With regard to the UCD data sets the figures show that the UCD-Dige images contain considerably less structure compared to the UCD-Stain data. Indeed, UCD-Dige images show very low contrast and in large image areas they even appear nearly homogeneous, while in comparison in the UCD-Stain images spots tend to be quite pronounced (Fig. 3).

Table 2 proves the RING detector as quite insensitive to the different types of transformations applied to the images. Absolute numbers of features detected as well as repeatability scores (data not shown) are quite stable between different transformation classes. This observation holds also for the other five detectors evaluated (data not shown) and suggests that the suitability of these detectors for registrations of gel images is rather independent of the type of transformation employed to model gel deformations. Consequently, we restrict the following evaluation to one transformation class, namely TPS, as the presented results can be transferred to the other two types of transformations.

Figure 2. Example gels of the LECB datasets from classes C_0, C_1 and C_2 (left to right), each deformed by TPS with $\sigma_D = 5$.

Table 2. Average numbers of spots detected by the RING detector applied to all transformed images (covering all five levels of distortions σ_D) for each of the five image data sets

RING	LECB			UCD	
	C_0	C_1	C_2	Dige	Stain
TPS	23.30	78.07	246.26	5.90	56.41
PBLT	23.43	77.90	246.03	5.99	56.42
CPT	24.30	79.96	250.12	7.16	57.13

Next we turn to performance characteristics of all six detectors, *i.e.* RING, Laplace detector (LAPLACE), meaningful boundaries (LEVEL), watershed detector (WATER), SIFT, and SURF. For this experiment, parameters of the spot detectors LAPLACE, WATER, and LEVEL were adjusted to detect as many correct spots as possible, while restricting the number of apparently incorrect spots to an amount that seems reasonable. This approach is subjective to some extent, but appropriate to define a rough common reference for the detectors due to the lack of ground truth data. For the RING detector, this adjustment is not feasible, as it does not allow for an easy way to control the number of spots detected. Hence, it is always applied with its standard parameters in the following. In Fig. 4, top row, prototypical detection results for all six detectors are shown for part of an undistorted sample gel from the LECB-C_1 image set. Table 3 gives the numbers of features detected on average over all original images in the five data sets. Although LAPLACE, WATER, and LEVEL were all manually adjusted to detect a large number of correct spots, the results differ significantly as the number of apparently incorrect spots increases differently for the detectors. RING and LEVEL detectors give the lowest numbers of detected spots. Both detectors are highly specific and enforce quite strict constraints on possible spots. Particularly, the RING detector finds only the most pronounced spots and misses many spots with smaller contrast. The low figures for the LEVEL detector are due to the fact that it often does not separate closely adjacent spots with low contrast, but rather merges these into a larger common spot region. LAPLACE and WATER yield significantly more spots, and their results are to a certain degree comparable. The slightly higher numbers of WATER are due to its tendency to split a single spot into more than one region due to violation of the convexity criteria along the common border.

In contrast to the spot detectors, general-purpose keypoint detectors SIFT and SURF are not restricted to detect spots only, but keypoints may describe image structure in general. For the results shown in Table 3 and Fig. 4 we therefore applied these detectors with parameters to yield numbers of features comparable to LAPLACE and WATER. While the majority of keypoints detected is directly related to spot centers, in addition also keypoints located at the borders of spots or between spots are detected. In the latter cases the keypoints not only characterize single spots, but rather constellations of several spots (see Fig. 4). This property provides the content-independent detectors with the potential for a larger flexibility to represent image structure, which may be of advantage for subsequent matching and registration.

The difference in performance between the four spot detectors confirms observations reported in recently published comparative studies (*e.g.*, [8]) that different spot detectors result in a large variation for the number of spots detected for same gels. In our experiments the variations mainly relate to different underlying spot models and detection criteria of the various detectors. LEVEL, WATER, and RING rely on more or less strict constraints on the form

Figure 3. Crops of a UCD-Dige gel (left) and a UCD-Stain gel (right) showing the different image characteristics.
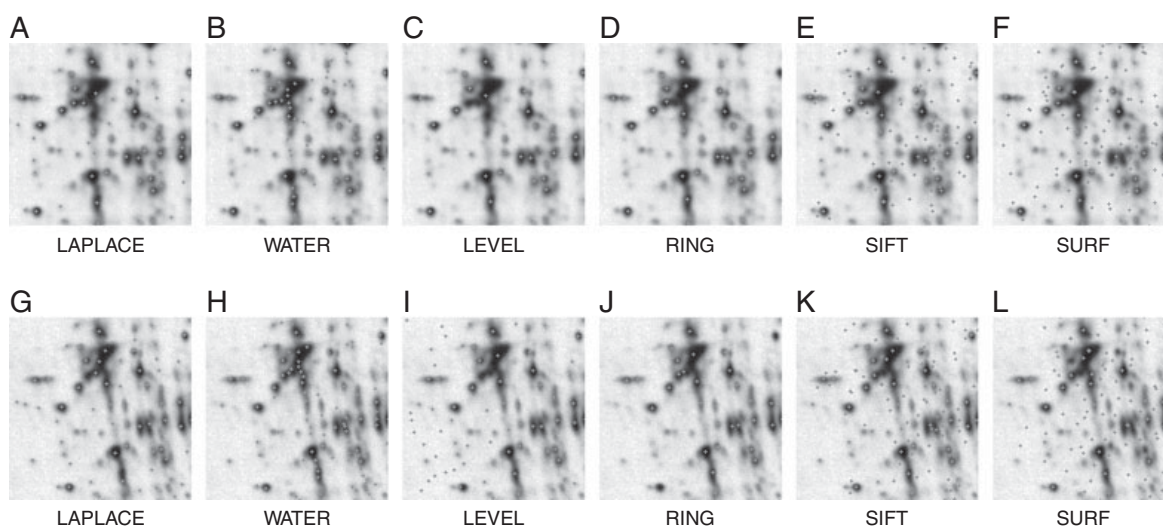

A	B	C	D	E	F
LAPLACE	WATER	LEVEL	RING	SIFT	SURF
G	H	I	J	K	L
LAPLACE	WATER	LEVEL	RING	SIFT	SURF

Figure 4. Prototypical comparison of detector results for part of a gel from set LECB-C_1. The top row shows results on the original image, the bottom row on a version deformed by TPS with $\sigma_D = 5$.

of spots and intensity characteristics, while LAPLACE assumes only some kind of convexity for spots.

In Table 4 the average numbers of detected features in deformed images are given for the TPS transformation and all data sets, averaging over all levels of deformation. The same parameter settings used for the numbers in Table 3 were applied. The numbers of spots detected in the transformed and original images are comparable in general. The LEVEL detector is the only exception, which shows this detector to be quite sensitive to noise added to the intensities. For the LECB sets, which exhibit a large dynamic range of intensity values and show high contrast, the numbers increase. On the contrary, for the UCD data with low contrast and a relatively moderate amount of structure, the number of spots detected decreases.

Finally, with regard to the UCD data there is another important aspect to mention. On both image sets, the RING detector yields only very few spots compared to all the other detectors. This is mainly due to the automatic threshold selection of the RING detector, which does not work well in cases where the background is dominating the image and foreground objects (here: spots) form only a very small fraction of all image pixels. This is an inherent problem of the RING detector and will impede acceptable results on certain kinds of gel images unless the automatic binarization stage is modified and made more flexible.

Table 3. Numbers of spots detected in the original images averaged over each image set[a]

	LECB			UCD	
	C_0	C_1	C_2	Dige	Stain
LAPLACE	105	202	503	311	407
WATER	114	291	802	266	427
LEVEL	70	174	225	185	298
SIFT	109	289	748	299	429
SURF	111	263	731	282	426
RING	22	75	236	8	58

a) Parameters were adjusted manually, see text.

Table 4. Numbers of spots detected in the transformed images, averaged over all distortion classes within each image set, applying the same parameter settings used for the original images shown in Table 3

	LECB			UCD	
	C_0	C_1	C_2	Dige	Stain
LAPLACE	106	202	503	315	417
WATER	170	309	865	234	444
LEVEL	222	444	288	95	217
SIFT	111	292	745	315	428
SURF	112	262	717	285	427
RING	23	78	246	6	56

In the second row of Fig. 4, exemplary detection results are shown visually for a deformed version of the sample gel, which was generated applying a TPS transformation with distortion level $\sigma_D = 5$. The results underline the formerly discussed feature numbers, *e.g.* the RING detector again locates only a small fraction of spots present; however, these are relocated very reliably. The same holds for WATER and LEVEL, besides the fact that both show a tendency to detect some more new spots in the transformed image in this case. This is true in general for the LECB image sets. Considering SIFT and SURF it should be noted that the formerly discussed flexibility in feature detection also shows in this example, as keypoints located between spots or at borders are robustly relocated to a large degree. The noise sensitivity of the LEVEL detector is also visible in this example, as many false spots are detected.

3.2 Detector comparison by repeatability scores

The numbers of spots detected give a first insight into the individual characteristics of the detectors. However, a more important property is the rate of spots that are re-detected in the transformed images with high localization accuracy, which is quantified with the repeatability score. In Fig. 5 repeatability scores for all detectors are shown for different amounts of distortion applied to the images. The same parameter settings as for the results in Tables 3 and 4 are used. As to be expected, the repeatability scores decrease with the increase of distortion. This is most pronounced for the SURF detector, where repeatability scores decline about 15% from distortion levels $\sigma_D = 1$ to $\sigma_D = 5$, almost independent of the image class. In general, however, this decrease is rather moderate with few percent in most cases. It can be concluded that, similar to the variation with regard to the class of transformations, the variation in performance due to different amounts of distortion is relatively small, except for the SURF detector. Since this is true in general, we summarize all distortion classes into average repeatability scores in the following discussions for conciseness. For the SURF detector these average numbers are subject to a larger variation between different distortion levels, which however does not invalidate the general performance assessment and validity of conclusions (see Supporting Information).

While the scores shown in Fig. 5 allow qualitative conclusions about detector robustness with regard to varying amounts of distortion, the absolute repeatability scores have to be interpreted carefully and at this stage do not allow for a direct comparison of detector performance. This is due to the fact that the numbers of features underlying the repeatability scores vary significantly, depending on the specific parameter settings used for each detector. In Fig. 6 the repeatability scores of all detectors for the image sets LECB-C_1, UCD-Stain, and UCD-Dige are shown for different numbers of features detected by varying parameter settings, which cover a reasonable range of feature numbers
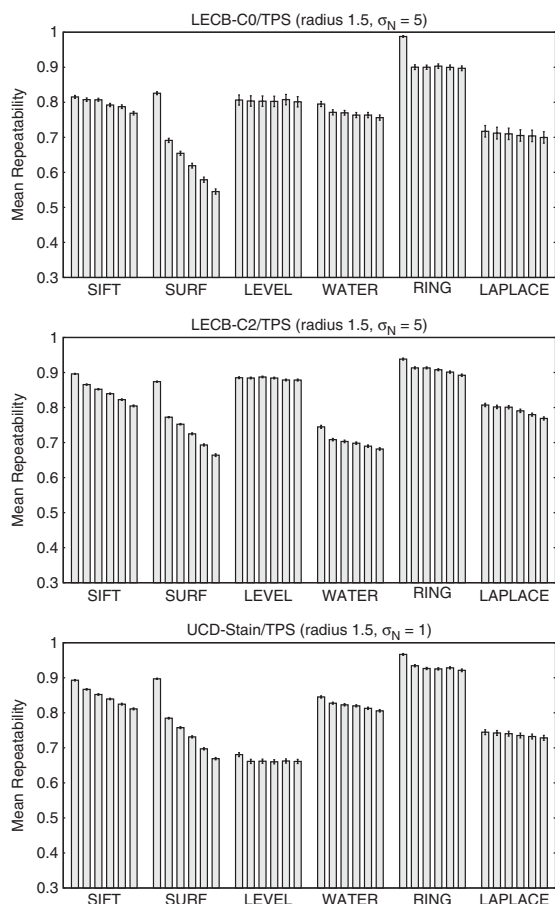
Figure 5. Repeatability scores *vs.* amount of distortion applying TPS transformations for image sets LECB-C_0 (top), LECB-C_2 (middle), and UCD-Stain (bottom). Scores are averaged over each distortion class in each image set and given for each detector for $\sigma_D = 1$ to $\sigma_D = 5$ from left to right. The error bars have an extent of ± 2 times the standard error and give an approximate 95% confidence interval.

Figure 6. Repeatability scores *vs.* numbers of features detected for image sets LECB-C_1 (top), UCD-Stain (middle), and UCD-Dige (bottom). Scores are averaged over all distortion classes applying TPS transformation. The error bars have an extent of ± 2 times the standard error and give an approximate 95% confidence interval.

for registration. Since for the RING detector adjustment of parameters is not easily possible, there is only a single point in each plot.

A general tendency of the curves is the decreasing repeatability scores with increasing numbers of features. There are two potential reasons for this behavior. First, an increase of feature numbers also increases the potential to detect multiple features within the radius *r* considered for the repeatability score, which in turn decreases the score. Second, if the detectors are required to return further features, these may not relate to the same spots (or spot configurations) in the original and transformed image. Thus, corresponding features are missing, which prevents relocalization and decrease the repeatability score. This effect is more severe as noise is added to intensities, and the LEVEL detector shows very sensitive to this problem on the low contrast UCD data sets. On the other hand, an increase
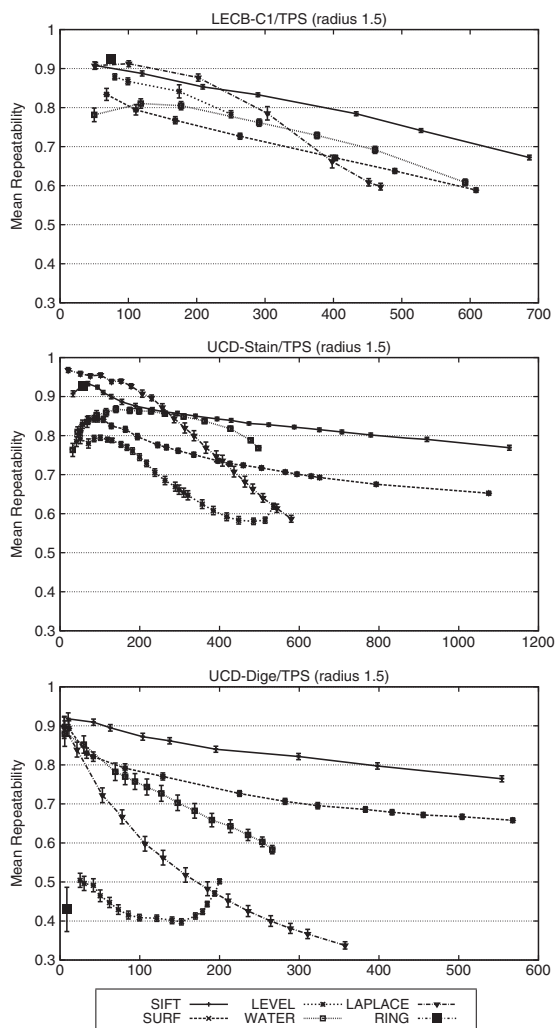
of features returned by a detector may also add corresponding features, which give the possibility for accurate relocation and increase of the repeatability score. This effect gives rise to increasing scores as the exceptions to the general trend.

For small numbers of features, most of the detectors have repeatability scores of around 80 to 95%. The performance of the LAPLACE detector shows the least robustness with regard to varying spot numbers as its scores quickly decrease to about 60%, on the UCD-Dige data to 34%. On the other hand, SIFT proves more stable with relatively high repeatability rates for all image sets, which are only moderately affected by the number of keypoints. While on

the LECB-C_1 images it decreases from 91 to 67%, on both UCD data sets the performance is constantly above 76%. Also, SURF shows satisfying scores for most cases; however, its repeatability scores are consistently about 10% below the scores of SIFT.

The performance of the WATER detector is comparable to the ones of SIFT and SURF for low numbers of features, always ranging between 75 and 85%. However, its drawback is its limited ability to provide higher numbers of spots, especially on the low contrast UCD-Dige data. Here, the score quickly decreases to 58%. The same disadvantage also holds for the RING detector. While on the LECB and UCD-Stain data its repeatability score is among the highest ones for comparable numbers of detected features, for the UCD-Dige data the performance is very weak. This renders the RING detector not a suitable alternative especially for low contrast images.

In summary, the answer to the question of which feature detector is best suited for gel image registration tasks is still not a general answer. A suitable detector must allow for large feature numbers as well as high repeatability scores, independent of applied image deformations. The number of features classical spot detectors can yield is limited by the number of spots in a gel image. Accordingly, in low-structured images their feature numbers will often be too small to allow for robust and accurate registration. Still, their robustness might render them suitable for estimating initial transformations based only on some few feature correspondences, which subsequently can be refined by data from other detectors. Image content-independent keypoint detectors also favor structured image regions for keypoint detection. However, as they are not restricted to spots, their larger flexibility and also scale-invariance allows for keypoints to be located elsewhere as well. Consequently, they can nearly be adjusted to yield any desired number of keypoints. And even if their repeatability scores decrease with increasing spot numbers, their scores are most of the time still superior to the ones that spot-specific detectors can achieve.

4 Concluding remarks

Pairwise registration of 2-DE gel images is a fundamental prerequisite for comparative proteome analysis. For feature-based and hybrid registration approaches robust and stable features detectable with high repeatability and localization accuracy are indispensable; however, valuable systematic evaluations and surveys of suitable detectors are rare.

The main contribution of this paper is a novel systematic qualitative and quantitative analysis of different feature detectors with regard to non-rigid gel image registration. Particularly and in contrast to earlier surveys, we not only consider spot-specific detectors but also include image-content independent detectors, not applied to gel image registration until now.

The basic outcomes of our study indicate a large variation in quantitative and qualitative detection results between spot-specific and general keypoint detectors, but also among spot-specific detectors themselves. In addition, repeatability scores as objective quality measure for detector performance identify general keypoint detectors as favorable compared to spot detectors with regard to feature number and robustness, yielding significantly larger numbers of features *per* image with a likewise higher repeatability. We assume that this mainly results from the keypoint detectors' ability to detect higher-order features like spot-constellations in addition to single spots. Since large numbers of stable features yield an important basis for robust correspondence detection and high-quality image registration, SIFT and SURF show advantages over conventional techniques and should no longer be ignored in the field of feature-based gel image registration.

5 References

[1] Speicher, D., *Proteome Analysis – Interpreting the Genome.* Elsevier, Amsterdam 2004.

[2] Dowsey, A., Dunn, M., Yang, G.-Z., *Proteomics* 2003, *3*, 1567–1596.

[3] Zitová, B., Flusser, J., *Image Vision Comput.* 2003, *21*, 977–1000.

[4] Wensch, J., Gerisch, A., Posch, S., *Image Vision Comput.* 2008, *26*, 1000–1011.

[5] Pleißner, K.-P., Hoffmann, F., Kriegel, K., Wenk, C., Wegner, S., Sahlstrm, A., Oswald, H. *et al.*, *Electrophoresis* 1999, *20*, 755–765.

[6] Rogers, M., Graham, J., Tong, R., 2D electrophoresis gel registration using point matching and local image-based refinement, *Proceedings of British Machine Vision Conference*, Kingston, UK, 2004.

[7] Srinark, T., Kambhamettu, C., *Electrophoresis* 2008, *29*, 706–715.

[8] Wheelock, Å. M., Buckpitt, A. R., *Electrophoresis* 2005, *26*, 4508–4520.

[9] Möller, B., Gress, O., Posch, S., A comparative study of robust feature detectors for 2D electrophoresis gel image registration, *Proceedings of German Conference on Bioinformatics*, LNI P-136, Dresden, Germany, 2008, pp. 138–147.

[10] Dowsey, A., Dunn, M. J., Yang, G.-Z., *Proteomics* 2006, *6*, 5030–5047.

[11] Sorzano, S., Thévenaz, P., Valdés, I., Beloso, A., Unser, M., Elastic image registration with applications to proteomics, *Proceedings of the 5th International Workshop on Information Optics* (*WIO'06*), Toledo, Spain, 2006, pp. 300–309.

[12] Clark, B., Gutstein, H., *Proteomics* 2008, *8*, 1197–1203.

[13] Rohr, K., Cathier, P., Wörz, S., Elastic registration of gel electrophoresis images based on landmarks and intensities, *International Symposium on Biomedical Imaging: Nano to Macro*, vol. 2, 2004, pp. 1451–1454.

[14] Cutler, P., Heald, G., White, I., Ruan, J., *Proteomics* 2003, *3*, 392–401.

[15] Langella, O., Zivy, M., *Proteomics* 2008, *8*, 4914–4918.

[16] Conradsen, K., Pedersen, J., *Biometrics* 1992, *48*, 1273–1287.

[17] Lemkin, P., GELLAB-II, A workstation based 2D electrophoresis gel analysis system, *Proceedings of 2D Electrophoresis*, VCH Press, Weinheim 1989, pp. 52–57.

[18] Baker, M., Busse, H., Vogt, M., An automatic registration and segmentation algorithm for multiple electrophoresis images, *Medical Imaging*, 2000, pp. 426–436.

[19] Hartley, R., Zisserman, A., *Multiple View Geometry in Computer Vision*, 2nd edn., Cambridge University Press, Cambridge 2004.

[20] Capel, D., *Image Mosaicing and Super-resolution*, Springer, Berlin 2004.

[21] Gracias, N., Zwaan, S., Bernardino, A., Santos-Victor, J., *IEEE J. Oceanic Eng.* 2003, *28*, 609–624.

[22] Mikolajczyk, K., Schmid, C., *Int. J. Comput. Vision* 2004, *60*, 63–86.

[23] Harris, C., Stephens, M., A combined corner and edge detector, *Proceedings of Alvey Vision Conference*, Manchester, England 1988, pp. 147–151.

[24] Smith, S., Brady, J., *Int. J. Comp. Vision* 1997, *23*, 45–78.

[25] Lowe, D., *Int. J. Comp. Vision* 2004, *60*, 91–110.

[26] Bay, H., Tuytelaars, T., van Gool, L., SURF: speeded up robust features, *Proceedings of European Conference on Computer Vision*, I, 2006, pp. 404–417.

[27] Garrels, J., *J. Biol. Chem.* 1989, *264*, 5269–5282.

[28] Rogers, M., Graham, M., *IEEE Trans. Image Processing* 2007, *16*, 624–635.

[29] Watanabe, Y., Takahashi, K., Nakazawa, M., Automated detection and matching of spots in autoradiogram images of two-dimensional electrophoresis for high-speed genome scanning, *International Conference on Image Processing*, vol. 3, 1997, pp. 496–499.

[30] Otsu, N., *Trans. Syst. Man Cybern.* 1979, *9*, 62–66.

[31] Roerdink, J., Meijster, A., *Fundamenta Informaticae* 2001, *41*, 187–228.

[32] Almansa, A., Gerschuni, M., Pardo, A., Preciozziand, J., Processing of 2D electrophoresis gels, *International Workshop on Computer Vision Application for Developing Regions (ICCV)*, 2007.

[33] Desolneux, A., Moisan, L., Morel, J., *Mathem. Imag. Vision* 2001, *14*, 271–284.

[34] Lester, E., Lemkin, P., Lipkin, L., *Ann. N Y Acad. Sci.* 1984, *428*, 158–172.

[35] Dowsey, A., Dunn, M., Yang, G.-Z., *Bioinformatics* 2008, *24*, 950–957.

[36] Veeser, S., Dunn, M. J., Yang, G.-Z., *Proteomics* 2001, *1*, 856–870.

[37] Salmi, J., Aillokallio, T., Westerholm, J., Griese, M., Rosengren, A., Nyman, T. A., Lahesmaa, R., Nevalainen, O., *Proteomics* 2002, *2*, 1504–1515.

[38] Pedersen, L., Analysis of two-dimensional electrophoresis gel images. *PhD thesis*, Informatics and Mathematical Modelling, Technical University of Denmark, 2002.

[39] Bookstein, F. L., *IEEE Trans. Pattern Anal. Mach. Intell.* 1989, *11*, 567–585.

Scale-adaptive Wavelet-based Particle Detection in Microscopy Images

Oliver Greß[1], Birgit Möller[1], Nadine Stöhr[2], Stefan Hüttelmaier[2], Stefan Posch[1]

[1]Institute of Computer Science, Martin Luther University Halle-Wittenberg
[2]ZAMED, Martin Luther University Halle-Wittenberg
oliver.gress@informatik.uni-halle.de

Abstract. Stress granules and processing bodies play a major role in analysing the physiology of cells under various environmental conditions. We present a fully automatic approach to detect such particles in fluorescence labeled microscope images. The detection is based on scale-adaptive analysis of wavelet coefficients allowing for an accurate detection of particles with a large variety in size. Results on real images illustrate the appropriateness of our approach and proof high quality.

1 Introduction

Systems biology on the cellular level requires detailed analysis of different particles in cells like stress granules (SGs) and processing bodies (PBs). They are suggested to be dynamically linked to places of mRNA sorting and storage or degradation [1]. To understand and clarify the physiological roles of SGs and PBs, it is important to investigate the alterations of number, size, shape or contacts of these particles under different physiological conditions. Consequently the fully automatic detection of SGs and PBs is an essential tool to gain deeper insights into their biological role and function.

To detect spot-like particles, global and local thresholding techniques are still used in microscopy [2, 3]. Further techniques include h-dome transform followed by clustering [4]. Level set methods [5] on the other hand assume an approximately constant grey level shared by all target entities which typically does not hold for microscopy images. In [6, 7] a method to detect particles in microscopy images based on wavelet coefficients is proposed, but best-suited to detect particles with limited variation in size. Wavelet-based approaches are assumed to be superior to Fourier-based ones as their basis functions have local support.

We present an extension to the approach in [6, 7] aiming at the variable size of target entities by automatic scale-adaptation and test-based hypothesis selection.

2 Materials and Methods

The method in [6] is based on multi-scale analysis of wavelet coefficients. The original image $I_0(x, y)$ is recursively smoothed yielding images $I_1(x, y), \ldots,$

$I_S(x, y)$. Wavelet coefficients $W_s(x, y)$ are derived as

$$W_s(x, y) = I_s(x, y) - I_{s-1}(x, y), \quad s \in \{1, \ldots S\} \tag{1}$$

For denoising, the amplitude-scale-invariant Bayes estimator [8] is applied, yielding adjusted coefficients $\tilde{W}_s(x, y)$. Features are represented by wavelet coefficients of adjacent scales. As the coefficients are correlated across scales due to the nature of the wavelet transform applied, adjacent scales are combined to a correlation image

$$c_{[a,b]}(x, y) = \prod_{s=a}^{b} \tilde{W}_s(x, y) \tag{2}$$

This correlation image is globally thresholded and the resulting connected components yield the final particles detected. The interval of scales $[a, b]$ used to correlate the wavelet coefficients defines the scales at which the particles of interest are represented. If all particles share the same characteristics, one interval is appropriate. In other cases one single interval either includes irrelevant scales or excludes important ones, in both cases often yielding incorrect particle size or shape, or missing particles.

To overcome these shortcomings we propose a new scale-adaptive technique which applies the wavelet-based segmentation to a set of – usually overlapping – intervals $[a_n, b_n]$, corresponding to different scales of particles. If for a target particle an appropriate interval is available, it is usually correctly detected in this interval. However, in many cases the particle is also found in adjacent intervals with incorrect size or shape (Fig. 1). This results in the necessity to select the correct one from overlapping and, thus, competing particle hypotheses. For this we propose an approach based on statistical hypothesis testing.

Typically hypotheses from the correlation image of a coarse scale result in larger regions than hypotheses from finer scales. Additionally, regions can be split up in finer scales due to the presence of multiple smaller particles or varying gray values inside larger particles.

We assemble competing hypotheses in trees whose nodes correspond to the particle regions detected in different intervals. The trees are build bottom-up
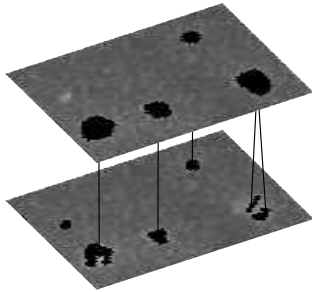

Fig. 1. Segmentation for two adjacent intervals (top: coarse, bottom: fine) and resulting hypotheses trees.
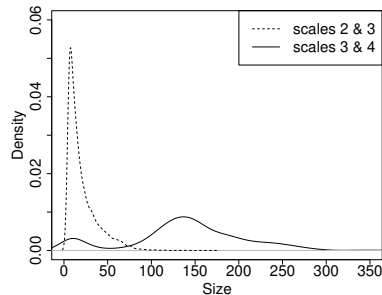
Fig. 2. Distribution of granule size detected in intervals $[a_1, b_1]$ and $[a_2, b_2]$.

starting with particle hypotheses from the finest scale. At each coarser level edges are inserted between a region and all overlapping regions of the next finer scale. Fig. 1 gives an example of the resulting hypotheses trees for two intervals.

In principle, two different regions at a coarse scale may both overlap with the same region at the next finer scale, destroying the tree characteristics of the hypothesis graph. However, this happened only once for the whole test data. Such constellations can for example be resolved using a simple criterion of largest overlap. The selection of one hypothesis out of several competing ones is accomplished using again a bottom-up procedure, starting from the fine scale leaves. Parent nodes are compared with their children and inferior nodes are deleted.

For comparison we employ the concept of meaningful events ([9]). This concept is tightly related to statistical hypothesis testing. In our case the null hypothesis H_0 models the case where no real particle is present at the location to be analyzed, rather a particle was detected due to noise or chance. To compute the likelihood $P(F_i \mid H_0)$ of a particle F_i detected under H_0, the observations at all pixels are assumed to be pairwise independent

$$P(F_i \mid H_0) = \prod_{(x,y) \in F_i} P\left(C_{[a_n,b_n]}(x,y) = c_{[a_n,b_n]}(x,y) \mid H_0 \right) \qquad (3)$$

where $C_{[a_n,b_n]}(x,y)$ are random variables modelling the correlation value observed at position (x,y). $P\left(C_{[a_n,b_n]}(x,y) = c_{[a_n,b_n]}(x,y) \mid H_0 \right)$ is the probability to observe the value $c_{[a_n,b_n]}$ at location (x,y) due to noise. Following [10] we estimate $P\left(C_{[a_n,b_n]}(x,y) \mid H_0 \right)$ as the discretized histogram of the complete correlation image for interval $[a_n, b_n]$.

The p-value $p(F_i)$ of F_i is the probability to observe a particle under the null hypothesis with correlation values at least as extreme as the ones of F_i. I.e., a particle with extremer values has at each pixel location in the correlation image a value larger than the one observed for F_i.
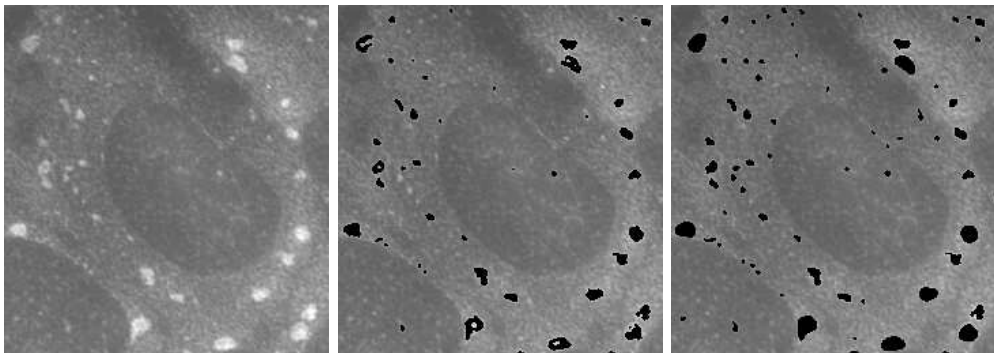

Fig. 3. Detection of stress granules. Left: Detail of image no. 8 with stress granules of varying sizes. Middle: Detection results using the method [6]. Right: Detection results using the scale-adaptive method (Image values scaled for visualization of results).

Still assuming independence of pixels this yields

$$p(F_i) = \prod_{(x,y) \in F_i} P\left(C_{[a_n, b_n]}(x, y) \geq c_{[a_n, b_n]}(x, y) \mid H_0 \right) \qquad (4)$$

We use this concept to compare a set of overlapping particle hypotheses to delete inferior nodes from the trees. We keep the particles with smallest p-value which consequently are assumed to be the ones most unlikely caused by chance. As p-values of particles with different size of support are compared, these raw p-values are normalized according to their support to allow fair comparison. In the case of multiple children their p-values are multiplied for comparison with the parent and we decide for F_i at the coarser scale if

$$p\left(F_i\right)^{\frac{1}{|F_i|}} < \prod_{\{k | F_k \text{ child of } F_i\}} p\left(F_k\right)^{\frac{1}{|F_k|}} \qquad (5)$$

and for particles F_k on the finer scale otherwise.

3 Results

The proposed approach is tested on 10 microscope images of U2OS osteosarcoma cells stressed with sodium arsenate for one hour before fixation. SGs were labeled by immunostaining of ZBP1 in red, and for 5 images PBs were labeled by immunostaining of DCP1A in green.

For our application, two overlapping intervals of scale $[a_1, b_1] = [2, 3]$ and $[a_2, b_2] = [3, 4]$ have shown to be sufficient and will be used for the experiments reported in the following. Fig. 3 shows a detail of image no. 8 with fluorescently labeled SGs, and segmentation results for the method [6] with scales $[a, b] = [2, 4]$ and the proposed scale-adaptive method, respectively. For the scale-adaptive method the distribution of the size of granules detected in each of the scale intervals is depicted in Fig. 2. In analogy to Fig. 3, Fig. 4 gives segmentation results for PBs in a part of image no. 2.
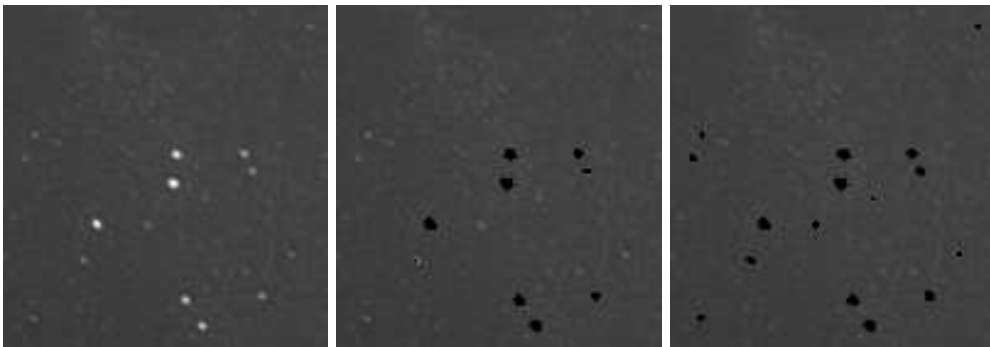

Fig. 4. Detection of processing bodies. Left: Detail of image no. 2 with processing bodies. Middle: Detection results using the method [6]. Right: Detection results using the scale-adaptive method. (Image values scaled for visualization of results).

4 Discussion

Stress granules show a large variety of different sizes and shapes. Using only one interval of wavelet coefficients as in [6] imposes the implicit constraint on similar shape and size for all granules. As can be seen from Fig. 3 (middle) this allows to detect a set of pronounced granules, however, misses several smaller granules and sometimes leads to incomplete segmentation for large granules, i.e. their shapes show deep convexities unusual for granules. In contrast, applying our new scale-adaptive approach based on selection of detection results from different scale intervals overcomes this problem. The detection is improved as detected granules cover a larger range of different scales and have more accurate contours (Fig. 3, right). The local adaptivity of our method shows also in Fig. 2. Detections from interval $[a_2, b_2]$ correspond mainly to large-sized granules, while small granules are detected predominantly in interval $[a_1, b_1]$.

Detection results for PBs also demonstrate the ability of our approach to automatically select features from the best scale. Compared to the results of method [6] our detection also includes PBs of less saliency (Fig. 4) avoiding the canceling effect of coarse scales. The variance in size among PBs is smaller than among SGs. Accordingly, a single fine-scale interval should be sufficient to detect the majority of PBs. Indeed, although intervals $[a_1, b_1]$ and $[a_2, b_2]$ are used for PBs detection, 99 % of the PBs are selected from interval $[a_1, b_1]$.

References

1. Kedersha N, Stoecklin G, Ayodele M, et al. Stress granules and processing bodies are dynamically linked sites of mRNP remodeling. Cell Biol. 2005;169(6):871–4.
2. Xavier J, Schnell A, Wuertz S, et al. Objective threshold selection procedure (OTS) for segmentation of scanning laser confocal microscope images. J Microbiol Methods. 2001;47(2):169.
3. Bolte S, Cordelieres FP. A guided tour into subcellular colocalization analysis in light microscopy. J Microsc. 2006;224(3):213–32.
4. Smal I, Niessen W, Meijering E. A new detection scheme for multiple object tracking in fluorescence microscopy by joint probabilistic data association filtering. In: Proc IEEE ISBI; 2008. p. 264–7.
5. Chan TF, Vese LA. Active contours without edges. IEEE Trans Image Process. 2001;10(2):266–77.
6. Olivo-Marin JC. Extraction of spots in biological images using multiscale products. Pattern Recognit. 2002;35(9):1989–96.
7. Dufour A, Meas-Yedid V, Grassart A, et al. Automated quantification of cell endocytosis using active contours and wavelets. In: Proc ICPR; 2008.
8. Figueiredo MAT, Nowak RD. Wavelet-based image estimation: an empirical Bayes approach using Jeffrey's noninformative prior. IEEE Trans Image Process. 2001;10(9):1322–31.
9. Desolneux A, Moisan L, Morel JM. A grouping principle and four applications. IEEE Trans Pattern Anal Mach Intell. 2003;25(4):508–13.
10. Desolneux A, Moisan L, Morel JM. Edge detection by helmholtz principle. J Math Imaging Vis. 2001;14(3):271–84.

Stress granules are dispensable for mRNA stabilization during cellular stress

Nadine Bley[1,2], Marcell Lederer[1], Birgit Pfalz[3], Claudia Reinke[1,2], Tommy Fuchs[1], Markus Glaß[1,2], Birgit Möller[4] and Stefan Hüttelmaier[1,2,*]

[1]Division of Molecular Cell Biology, Institute of Molecular Medicine, Martin Luther University Halle-Wittenberg, Heinrich-Damerow-Strasse 1, 06120 Halle, Germany, [2]Core Facility Imaging (CFI) of the Medical Faculty, Institute of Molecular Medicine, Martin Luther University Halle-Wittenberg, Heinrich-Damerow-Strasse 1, 06120 Halle, Germany, [3]Genome Biology Unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany and [4]Institute of Computer Science, Martin Luther University Halle-Wittenberg, Von-Seckendorff-Platz 1, 06099 Halle, Germany

ABSTRACT

During cellular stress, protein synthesis is severely reduced and bulk mRNA is recruited to stress granules (SGs). Previously, we showed that the SG-recruited IGF2 mRNA-binding protein 1 (IGF2BP1) interferes with target mRNA degradation during cellular stress. Whether this requires the formation of SGs remained elusive. Here, we demonstrate that the sustained inhibition of visible SGs requires the concomitant knockdown of TIA1, TIAR and G3BP1. FRAP and photo-conversion studies, however, indicate that these proteins only transiently associate with SGs. This suggests that instead of forming a rigid scaffold for mRNP recruitment, TIA proteins and G3BP1 promote SG-formation by constantly replenishing mRNPs. In contrast, RNA-binding proteins like IGF2BP1 or HUR, which are dispensable for SG-assembly, are stably associated with SGs and the IGF2BP1/HUR-G3BP1 association is increased during stress. The depletion of IGF2BP1 enhances the degradation of target mRNAs irrespective of inhibiting SG-formation, whereas the turnover of bulk mRNA remains unaffected when SG-formation is impaired. Together these findings indicate that the stabilization of mRNAs during cellular stress is facilitated by the formation of stable mRNPs, which are recruited to SGs by TIA proteins and/or G3BP1. Importantly, however, the aggregation of mRNPs to visible SGs is dispensable for preventing mRNA degradation.

INTRODUCTION

In response to environmental stress or infection, bulk protein synthesis is severely reduced due to the impairment of the initiation step of mRNA translation. This is mainly facilitated by the stress-dependent activation of kinases phosphorylating the translation initiation factor eIF2α (EIF2S1). The phosphorylation of eIF2α impairs the assembly of the ternary eIF2/tRNA$_i^{Met}$/GTP complex, which is essential for the initiation of mRNA translation and thus results in the stalling of bulk mRNA in 48S 'pre-initiation' complexes (1–3). These associate with various RNA-binding proteins (RBPs) in cytoplasmic mRNPs, which transiently assemble into cytoplasmic stress granules (SGs). This stress-induced assembly or aggregation of mRNPs was suggested to essentially rely on the aggregation of RBPs comprising prion-like or low complexity (LC) regions, which are frequently observed in intrinsically disordered (ID) proteins [reviewed in (4,5)]. In accord, suggested ID-like RBPs including TIA1, TIAR (TIAL1) or G3BPs were demonstrated to induce the formation of SG-like cytoplasmic granules at high cytoplasmic concentration, for instance, when transiently overexpressed in tissue cultured cells (6,7). *In vitro*, some of the ID-like proteins found in SGs were shown to induce the formation of hydrogels, which resemble RNA germ cell granules in *Caenorhabditis elegans* (8). The latter behave like liquid droplets formed or maintained by the transient interaction of LC-containing RBPs, which presumably associate with mRNPs containing translationally stalled mRNAs. Like in germ cells, the dynamic association of LC-containing RBPs and mRNPs in stressed cells was proposed to induce a 'demixing phase transition' resulting in the formation of hydrogel-like cytoplasmic SGs [reviewed in (4,5)]. Recent studies propose that the increase of non-polysome associated mRNA during cel-

lular stress is essential and sufficient to trigger the formation of SGs (9).

Although various SG-recruited proteins have been identified to date, the physiological role of SGs remains largely elusive. Initially, it was proposed that SGs are essential to facilitate the block of bulk mRNA translation during cellular stress (1). However, even when the formation of SGs visible by fluorescence microscopy was impaired by the depletion of factors essential for their assembly, the stress-induced inhibition of bulk protein synthesis appeared largely unaffected (10,11). Moreover, recent studies have indicated that SGs disassemble before mRNA translation is fully restored in cells recovering from cold shock (12). These findings strongly suggest that the block of mRNA translation observed during cellular stress can be facilitated in a largely SG-independent manner. Despite a potential involvement of SGs in reorganizing the 'stress-translatome', it was proposed that the formation of SGs promotes cell viability (13). However, it has to be noted that the vast majority of studies addressing the biological role of SGs rely on the depletion of factors proposed to be essential for the formation of these foci. These factors are likely to serve SG-independent roles in enhancing cell viability and thus their depletion may impair cellular fitness in a largely SG-independent manner.

Various recent findings suggest that SGs serve an essential role in modulating intracellular signaling during cellular stress. This view is supported by the SG-recruitment of various key signaling factors and the observation that perturbing specific signaling cascades impairs the formation of SGs, respectively. For instance, it was demonstrated that RSK2 (p90 ribosomal S6 kinase) depletion interferes with the formation of SGs and reduces cell survival in response to cellular stress (14). Notably, RSK2 is recruited to SGs by associating with the putative prion-like C-terminus of TIA1, a proposed key factor in the assembly of SGs. The recruitment of RSK2 to SGs was suggested to prevent or 'slow down' the induction of apoptosis during cellular stress. Similar mechanisms were also proposed for the stress-induced recruitment of RACK1 or ROCK1 to SGs. The SG-facilitated sequestration of these factors was associated with reduced activation of JNK or JIP-1 dependent apoptosis, respectively [reviewed in (4)]. Notably in this respect, the depletion of O-GlcNac transferase, which modifies various cellular proteins including key signaling factors like RACK1, was shown to impair the formation of SGs (10). Consistent with a role of SGs in modulating the ROCK1-facilitated control of apoptosis, it was proposed that the Wnt-dependent activation of Rac1 interferes with the RhoA-dependent activation of SG-assembly [reviewed in (4)]. Most recently, it was revealed that the inhibition of mTORC1 by the sequestering of RAPTOR in SGs prevents or severely limits the apoptosis of cancer cells during cellular stress (15). In summary, it appears tempting to speculate that the assembly of SGs interconnects the stress-induced block of bulk mRNA translation and key signaling cascades modulating cell viability and apoptosis [reviewed in (4)].

In previous studies, we demonstrated that the IGF2 mRNA binding protein 1 (IGF2BP1), also termed ZBP1 (Zipcode binding protein 1), is recruited to SGs together with its target transcripts (16,17). This transient enrichment in SGs was correlated with the selective stabilization of IGF2BP1-target mRNAs including MYC, ACTB and MAPK4 (16,18). These findings prompted us to speculate that the selective stabilization of target mRNAs during cellular stress allows the identification of novel RBP-target transcripts. Accordingly, we analyzed if the transient depletion of IGF2BP1 allows the identification of target mRNAs in stressed tumor-derived cells. As expected, we identified almost a hundred novel target candidate transcripts, which were selectively decreased in stressed cells upon the knockdown of IGF2BP1 (18). Notably, these target mRNAs are also subjected to IGF2BP1-dependent regulation of mRNA translation and/or turnover in non-stressed cells (18,19). Hence, the selective stabilization of mRNAs by RBPs during cellular stress allows the identification of target mRNAs regulated by the same proteins, presumably via cytoplasmic mRNPs (20), also in non-stressed cells. However, it remained elusive if the selective stabilization of mRNAs observed during cellular stress is facilitated via SGs.

In this study, we addressed the role of SGs in the stress-induced stabilization of mRNAs. To this end we aimed at establishing protocols allowing the impairment of SG-formation without perturbing stress signaling. In remarkable contrast to various previous reports, our studies reveal that the sustained inhibition of SGs visible by fluorescence microscopy requires the concomitant depletion of TIA proteins and G3BP1. FRAP and photo-conversion studies, however, indicate that these proteins rapidly shuttle between SGs and the cytoplasm with barely any immobile fraction in SGs. The opposite is observed for RBPs dispensable for SG-formation including IGF2BP1, YB1 (YBX1) or HUR (ELAVL1). This provides further evidence that the recruitment of mRNPs into SGs by TIA proteins and G3BPs relies on their transient association/aggregation instead of forming a rigid and largely non-dynamic scaffold for mRNP-association. Finally, our studies reveal that the stabilization of bulk mRNA including IGF2BP1 target transcripts is largely independent of SG-formation. Taken together these findings provided strong evidence that the aggregation of mRNPs in SGs neither required for the control of bulk mRNA translation nor the stabilization of translationally stalled transcripts.

MATERIALS AND METHODS

Cell culture, transfections and treatments

U2OS, HUH7 and HEK293 cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum (FBS). Where stated, plasmids (48 h) or siRNAs (72 h) were transfected, according to manufacturer's instructions, using Lipofectamine 2000 or RNAiMax (Life Technologies), respectively. Cells were transduced by lentiviral vectors, as previously described (18,19). All siRNAs and plasmids are depicted in Supplementary Table S1. All transfected cells were splitted 24 h post-transfection to allow application of different assays from the same transfection and/or the analysis or several (stress) conditions. Oxidative or endoplasmic reticulum (ER)-stress was induced using sodium-arsenate (2.5 mM) or thapsigargin (1 μM) for indicated time. Actinomycin D (ActD) was added (5 μM) to block transcription for the time indicated.

Western blotting

Western blotting was essentially performed as recently described using RIPA-buffer (20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% NP-40, 1% sodium deoxycholate (DOC), 2.5 mM sodium pyrophosphate, 1 mM beta-glycerophosphate, 1 mM Na_3VO_4; supplemented with protease inhibitor cocktail (Sigma Aldrich)) to lyse cells (18,19). For used primary antibodies please refer to Supplementary Table S2. All secondary antibodies used were described before (18).

Co-immunopurification

Cells were lysed in gradient buffer (10 mM Hepes pH7.4, 150 mM KCl, 5 mM $MgCl_2$, 0.5% NP40; supplemented with protease and phosphatase inhibitors). Proteins were immune-purified by indicated antibodies immobilized to Dynabeads® Protein G (Life Technologies). After intense washing, proteins were eluted using sodium dodecyl sulphate sample buffer and analyzed by western blotting, essentially as described previously (17).

Metabolic labeling by ^{35}S-methionine

U2OS cells transfected with indicated siRNAs were pre-incubated with methionine-free medium overnight 48 h post-transfection. Protein *de novo* synthesis was monitored by the addition of 100 μCi ^{35}S methionine per 6-well for 1 h. Where indicated cells were simultaneously stressed by arsenate. Upon extensive washing cells were extracted in RIPA buffer. Equal amounts of total protein were analyzed by western blotting using Ponceau-staining to analyze protein loading and phosphor-imaging to determine newly synthesized proteins.

Sucrose gradient centrifugation

U2OS cells (~1.5 Mio cells) transfected with indicated siR-NAs and stressed by arsenate when stated were lysed in gradient buffer 72 h post-transfection. Total protein (D_C protein assay, Bio-Rad) and RNA (OD_{260}) concentrations were determined to ensure equal loading of the gradients. Linear 15–45% sucrose (w/v) gradients in gradient buffer lacking NP-40 were centrifuged in a Beckman SW-40 rotor at 30 000 revolutions per minute for 2 h. Gradients were fractionated and ultraviolet-profiles were monitored by a Foxy Jr. fraction collector (Teledyne) with syringe pump (Brandel) (18).

Quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR) and microarray analyses

Changes in RNA abundance were determined by qRT-PCR as recently described (18,19). All gene-specific primer pairs used are shown in Supplementary Table S1. For microarray analyses, total RNA was extracted using TRIZOL and further purified using RNeasy MinElute Cleanup Kit (QIAGEN). RNA integrity and concentration was then examined on an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) using the RNA 6.000 LabChip Kit (Agilent Technologies). Array analyses were performed

at the microarray core facility of the IZKF (Leipzig, Germany) essentially as recently described using two independent chip systems (Affimetrix and Illumina) (18). For the analysis with Affimetrix HG133plus 2.0 chips U2OS cells were transfected with control (siC) siRNAs or siRNAs directed against TIA-1, TIAR and G3BP1 (siSGs). Where indicated cells were treated with arsenate and ActD for 2 h (stress). All samples were analyzed in duplicates. Raw, Mas5- or RMA-normalized data of Affimetrix chips were analyzed using Bioconductor (www.bioconductor.org) and R (www.r-project.org). Reliably detected transcripts in both untreated samples identified by Mas5 present/absent calls were further analyzed. Ratios (knockdown versus control) of log2 expression data from non-stressed cells were plotted against the corresponding ratios of stressed cells to determine stress- and knockdown-dependent changes in RNA abundance.

For the analysis using Solexa HumanHT-12 chips (Illumina) U2OS cells were transfected with two independent sets of TIA-1, TIAR and G3BP1-directed (siSGs) or control (siC) siRNAs and treated with arsenate and ActD for indicated time. Reliably detected transcripts in all untreated samples were identified by a *P*-value less than 0.001. A linear regression was applied to the quantile-normalized and background corrected non-logarithmic expression data to determine the mRNA degradation over time (slope) for siSGs versus siC transfected cells. Pearson's correlation analyses were used to determine how the inhibition of SG-formation affects bulk mRNA degradation.

Microscopy and image analyses

Indirect immunostaining was essentially performed as previously described (16). For primary antibodies please refer to Supplementary Table S2. All secondary antibodies were previously described (18). Images were acquired on a Leica TCS-SP5X CLSM equipped with a Ludin live chamber or a Nikon TE-2000E fluorescence microscope using 63× magnification and standardized settings. The area fraction representing the number and size of SGs was automatically quantified using the Mica2D particle detector of MiToBo (www.informatik.uni-halle.de/mitobo), an extension package for ImageJ (www.imagej.nih.gov/ij/). To allow an assessment of SG parameters for individual cells the cell area was manually labeled.

For FRAP and photo-conversion analyses, U2OS cells were transiently transfected for 24 h before seeding on glass bottom dishes (MatTek). Where indicated stably expressing cell clones were generated by G418 and at least two distinct clones were included in the studies. If not indicated otherwise, cells were stressed by arsenate for 25 min. Imaging was conducted up to 1 h after arsenate treatment. FRAP analyses were performed on the TCS-SP5X using the provided FRAP wizard. The region of interest was selected to cover a single SG and fluorescence was bleached using the Argon laser (488 nm) at maximal power. Photo-conversion analyses were performed using the FRAP wizard using standard settings for the concomitant detection of green-fluorescent protein (GFP) (Argon laser: 488 nm) and red-fluorescent protein (RFP) (DPSS laser: 561 nm) fluorescence. Photo-conversion was induced by using the bleach point function

of the LAS AF software package at maximal laser power (Argon laser: 488 nm) for 100 ms. The bleach point was chosen in close proximity to a single SG without affecting additional SGs. For both, FRAP as well as photo-conversion analyses, the recovery or loss of fluorescence signal was recorded at a 300 ms time interval for five frames before and 100 or 300 frames after bleaching or conversion, respectively. The wizard software application was used to normalize the fluorescence intensities for background bleaching or conversion. The traffic model for molecules moving in and out of SGs can be described by first-order kinetics, which are $k_1 M_{IN} = dM_{IN}/dT$ and $k_2 M_{OUT} = dM_{OUT}/dT$. $k_1 M_{IN}$ and $k_2 M_{OUT}$ describe the traffic constants, whereas M_{IN} and M_{OUT} are the number of molecules moving into and out of SGs. At steady-state levels incoming and outgoing molecules are expected to be balanced and thus the corresponding half-lives are considered to be $t_{1IN} = t_{2OUT}$. FRAP parameters determined by first-order kinetics are summarized in Supplementary Figure S8A.

RESULTS

Inhibiting SG-formation without affecting stress signaling

To analyze whether the IGF2BP1-dependent stabilization of target mRNAs during cellular stress is essentially facilitated via SGs, we aimed at inhibiting or severely reducing SG formation without affecting the phosphorylation of eIF2α or the block of bulk mRNA translation (1,3). Initially, we focused our efforts on approaches previously reported to impair the assembly of SGs visible by fluorescence microscopy. These included: (i) the knockdown of FMRP (21); (ii) the transient depletion of ataxin-2 (22); (iii) the knockdown of RSK2 (14); (iv) the forced expression of TIA1-ΔRRM (3,6); (v) the depletion of HDAC6 (23); (vi) the knockdown of TIA-proteins and/or G3BP1, since these proteins were suggested as key factors of SG assembly (6,24). For all these approaches, SG formation was monitored in U2OS and/or Huh7 cells upon the transient depletion or overexpression of the respective protein factors. Initially, stress was induced by 1 h of arsenate treatment and SG formation was monitored by fluorescence microscopy upon immunostaining of key SG-components including IGF2BP1, YB1, TIA proteins and/or G3BP1, respectively. The number of SG-positive cells as well as the SG-area fraction indicating the percentage of cell area covered by SGs was determined manually. In the following we refer to SGs when mRNP-aggregates were detectable by fluorescence microscopy.

The role of fragile X mental retardation protein (FMRP) in SG formation was monitored in tumor-derived U2OS cells upon the siRNA-directed depletion of FMRP (Supplementary Figure S1A and B). Although the endogenous FMRP was recruited to SGs, SG assembly induced by arsenate remained largely unchanged by the knockdown of FMRP (Supplementary Figure S1A and B). To exclude bias due to knockdown efficiencies or the transfection procedure we extended our analyses to immortalized FMRP (-/-) mouse embryonic fibroblasts (MEFs) versus KO-MEFs stably transduced with Flag-tagged FMRP (25). As in U2OS cells, the exogenous FMRP was recruited to SGs, but SG assembly induced either by arsenate (1 h), arsenite (30 min)

or thapsigargin (1 h) (data not shown for arsenite and thapsigargin) appeared largely unaffected by the loss of FMRP (Supplementary Figure S1C and D). Likewise, the number of SG-positive U2OS or Huh7 cells remained essentially unchanged upon the knockdown of ataxin-2 (Supplementary Figure S2A–D). Surprisingly, in Huh7 SG formation appeared even enhanced upon ataxin-2 depletion as evidenced by an increase in the apparent size of SGs. The transient depletion of RSK2 impaired the arsenate-dependent formation of SGs as suggested (data not shown). However, this inhibition was correlated with an impaired arsenate- or thapsigargin-induced phosphorylation of eIF2α (Supplementary Figure S3A). This confirmed previous reports indicating RSK2 to modulate the activation of PKR and thus the stress-dependent activation of eIF2α (26,27). TIA1 is a key marker of SGs, which was shown to facilitate SG-formation via a C-terminal prion-like domain (3,6). However, upon deletion of the N-terminal RRM-domains, the truncated TIA1-ΔRRM protein was proposed to impair the formation of SGs when transiently expressed in tissue-cultured cells. As reported, TIA1-ΔRRM appeared to impair the formation of 'regular' IGF2BP1-containing SGs in ~45% of transiently transfected cells (Supplementary Figure S3B). However, in the majority of cells, 'regular' SGs were formed. Surprisingly, these also contained TIA1-ΔRRM. Most notably, however, we also observed an impairment of SG formation in ~20% of GFP-transfected cells and in cells expressing high levels of IGF2BP1, HUR or YB1 (data not shown). Consistent with recent studies (9), this suggested that the transient expression of proteins, in particular RBPs, can interfere with the formation of SGs (Supplementary Figure S3B). The knockout of the SG-recruited histone deacetylase 6 (HDAC6) was reported to prevent the formation of SGs in stressed MEFs (23). However, in U2OS cells we failed to confirm the recruitment of endogenous as well as GFP-tagged HDAC6 to thapsigargin- (data not shown) or arsenate-induced SGs (Supplementary Figure S4A and B). Moreover, the assembly of SGs was unchanged in cells depleted of HDAC6 suggesting that SG-formation is largely uncoupled from HDAC6 function in cancer-derived cells (Supplementary Figure S4C and D). Finally, we investigated if the formation of SGs can be inhibited by the depletion of TIA proteins and/or G3BP1. In previous studies, it was demonstrated that the formation of SGs is significantly impaired in TIA1 KO-MEFs stressed by arsenite for 30 min (6). Moreover, it was proposed that the G3BP1 protein is an essential facilitator of SG formation (7). However, the assembly of SGs in U2OS cells treated with arsenate for 1 h was only modestly decreased by the depletion of TIA1 or TIAR (Supplementary Figure S5A–D). This supported recent findings suggesting TIA1 only plays a minor role in SG formation (9). In cells transfected with G3BP1-directed siRNAs, the impairment of SG-formation was significantly more pronounced with ~40% of cells lacking visible SGs after 1 h of arsenate treatment. Consistently, the 'area fraction of SGs' was only modestly reduced by the knockdown of TIA1 or TIAR yet significantly decreased by the depletion of G3BP1. However, the depletion of individual factors appeared insufficient for severely inhibiting SG formation, as previously suggested for G3BP1/2 depletion (28). This suggested that

the concomitant depletion of these factors enhances the impairment of SG formation even at late time points of stress application. To evaluate this by quantitative means and in a time-resolved manner, we adapted an automated tracing algorithm previously used for identifying focal contacts (18). This allowed the automated tracing of SGs labeled by indirect immunostaining of YB1 and IGF2BP1 in U2OS cells treated with arsenate for 30–120 min (Figure 1A–C). In U2OS cells transfected with control siRNAs (siC), ∼80% of cells contained SGs already 30 min after arsenate application (Figure 1B). Consistent with previous studies (6), the depletion of TIA1 or TIAR essentially abolished SG formation after 30 min of stress induction but ∼80% of cells contained SGs after 1 h of arsenate treatment. This delay of SG assembly was significantly enhanced upon the depletion of G3BP1 with only 70% of SG-positive cells after 2 h of arsenate treatment. The concomitant knockdown of all three factors in U2OS, termed SG-knockdown/-depletion (siSGs, SGD), severely impaired SG formation and reduced the number of SG-positive cells to ∼20% of control levels after 2 h of arsenate-induced stress (Figure 1A and B). In agreement with a severely reduced assembly of SGs, we observed a striking reduction in the SG-area fraction upon the depletion of all three proteins (Figure 1C). This was confirmed by an additional set of siRNAs, in another cell line (Huh7) and thapsigargin to test an additional stressor (Supplementary Figure S6A–C). As observed in U2OS cells, SG formation was essentially abolished by the concomitant depletion of TIA1, TIAR and G3BP1 irrespective of the used stressor or cell line. To exclude that the impairment of SG formation resulted from aberrant stress signaling, the phosphorylation of eIF2α was monitored upon the SGD (Figure 1D; Supplementary Figure S6B). In both cell lines analyzed, U2OS as well as Huh7, the upregulation of eIF2α phosphorylation by arsenate or thapsigargin appeared unchanged by the SGD suggesting that 'stress signaling' was largely unaffected. Consistently, sucrose gradient centrifugation revealed that the depletion of polysomes and the shift of bulk (m)RNA to pre-polysomal fractions was preserved in arsenate-stressed cells when SG assembly was abolished by the SGD (Figure 1E). UV260 reads of indicated fractions were averaged over cell populations transfected with control or TIA1-/TIAR-/G3BP1-directed siRNAs. Thus, the small variation in UV reads indicated that the SGD neither affected mRNA translation in non-stressed (Figure 1E, gray) nor the block of bulk mRNA translation in arsenate-stressed cells (Figure 1E, black). Finally, this was evaluated by the metabolic labeling of stressed (+, arsenate) or non-stressed (−, arsenate) cells transfected with control (C) or TIA1-/TIAR-/G3BP1-directed (SG) siRNAs (Figure 1F). In agreement with the polysomal profiling studies, protein synthesis was massively reduced in response to arsenate treatment in both, the control as well as SGD populations.

In summary, the analyses indicated that the combined knockdown of TIA1, TIAR and G3BP1 (SGD) substantially impaired the formation of visible SGs without affecting the upregulation of eIF2α phosphorylation or the stress-induced block of bulk mRNA translation.

SG assembly: dynamic recruitment of mRNPs versus rigid prion-like scaffolds

Overall our studies were in agreement with the current view that TIA proteins and G3BP1 are essential for the formation of SGs. However, the depletion of TIA proteins or G3BP1 only delayed the formation of SGs, whereas their combined knockdown severely diminished SG formation in a sustained manner. This suggested that either these proteins serve redundant functions as prion-like scaffolding or glue-like factors in the assembly of SGs or that they act redundantly in dynamically recruiting mRNPs to SGs. In line with the prion-like scaffolding scenario one would expect the three factors have an average SG half-life longer than observed for other SG-localized RBPs for which no obvious role in SG assembly has been reported, e.g. IGF2BP1 (16). Moreover, it is tempting to speculate that a prion-like scaffolding factor shows some sort of an immobile fraction in SGs, whereas this should be less pronounced for the transiently stored 'cargo', the mRNPs comprising translationally stalled mRNAs and RBPs like IGF2BP1. Aiming to evaluate these aspects, we analyzed the SG dynamics of the prion-like TIA1, TIAR and G3BP1 proteins versus three RBPs (IGF2BP1, YB1 and HUR) that are localized to SGs. As also demonstrated for HUR and YB1 (9), all three analyzed RBPs were dispensable for the assembly of SGs, since SG formation remained unaffected upon the depletion of each protein (IGF2BP1, HUR and YB1) alone as well as the concomitant knockdown of all three RBPs (Supplementary Figure S7A–H).

The dynamics of all proteins were first analyzed by FRAP in cells transiently and/or stably expressing the GFP-tagged fusion protein of interest (Figure 2A and B; Supplementary Figure S8A). Although we aimed at analyzing stably expressed proteins where feasible, we did not observe significant kinetic differences between stably versus transiently expressed proteins (Supplementary Figure S8). Surprisingly, however, TIA1, TIAR as well as G3BP1 showed a short half-live of ∼2–3 s and an insignificant immobile fraction in SGs (Figure 2A, red; Supplementary Figure S8). The latter was consistent with previous studies showing an almost complete fluorescence recovery for TIA1, G3BP, TTP and CPEB (24,29). However, previous studies suggested that the immobile fraction of CPEB in SGs was dependent on whether granules were induced by arsenite or CPEB overexpression (29). In contrast, we did not observe any sort of stressor-dependent differences in G3BP1 dynamics (Supplementary Figure S8A). Notably, GFP-G3BP1 signal recovered fast and almost complete in both, early and small (5–15 min after stress-induction) as well as late and large SGs (30–60 min after stress induction; Supplementary Figure S8B and D). However, it has to be noted that the kinetic models fitted to the FRAP data rely on the assumption that the state of equilibrium has been reached. This severely biases studies at early time points of stress induction and thus the phase of SG initiation. Aiming to compare SG dynamics with the movement of proteins in the cytoplasm, the latter was analyzed by FRAP in the SG-free cytoplasm of stressed cells (Figure 2A, black). Strikingly, the dynamics of TIA1 and TIAR were essentially indistinguishable between SGs and SG-free cytoplasm suggesting that both proteins
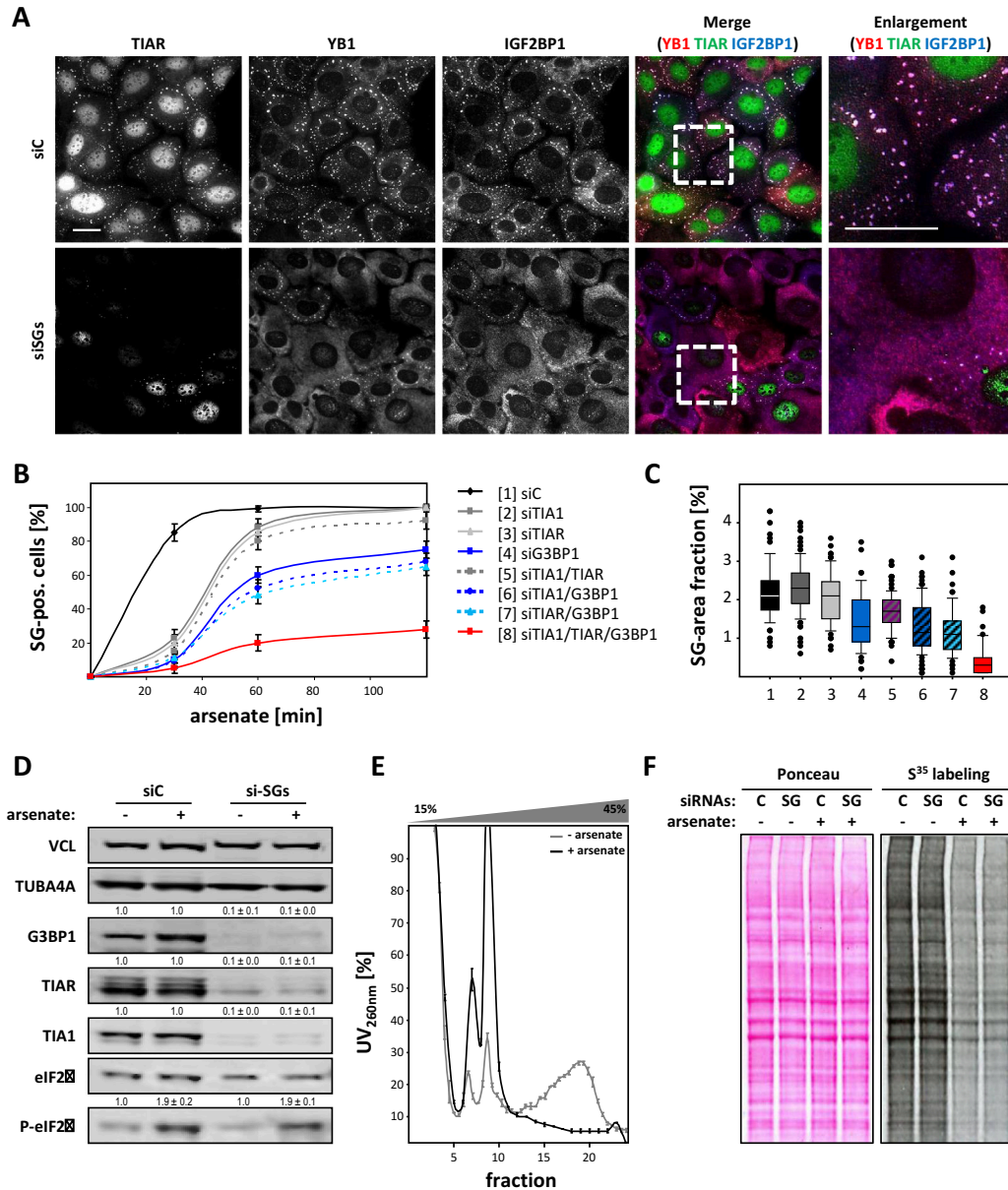
Figure 1. The concomitant depletion of TIA proteins and G3BP1 impairs SG formation. (**A**) U2OS cells transfected with indicated siRNAs (siC, control; siSG, siTIA1, siTIAR and siG3BP1) for 72 h were stressed by arsenate (2.5 μM) for 2 h before immunostaining of indicated proteins. Enlargements of the boxed regions in the merged images are shown in the right panel. Bars, 25 μm. (**B** and **C**) The average number of SG containing cells (B) and the SG area fraction (C) was analyzed by immunostaining for IGF2BP1 and YB1 in U2OS cells transfected with indicated siRNAs. Both parameters were determined by an automated particle detection tool, adapted from (18), after indicated times of arsenate stress. Error bars indicate SD determined by analyzing at least 100 cells per condition in three independent experiments. (**D**) The phosphorylation of eIF2α in non-stressed (−) or arsenate (+, as in A) stressed U2OS cells transfected with control (siC) or siSG (as in A) siRNAs was determined by western blotting with indicated antibodies. VCL and TUB4A4 served as loading controls to determine knockdown efficiencies as indicated by numbers above each panel. Standard deviation was determined from three independent experiments. (**E**) The association of bulk (m)RNA with polysomes was monitored by linear (15–45% w/v) sucrose gradient centrifugation in stressed (+, arsenate) versus non-stressed (−, arsenate) U2OS cells transfected with control (siC) or siSG siRNAs, as in (A). The distribution of RNA was monitored by UV spectroscopy and is shown as the average absorbance determined for individual fractions isolated from siC- and siSG-transfected samples. Error bars indicate SD determined in three independent studies for siC- and siSG-transfected cells. (**F**) Protein synthesis in arsenate-stressed (+) versus non-stressed (−) U2OS cells transfected as in (A) was analyzed by metabolic labeling using S35-methionine. The fraction of newly synthesized proteins was determined by western blotting using autoradiography (right panel). Equal loading was controlled by Ponceau staining (left panel).
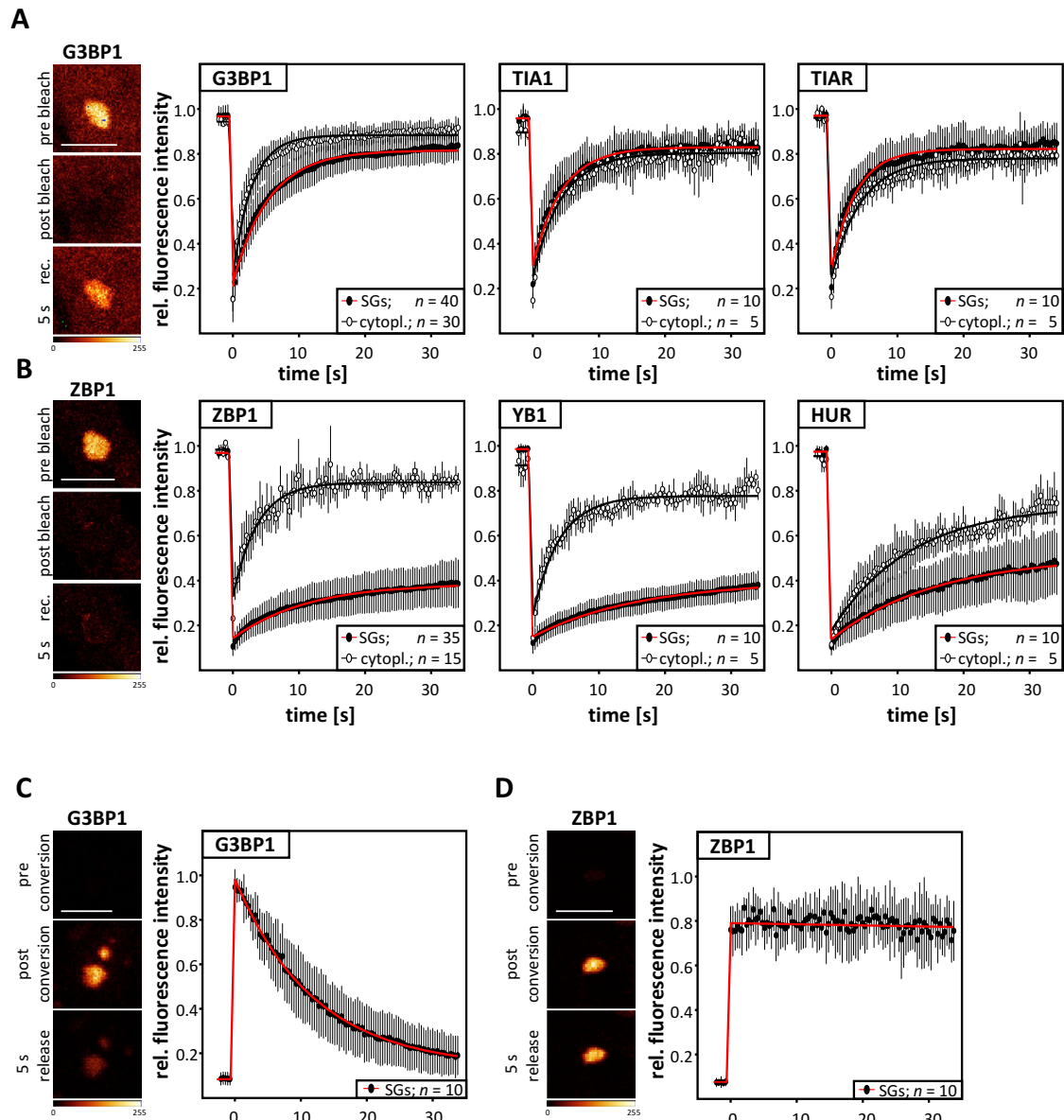
Figure 2. SG dynamics distinguish two classes of SG-associated RBPs. (**A** and **B**) The averaged recovery of GFP-fluorescence for the indicated proteins in SGs (red) or SG-free cytoplasm (black) in U2OS cells stably expressing and/or transiently transfected with the indicated proteins was determined by FRAP. Representative images of fluorescence intensities observed in SGs for GFP-G3BP1 (A) or GFP-ZBP1 (B) at indicated times of FRAP studies are shown in pseudo-colors (glow over/under) in the left panels. (**C** and **D**) The change in SG-localized photo-converted Dendra-fused G3BP1 (C) or ZBP1 (D) was determined by time-lapse microscopy. The averaged change of fluorescence intensities was determined over 30 s after photo-conversion. Representative images of fluorescence intensities observed at indicated time after photo-conversion are shown in left panels. Error bars indicate SD determined for the number of analyses summarized together with kinetic data in Supplementary Figure S8A. Bars, 5 μm.

are rapidly exchanged and remain highly dynamic irrespective of their subcellular localization. Although slightly less mobile in SGs, this was also observed for G3BP1 (Figure 2A; Supplementary Figure S8).

In sharp contrast to the rapid exchange of TIA proteins and G3BP1, the three RBPs dispensable for SG formation (IGF2BP1, YB1 and HUR) showed significantly longer half-lives (~15–20 s) and a striking immobile fraction (~60–70%) in SGs (Figure 2B, red, Supplementary Figure S8A). Although less mobile in the cytoplasm when compared to TIA proteins or G3BP1, the analyzed RBPs were substantially more dynamic in the SG-free cytoplasm than in SGs (Figure 2B, black; Supplementary Figure S8A). Notably, the substantially distinct dynamics of TIA proteins and G3BP1 versus other RBPs like IGF2BP1 were observed in early as well as late SGs (Supplementary Figure S8B–D). As for G3BP1, the recovery time of RBPs like IGF2BP1/ZBP1 in early SGs was shorter. This suggests that the accumulation of mRNPs in SGs could be pronounced at early time point of stress induction. This conclusion is of course limited due the fact that an equilibrium has not been reached and thus the conclusiveness of FRAP studies remains limited, as out lined above. Taken together, the presented findings indicated that the prion-like scaffolding TIA proteins and G3BP1 were rapidly turned over in SGs, whereas the RNA- and thus 'cargo'-associated RBPs were substantially less dynamic. This was largely consistent with previous reports indicating complete fluorescence recovery for G3BP1 or TIA1 but a substantial immobile fraction for the PABP as well as slow and incomplete exchange of CPEB between arsenite-induced SGs and cytoplasm (24,29).

To access SG association more directly, the 'residence-time' of one representative of the two classes of SG-localized proteins, G3BP1 versus IGF2BP1, were transiently expressed as Dendra-fusion proteins in U2OS cells (Figure 2C and D). Dendra is a GFP, which is photoconvertible to a RFP by a UV-irradiation or short wavelength lasers (e.g. 405 or 488 nm). This allowed analyzing how rapidly the SG-localized proteins were exported to the cytoplasm when the Dendra-fusion protein had been photo-converted in a single SG (green to red). Consistent with a rapid exchange determined by FRAP, photoconverted Dendra-G3BP1 was cleared from SGs within seconds. In contrast, barely any SG-localized Dendra-IGF2BP1 was lost 30 s after photo-conversion.

In summary, the FRAP and photo-conversion analyses indicated at least two classes of SG-recruited RBPs: (i) proteins essential for the assembly of SGs including G3BP1 and TIA proteins which were rapidly exchanged; (ii) RBPs like IGF2BP1, YB1 or HUR which were dispensable for SG assembly but barely exchanged with the SG-free cytoplasm.

G3BP1 promotes SG assembly by the RNA-dependent association with RBPs

The dynamics of TIA proteins or G3BP1 suggested that these factors could facilitate the dynamic aggregation of mRNPs in SGs by delivering protein-RNA complexes to these *foci*. Accordingly, the association of these 'mRNP-movers' with mRNAs and/or other RBPs like IGF2BP1

should be pronounced during cell stress. Since IGF2BP1 was previously reported to associate with G3BP1 in differentiated P19 neuronal cells (30), we characterized the association of IGF2BP1 and G3BP1 during cellular stress. HEK293 cells were used for these studies due to the high abundance of IGF2BP1 (20).

Consistent with previous studies, IGF2BP1 copurified with G3BP1 in a RNA-dependent manner from both, stressed as well as non-stressed HEK293 cells (Figure 3A). In stressed cells, the amount of IGF2BP1 copurified with G3BP1 in a RNA-dependent manner was modestly but reproducibly increased, the amount of IGF2BP1 copurified with G3BP1 in a RNA-dependent manner was modestly but reproducibly increased. This was also observed for the association of G3BP1 with HUR. In agreement, G3BP1 also associated with stably expressed Flag-tagged ZBP1, the chicken ortholog of human IGF2BP1 (Figure 3B), which was previously shown to recover or mimic IGF2BP1-dependent phenotypes (17). Binding was abolished when all four KH-domains were inactivated by point mutations (ZBP1-KH1-4; (17)) in the GXXG loop indicating that IGF2BP1/ZBP1 and G3BP1 associate indirectly via (m)RNA. Notably, we previously demonstrated that the ZBP1-KH1-4 mutant does not localize to SGs suggesting that the SG recruitment of IGF2BP1/ZBP1 essentially relies on RNA binding (17).

The role of G3BP1 in modulating SG formation was suggested to rely on the dephosphorylation of the protein at S149 as well as RNA binding (7). Accordingly, we investigated whether the association of IGF2BP1 and G3BP1 is modulated via the acidic domain comprising S149, involves the RRM and/or the C-terminal RGG domain (Figure 3C). To this end, GFP-tagged wild-type and mutant G3BP1 proteins were probed for association with IGF2BP1 by co-immunoprecipitation (Figure 3D). G3BP1 and the non-phosphorylatable G3BP1-S149A copurified with IGF2BP1 from transiently transfected HEK293 cells. This was also observed for the S149E mutant protein (data not shown). The copurification of the RRM mutant protein (FFVV: F380,382V) or the RGG-lacking mutant (ΔRGG) was either severely diminished or abolished, respectively. These findings supported the view that the association of G3BP1 with IGF2BP1 is facilitated indirectly via (m)RNA but does not involve the modification of S149 in G3BP1.

G3BPs are considered to act as prion-like nucleators of SGs by enhancing the formation of these foci via their oligomerization (7,28,31). The latter is controlled by the modification of S149, as previously proposed (7). In contrast to the view that the prion-like aggregation of G3BPs and/or TIA proteins provides a rigid and largely non-dynamic scaffold for SG assembly, our analyses of G3BP1 protein dynamics and stress-dependent protein association suggested that the protein promotes SG formation by the dynamic and RNA-dependent recruitment of protein-RNA complexes. To investigate this in further detail, the G3BP1 wild-type and mutant proteins were transiently and stably expressed in U2OS cells. The formation of SGs and recruitment of proteins was monitored by indirect immunostaining of TIA1 and IGF2BP1 (Figure 4A; only shown for stably expressing cells). Consistent with previous reports, transiently expressed G3BP1 induced IGF2BP1- and TIA1-
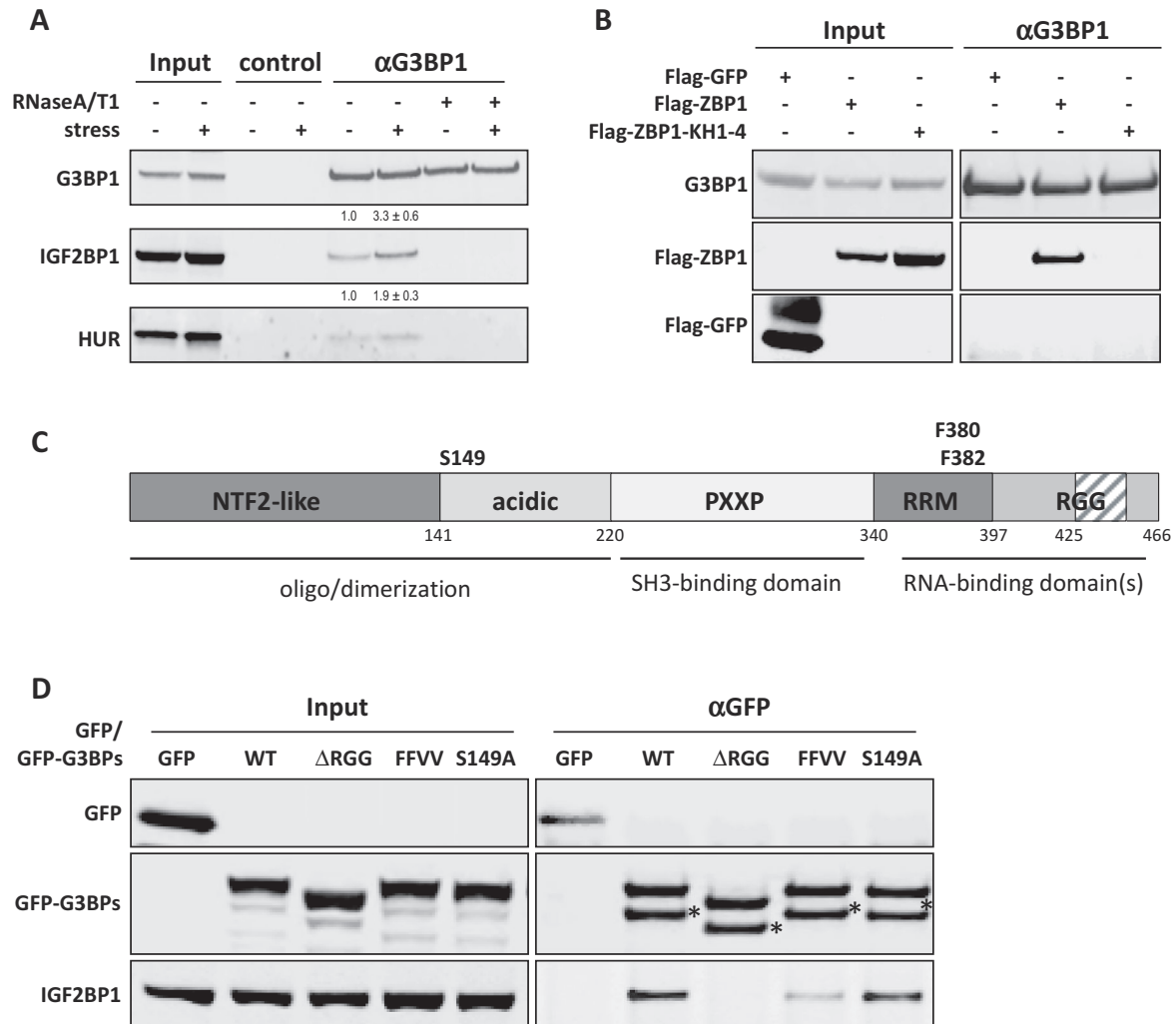
Figure 3. The association of G3BP1 with IGF2BP1 and HUR is enhanced during cellular stress. (**A**) The co-immunopurification of IGF2BP1 or HUR with G3BP1 from arsenate-stressed (+) or non-stressed (−) HEK293 cells was analyzed by western blotting. Where indicated, lysates were supplemented with RNaseA/T1. ProteinG dynabeads served as negative control. Copurification of IGF2BP1 or HUR with G3BP1 was quantified relative to immunopurified G3BP1 amounts by quantitative western blotting. The ratio of IGF2BP1/HUR copurified with G3BP1 was set to one, as indicated above by numbers above lanes. Standard deviation of copurification was determined from three independent experiments. (**B**) The co-immunopurification of endogenous G3BP1 with indicated proteins transiently expressed in HEK293 cells was analyzed by western blotting. Note that RNA binding of ZBP1-KH1-4 is substantially impaired by point mutation in all four KH domains (17). Flag-GFP served as negative control. (**C**) Schematic of G3BP1 domains, putative functions of the indicated domains and relative position of domains indicated by the numbering of residues. (**D**) Co-immunopurification of endogenous IGF2BP1 with indicated stably expressed GFP or GFP-fused G3BP mutant proteins. G3BP proteins analyzed: *WT*, wild type G3BP1; *ΔRGG*, G3BP1 lacking amino acids 425–466; *FFVV*, full-length G3BP1 with F-V conversion at residues 380 and 382; *S149A*, full-length G3BP1 with S-A conversion at residue 149. Western blotting for indicated proteins is shown for the input or co-immunopurified (α-GFP) protein fraction. HEK293 cells stably expressing GFP served as negative controls. * indicates degradation product.

positive SGs in ∼40% of U2OS cells (Supplementary Figure S9A). Although the induction of SGs appeared less pronounced for G3BP1-S149E, both transiently expressed S149 mutant proteins induced the assembly of SGs. Strikingly, however, all G3BP1 mutants with changes in the RNA-binding domains failed to induce SGs when transiently expressed. Although all stably expressed G3BP1 proteins were recruited to SGs (Figure 4A), they failed to in-

duce SG formation in the absence of stress (Supplementary Figure S9A). In agreement, the stress-induced upregulation of eIF2α phosphorylation remained largely unaffected by the stable expression of all proteins (Supplementary Figure S9B). However, compared to GFP-expressing controls, the number of SG-positive cells and more prominently the SG-area fraction were increased in cells stably expressing wild-type G3BP1 (Figure 4A–C). In contrast,
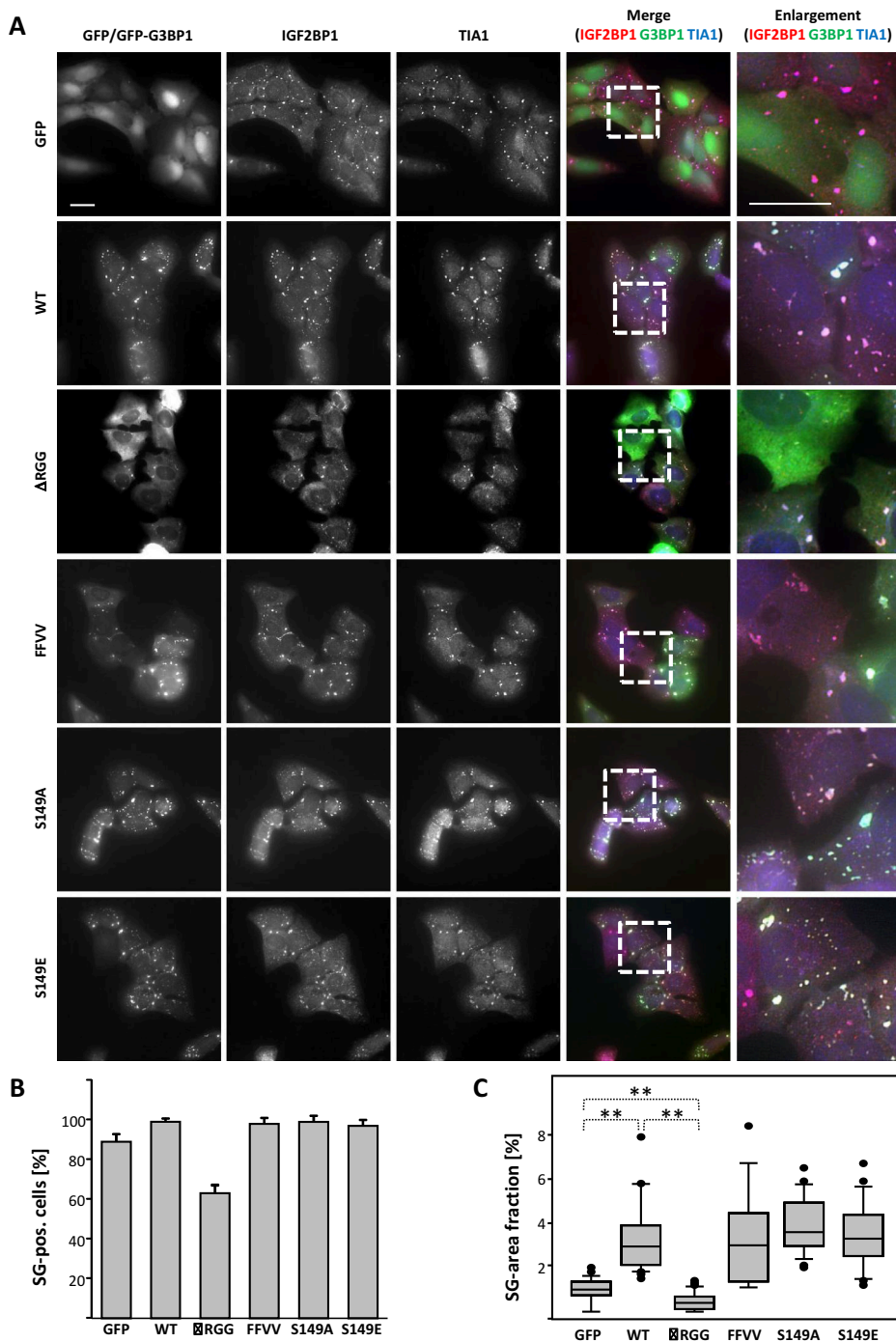
Figure 4. The forced expression of G3BP1 protein mutants modulates SG formation. (**A**) The formation of SGs was monitored in arsenate stressed (1 h) U2OS cells stably expressing GFP or indicated GFP-fused G3BP1 mutant proteins using immunostaining of SG-localized IGF2BP1 and TIA1. Enlargements of boxed regions depicted in the merged images are shown in the right panel. Bars, 25 μm. (**B** and **C**) The number of SG-positive cells (**B**) as well as the SG-area fraction (**C**) was determined as described in Figure 1B and C. Error bars indicate SD of at least three independent analyses including at least 30 cells per condition. Statistical significance was determined by Student's *t*-test: **P < 0.005.

both these parameters were significantly decreased in cells expressing the ΔRGG mutant, which also failed to associate with IGF2BP1. The number of SG-positive cells as well as the SG-area fraction was enhanced in cells stably expressing the other G3BP1 mutant proteins.

In conclusion, these findings supported the view that G3BP1 promotes SG assembly in a dose and RNA-binding-dependent manner. In contrast to previous reports (7), the proposed phosphorylation-dependent control of G3BP1-aggregation appeared irrelevant for the formation of SGs in the cells analyzed here. This provided further evidence that instead of acting as a rigid prion- or glue-like scaffold, G3BP1 promotes SG formation by the dynamic RNA-dependent recruitment of mRNPs.

SGs are dispensable for mRNA stabilization during cellular stress

The concomitant knockdown of TIA proteins and G3BP1 allowed the sustained inhibition of SG formation without affecting stress signaling. This provided a *bona fide* protocol for testing the role of SGs in preventing bulk mRNA degradation during cell stress.

At first, we monitored the turnover of six mRNAs in arsenate-stressed U2OS cells concomitantly transfected with G3BP1-, TIA1- and TIAR-directed siRNAs (SGD; siSGs). The decay of mRNAs was determined upon the block of transcription by actinomycin D (ActD) using qRT-PCR (Figure 5A). Compared to control transfected cells (siC, black), the turnover of three none IGF2BP1-associated transcripts (RPLP0, PPIA and VCL) as well as three IGF2BP1-target mRNAs (ACTB, MAPK4 and MYC) was essentially unchanged by the SGD (siSG, green). Although instable mRNAs like MYC are stabilized during stress, they are still partially degraded. The stress-dependent stabilization of such mRNAs presumably occurs irrespective of SGs, since mRNA decay during cellular stress appeared unchanged when the formation of SGs was impaired by the SGD (Figure 5A). To monitor how the SGD affects the turnover of transcripts induced during the stress response, we analyzed the decay of three mRNAs encoding heat-shock proteins (HSPs), HSP90 (HSP90AA) and HSP70 transcripts (HSPA1A/B; HSPA2), during arsenate-induced cellular stress. For these analyses, cells were stressed by arsenate for 1 h to induce or boost the synthesis of HSP-encoding transcript before blocking transcription by ActD. Like for none stress-induced mRNAs, transcript turnover was monitored by qRT-PCR (Figure 5B). The synthesis of the analyzed HSP-encoding transcripts was enhanced by arsenate confirming the induction of the cellular stress response (Figure 5B; gray). Notably, the induction of mRNA synthesis appeared largely unaffected by the SGD (Figure 5B; compare black to green). Strikingly, HSP transcript levels remained stable over 4 h of ActD treatment in both cell populations, controls (siC, black) versus SGD cells (siSGs, green). However, we cannot exclude that elongated stress conditions might reveal significant differences for HSP-encoding transcripts due to either the inhibition of SG formation or depletion of TIA/G3BP proteins, respectively. Together our data suggested that the stabilization of specific mRNAs could be uncoupled from the for-

mation of SGs during cellular stress. Aiming to monitor bulk mRNA turnover in stressed cells, steady-state mRNA levels in non-stressed and arsenate-treated control versus SGD cells were monitored by comparative microarray analyses (Figure 5C–E; Affymetrix HG133plus2.0). The SGD-induced change in mRNA abundance before stress induction (no stress) or in cells treated with arsenate and ActD for 2 h (stress) was monitored by normalization to siC-transfected controls. These analyses revealed that the SGD failed to induce a significant shift in bulk mRNA levels in stressed or non-stressed cells (Supplementary Figure S10A). This was analyzed in further detail by determining the log2-fold change in transcript abundance induced by the SGD in stressed versus non-stressed cells (Figure 5D and E). Irrespective of the normalization method used, the SGD led to an at least 2-fold up- or down-regulation of 1168 (6.7%) out of 17 398 reliably detected hits (Figure 5D, green). However, the significantly deregulated hits were essentially equally distributed over all four quadrants depicted in the complex dot plot (Figure 5D and E). Moreover, only 236 (1.2%) of all hits were more than 2-fold down-regulated in stressed cells due the SGD (Figure 5D, red). However, the same number of hits was upregulated or reduced by the SGD in the absence of stress (Figure 5D, blue). These results remained essentially unchanged by the normalization method or evaluation platform used for analyzing the microarray data suggesting that steady-state bulk mRNA abundance was largely unaffected by the SGD in stressed and non-stressed cells (Figure 5E). To test how the SGD affects the turnover of bulk mRNA, transcript abundance was monitored in cells treated with arsenate and ActD for 30–120 min. As before, mRNA abundance was monitored by microarrays, this time using another type of chip (Illumina; Solexa HumanHT-12). The decay rate of individual transcripts was calculated by the decline of signal intensities determined by microarray studies (Supplementary Figure S10B). For each of the reliably detectable hits (14 338), the 'slope' of mRNA degradation in control cells (siC-transfected) was plotted versus the rate of degradation determined in SGD cells. Consistent with barely affected steady-state levels, the decay of bulk mRNA appeared largely unaffected by preventing the formation of SGs. This was also validated by Pearson's correlation analyses of the determined decay rates (Supplementary Figure S10B).

In conclusion, the presented findings provided strong evidence that bulk mRNA turnover remained largely unaffected by preventing the formation of visible SGs by the concomitant knockdown of TIA proteins and G3BP1.

IGF2BP1-directed mRNA stabilization during cellular stress is independent of SGs

After having shown that the formation of visible SGs is dispensable for bulk mRNA stabilization during cellular stress, it remained to be addressed if the selective stabilization of IGF2BP1-target mRNAs is maintained when SG formation is impaired. Therefore, we monitored the fate of three IGF2BP1-target transcripts (MYC, ACTB and MAPK4) as well as one none IGF2BP1-associated mRNA (PPIA) upon preventing SG formation in arsenate-stressed cells depleted
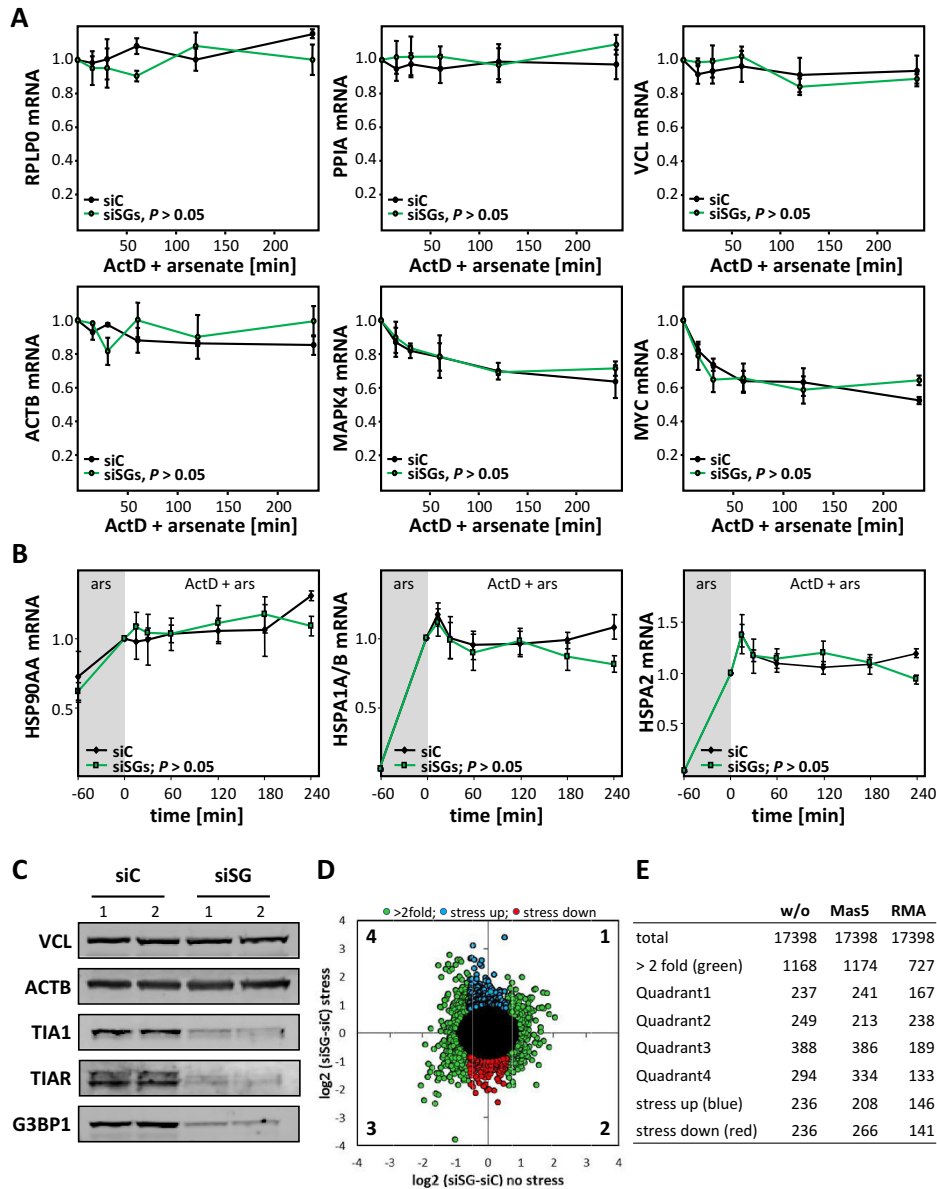
Figure 5. The control of mRNA turnover is independent of SG formation. (**A**) The turnover of indicated mRNAs was analyzed in U2OS cells transfected with indicated siRNAs (siC, control; siSG, siTIA1, siTIAR and siG3BP1) for 72 h. Cells were treated with arsenate and actinomycin D (ActD + arsenate) for indicated times. For RPLP0, PPIA and VCL mRNA abundance was determined relative to untreated samples by the ΔC_t method. For ACTB, MAPK4 and MYC, changes in mRNA levels relative to input (untreated) controls were determined by cross-normalization to RPLP0 using the $\Delta\Delta C_t$ method. (**B**) The turnover of indicated stress-induced mRNAs was analyzed in U2OS cells transfected as in A. Before monitoring mRNA turnover by ActD addition, cells were pre-stressed by arsenate for 1 h (gray). RNA levels observed after 1 h of arsenate stress were set to one. The decay of mRNAs was analyzed by qRT-PCR using the $\Delta\Delta C_t$ method and RPLP0 for internal cross-normalization as in A. Error bars indicate SD of at least three independent analyses. Statistical significance was determined by Student's t-test. (**C–E**) The abundance of RNAs in stressed (arsenate and ActD for 2 h) and non-stressed U2OS cells transfected as in A was analyzed by comparative microarray analyses. The concomitant knockdown of indicated proteins was confirmed by western blotting with VCL and ACTB serving as internal controls (C). The change of RNA abundance in response to the triple knockdown (siSG, as in A) was determined relative to siC-transfected controls (D). The siSG/siC ratio of transcripts is shown for the average of two independent analyses in stressed and non-stressed U2OS cells. Transcripts with an at least 2-fold distance to the origin in any direction are indicated in green. Transcripts selectively decreased at least 2-fold in the siSG-transfected populations during stress without significantly changed abundance under non-stressed conditions are indicated in red. Transcripts selectively upregulated during stress are indicated in blue. The number of transcripts in the depicted color-coded classes (D) or with an at least 2-fold distance to the origin in the four quadrants (D) was determined without any normalization of array data (E, w/o) or using indicated tools for normalization (E, Mas5 or RMA).

for IGF2BP1 (Figure 6A and B). Consistent with previous studies (16), target mRNA levels were reduced by the knockdown of IGF2BP1, whereas PPIA transcript abundance remained largely unaffected (Figure 6B, gray). The inhibition of SG formation (SGD) itself had no effect on the abundance of any of the analyzed transcripts supporting the view that preventing SG formation does not affect bulk mRNA turnover (Figure 6B, black). Most strikingly, however, only the IGF2BP1 target mRNAs were decreased in their steady-state levels when the SGD was combined with the knockdown of IGF2BP1 (Figure 6B, white). Compared to the knockdown of IGF2BP1 alone, the decrease in target mRNA abundance was similar to the quadruple knockdown of TIA1, TIAR, G3BP1 and IGF2BP1. This suggested that the target mRNA-specific stabilization by IGF2BP1 is facilitated via IGF2BP1-containing mRNPs but independent of SGs. In line with this conclusion one would expect that increasing the abundance of the stabilizing factor IGF2BP1 promotes the stability of target transcripts during cellular stress due to their increased recruitment into mRNPs. This was tested by monitoring mRNA turnover in stressed U2OS cells stably expressing GFP or GFP-ZBP1 (Figure 6C). The overexpression of GFP-ZBP1 (gray) interfered with the turnover of its target mRNAs MYC, ACTB and MAPK4, as evidenced by significantly elevated mRNA abundance compared to GFP-expressing controls. In contrast, the turnover of none IGF2BP1-associated PPIA mRNA remained largely unaffected by the stable expression of GFP-ZBP1.

In conclusion, our findings indicated that the selective stabilization of target mRNAs by IGF2BP1 was dose-dependent but facilitated in a SG-independent manner.

DISCUSSION

In this study, we demonstrate that the stabilization of mRNAs during cellular stress can be uncoupled from the aggregation of mRNPs to cytoplasmic SGs visible by fluorescence microscopy. However, RBPs like IGF2BP1 protect their target mRNAs from decay during cellular stress. This mRNA stabilization is facilitated in a dose-dependent manner suggesting that the RBPs recruit target mRNAs to protective mRNPs, as proposed for IGF2BP1 previously (20,32,33). The protective mRNPs apparently form stable RNA-protein complexes, which can transiently assemble into SGs. Consistently, FRAP and photo-conversion studies reveal that RBPs like IGF2BP1 shielding target mRNAs from degradation, presumably by recruiting these into mRNPs, are stably incorporated in SGs. In contrast, RBPs essential for the assembly of SGs but dispensable for bulk mRNA stabilization, for instance, TIA proteins and G3BPs, only transiently associate with SGs. This suggests that they continuously and dynamically recruit mRNPs to SGs instead of forming a rigid and largely non-dynamic protein scaffold for the assembly of SGs.

SGs are dispensable for the control of bulk mRNA fate in stressed cells

There is accumulating evidence that the assembly of mRNPs in super-structures during cellular stress is largely

dispensable for the control of bulk mRNA translation, whereas relatively little is known about the role of SGs in controlling mRNA turnover [reviewed in (13,34)].

A putative role of SGs in controlling protein synthesis is largely based on the observation that key factors modulating mRNA translation accumulate or transiently associate with SGs [reviewed in (2,24)]. Experimental proof supporting roles of SGs in controlling mRNA translation, however, remain sparse. On the contrary, recent findings provide strong evidence that the assembly of mRNPs in SGs is even dispensable for controlling mRNA translation during cellular stress. When SG formation is inhibited by the depletion of factors essential for their assembly, the block of mRNA translation observed during cellular stress is maintained (10,11). Moreover, recent findings indicate that during the recovery from hypothermia the complete resumption of mRNA translation is largely uncoupled from SG disassembly (12). Consistently, we demonstrate that the inhibition of SG formation by the combined depletion of G3BP1 and TIA proteins neither affects the phosphorylation of eIF2α nor interferes with the stress-induced block of bulk mRNA translation.

As observed for the control of mRNA translation, it was proposed that SGs also modulate the stabilization of bulk mRNA during cellular stress [reviewed in (13)]. However, in yeast the depletion of factors essential for the assembly of SGs does not impair bulk mRNA stabilization during cellular stress (35). In cancer-derived cells, IGF2BP1 is essential for stabilizing target mRNAs during cellular stress, as shown here and previously (16). However, although IGF2BPs are recruited to SGs (16,17), their role in mRNA stabilization during cellular stress is maintained when the formation of SGs is impaired. Likewise, we observe that the turnover of stress-induced transcripts, for instance, HSP-encoding mRNAs, as well as bulk mRNA turnover remains essentially unchanged when preventing the formation of SGs. In conclusion, these findings provide strong evidence that the assembly of mRNPs in SGs is dispensable for modulating mRNA turnover in cancer-derived cells. This suggests that cytoplasmic mRNA fate during cellular stress is largely if not exclusively determined by the recruitment of mRNAs into comparatively stable protein-mRNA complexes termed mRNPs. We propose that the 'caging' of mRNAs in mRNPs transiently protects associated transcripts from being degraded during cellular stress. In support of this view, the increased abundance of mRNP 'guards', for instance, IGF2BP1, enhances the stress-dependent stabilization of mRNAs. Moreover, the formation of comparatively stable 'mRNA cages' is supported by FRAP and photo-conversion analyses. These provide strong evidence that 'protective' RBPs, for instance, IGF2BP1, are barely turned over in SG-recruited mRNPs. Although these mRNPs are assumed to partially cycle between SGs, the cytoplasm and remaining polysomes, they were suggested to be mainly concentrated in the cytoplasm outside SGs (29). How mRNAs are released from their cages and why mRNPs associate into SGs remains to be addressed. However, it is tempting to speculate that posttranslational protein modifications of RBPs provide regulatory triggers allowing the control of mRNA association and mRNP-recruitment to SGs.
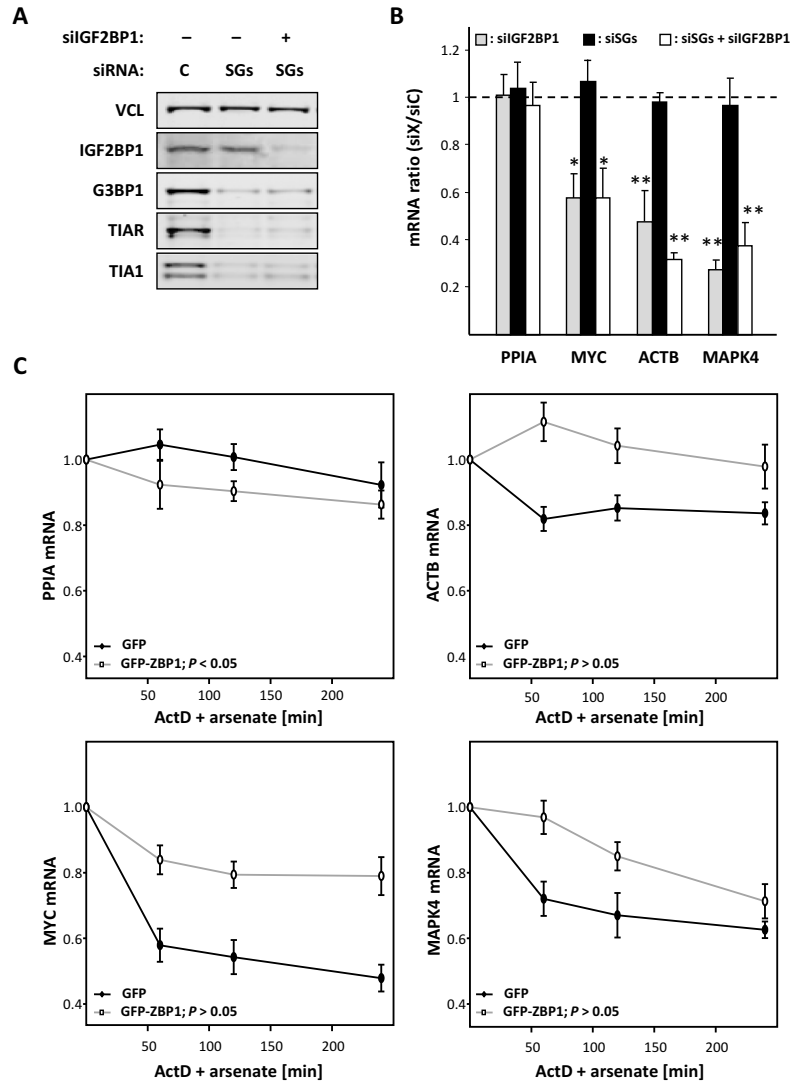
Figure 6. IGF2BP1 stabilizes target mRNAs in a SG-independent manner. (**A**) The knockdown of indicated proteins in U2OS cells transfected with indicated siRNAs or siRNA mixtures (C, control; SG, siTIA1, siTIAR and siG3BP1) for 72 h was analyzed by western blotting. VCL served as loading control. (**B**) The abundance of indicated mRNAs in cells transfected as in (A) and stressed by arsenate for 2 h was determined by qRT-PCR relative to controls (siC) using the $\Delta\Delta C_t$ method and RPLP0 for internal cross-normalization. (**C**) The turnover of indicated mRNAs was analyzed in U2OS cells stably expressing GFP (black) or GFP-tagged ZBP1 (gray) treated with ActD and arsenate for indicated time. The abundance of indicated mRNAs was analyzed by qRT-PCR relative to untreated samples using the $\Delta\Delta C_t$ method and RPLP0 for internal cross-normalization. Error bars indicate SD of at least three independent analyses. Statistical significance was determined by Student's t-test. *$P < 0.05$; **$P < 0.005$.

Prion-like scaffolding versus dynamic recruitment of mRNPs in SGs

SG formation is considered to essentially rely on the aggregation of proteins comprising LC, also termed prion-like, domains via which they assemble into larger structures. Similar to germ cell granules, this aggregation is expected to induce a 'demixing phase transition' resulting in the formation of hydrogel-like structures with a granular morphology, in this case SGs [reviewed in (4,5)]. In agreement with this view, LC domains have been characterized in various RBPs localized to SGs and for some of these proteins aggregation into hydrogel-like superstructures has been demonstrated *in vitro*, e.g. TIA1 (6). Notably, there is accumulating evidence that protein-(RNA) aggregates observed in some human neurodegenerative diseases result from a disturbed SG homeostasis [reviewed in (5)]. The deregulated subcellular sorting and/or mutations in LC domains were suggested to induce aberrant aggregation of LC domain containing proteins like TIA1 or TDP-43 in nuclear and/or cytoplasmic

foci. However, recent findings nuance this view by providing evidence that SG formation essentially relies on excessive, non-polysome associated (m)RNA and consistently is induced by free RNA or single-stranded DNA (9).

If TIA proteins and/or G3BPs would facilitate the assembly of SGs by forming a rigid and barely dynamic prion-like scaffold one would expect a reduced turnover rate of these factors in SGs. At least they should reside in SGs longer than proteins, which are dispensable for the formation of SGs, for instance, IGF2BP1. However, we observe the exact opposite, with TIA proteins and G3BP1 having comparatively short half-lives and an insignificant immobile fraction in SGs. The latter is also supported by previous studies reporting insignificant immobile fractions for TIA proteins and G3BP1 (24). Furthermore, the dynamics of TIA/G3BP proteins in SGs were comparable to those measured in the SG-free cytoplasm. In contrast, RBPs found to be dispensable for SG formation but essential for the selective stabilization of mRNAs during cellular stress, are associated with SGs at 4–5 times longer half-lives and strikingly increased immobile fractions. These observations indicate that instead of forming a rigid glue-like scaffold, TIA proteins and G3BP1 dynamically recruit mRNPs to SGs and thus act as 'mRNP movers'. In support of this view, we and other labs largely failed to purify SGs, whereas the subcomplexes forming SGs, namely, mRNPs, can be isolated from stressed cells by density centrifugation (data not shown). Moreover, we demonstrate that the amount of IGF2BP1 and HUR copurified with G3BP1 during stress is increased. We therefore conclude that G3BP1 and presumably TIA proteins promote the assembly of SGs by the dynamic recruitment of mRNPs to sites of SG formation instead of acting like rigid scaffolds formed by prion-like aggregation. This dynamic recruitment may involve the transient assembly of TIA proteins and G3BPs via ID protein domains. In accord with a redundant role of TIA proteins and G3BPs, the assembly of SGs is only delayed but not prevented when the factors are depleted separately, whereas it is substantially impaired by their concomitant knockdown.

If the presented findings here also apply to cytoplasmic granules observed in neurodegenerative diseases remains to be investigated. Although the formation of SGs in neurons essentially relies on the same proteins (TIA1, TIAR, G3BP1/2 and TTP), their maturation to pathological structures was proposed to involve secondary RBPs like TDP-43, ATXN2 or FUS (36–38). This suggests the assembly and persistence of such granules is modulated in a distinct manner.

SGs: connecting protein synthesis to stress-signaling?

In previous studies we suggested that IGF2BPs prevent target mRNA degradation during cell stress by recruiting target transcripts to SGs (16). Here, we demonstrate that the stabilization of mRNAs during cellular stress can be uncoupled from the assembly of mRNPs in SGs. Moreover, we propose that the formation of SGs requires the dynamic recruitment of cytoplasmic mRNPs. Notably, these findings do not contradict the suggested role of SGs in modulating cellular signaling, for instance, by sequestering protein-kinases, which in the case of mTOR signaling

can antagonize signaling pathways during cellular stress [reviewed in (4,5)]. Moreover, our findings support the hypothesis that SGs play a minor role in modulating mRNA fate during cellular stress. The latter presumably is largely and sufficiently facilitated by mRNPs, as supported by here presented and previous analyses [reviewed in (4,5,13)]. In conclusion, this suggests that SGs, although dispensable for controlling mRNA translation and/or turnover, orchestrate stress signaling and the crosstalk of stalled bulk protein synthesis with protein unfolding during cellular stress. This is supported by the finding that a variety of signaling components, e.g. the mTORC1 complex, RACK1 or the serine/threonine kinase FAST, were recruited to SGs (24,39,40). The latter was shown to have a significant immobile fraction in SGs suggesting a substantial retention time for this kinase in SGs (24). This sequestering of signaling components was, for instance, proposed to enhance cell viability by modestly promoting autophagy via the inhibition of mTOR signaling (41). Furthermore, SGs could provide a dynamic platform for stress-dependent protein modifications by localizing and concentrating signaling molecules of distinct parts of the cell. The mechanisms and consequences of SG-directed regulation of stress signaling remain to be addressed in further detail. Moreover, future studies have to investigate whether protein-RNA aggregates observed in neurodegenerative diseases serve roles in controlling mRNA fate or rather act as signaling modulators as proposed for SGs.

REFERENCES

1. Anderson,P. and Kedersha,N. (2008) Stress granules: the Tao of RNA triage. *Trends Biochem. Sci.*, **33**, 141–150.
2. Kedersha,N., Chen,S., Gilks,N., Li,W., Miller,I.J., Stahl,J. and Anderson,P. (2002) Evidence that ternary complex (eIF2-GTP-tRNA(i)(Met))-deficient preinitiation complexes are core constituents of mammalian stress granules. *Mol. Biol. Cell.*, **13**, 195–210.
3. Kedersha,N.L., Gupta,M., Li,W., Miller,I. and Anderson,P. (1999) RNA-binding proteins TIA-1 and TIAR link the phosphorylation of eIF-2 alpha to the assembly of mammalian stress granules. *J. Cell. Biol.*, **147**, 1431–1442.

4. Kedersha,N., Ivanov,P. and Anderson,P. (2013) Stress granules and cell signaling: more than just a passing phase? *Trends Biochem. Sci.*, **38**, 494–506.

5. Ramaswami,M., Taylor,J.P. and Parker,R. (2013) Altered ribostasis: RNA-protein granules in degenerative disorders. *Cell*, **154**, 727–736.

6. Gilks,N., Kedersha,N., Ayodele,M., Shen,L., Stoecklin,G., Dember,L.M. and Anderson,P. (2004) Stress granule assembly is mediated by prion-like aggregation of TIA-1. *Mol. Biol. Cell.*, **15**, 5383–5398.

7. Tourriere,H., Chebli,K., Zekri,L., Courselaud,B., Blanchard,J.M., Bertrand,E. and Tazi,J. (2003) The RasGAP-associated endoribonuclease G3BP assembles stress granules. *J. Cell. Biol.*, **160**, 823–831.

8. Brangwynne,C.P., Eckmann,C.R., Courson,D.S., Rybarska,A., Hoege,C., Gharakhani,J., Julicher,F. and Hyman,A.A. (2009) Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science*, **324**, 1729–1732.

9. Bounedjah,O., Desforges,B., Wu,T.D., Pioche-Durieu,C., Marco,S., Hamon,L., Curmi,P.A., Guerquin-Kern,J.L., Pietrement,O. and Pastre,D. (2014) Free mRNA in excess upon polysome dissociation is a scaffold for protein multimerization to form stress granules. *Nucleic Acids Res.*, **42**, 8678–8691.

10. Ohn,T., Kedersha,N., Hickman,T., Tisdale,S. and Anderson,P. (2008) A functional RNAi screen links O-GlcNAc modification of ribosomal proteins to stress granule and processing body assembly. *Nat. Cell Biol.*, **10**, 1224–1231.

11. Mokas,S., Mills,J.R., Garreau,C., Fournier,M.J., Robert,F., Arya,P., Kaufman,R.J., Pelletier,J. and Mazroui,R. (2009) Uncoupling stress granule assembly and translation initiation inhibition. *Mol. Biol. Cell.*, **20**, 2673–2683.

12. Hofmann,S., Cherkasova,V., Bankhead,P., Bukau,B. and Stoecklin,G. (2012) Translation suppression promotes stress granule formation and cell survival in response to cold shock. *Mol. Biol. Cell.*, **23**, 3786–3800.

13. Buchan,J.R. and Parker,R. (2009) Eukaryotic stress granules: the ins and outs of translation. *Mol. Cell*, **36**, 932–941.

14. Eisinger-Mathason,T.S., Andrade,J., Groehler,A.L., Clark,D.E., Muratore-Schroeder,T.L., Pasic,L., Smith,J.A., Shabanowitz,J., Hunt,D.F., Macara,I.G. *et al.* (2008) Codependent functions of RSK2 and the apoptosis-promoting factor TIA-1 in stress granule assembly and cell survival. *Mol. Cell*, **31**, 722–736.

15. Thedieck,K., Holzwarth,B., Prentzell,M.T., Boehlke,C., Klasener,K., Ruf,S., Sonntag,A.G., Maerz,L., Grellscheid,S.N., Kremmer,E. *et al.* (2013) Inhibition of mTORC1 by astrin and stress granules prevents apoptosis in cancer cells. *Cell*, **154**, 859–874.

16. Stohr,N., Lederer,M., Reinke,C., Meyer,S., Hatzfeld,M., Singer,R.H. and Huttelmaier,S. (2006) ZBP1 regulates mRNA stability during cellular stress. *J. Cell. Biol.*, **175**, 527–534.

17. Wachter,K., Kohn,M., Stohr,N. and Huttelmaier,S. (2013) Subcellular localization and RNP formation of IGF2BPs (IGF2 mRNA-binding proteins) is modulated by distinct RNA-binding domains. *Biol. Chem.*, **394**, 1077–1090.

18. Stohr,N., Kohn,M., Lederer,M., Glass,M., Reinke,C., Singer,R.H. and Huttelmaier,S. (2012) IGF2BP1 promotes cell migration by regulating MK5 and PTEN signaling. *Genes Dev.*, **26**, 176–189.

19. Zirkel,A., Lederer,M., Stohr,N., Pazaitis,N. and Huttelmaier,S. (2013) IGF2BP1 promotes mesenchymal cell properties and migration of tumor-derived cells by enhancing the expression of LEF1 and SNAI2 (SLUG). *Nucleic Acids Res.*, **41**, 6618–6636.

20. Bell,J.L., Wachter,K., Muhleck,B., Pazaitis,N., Kohn,M., Lederer,M. and Huttelmaier,S. (2013) Insulin-like growth factor 2 mRNA-binding proteins (IGF2BPs): post-transcriptional drivers of cancer progression? *Cell. Mol. Life Sci.*, **70**, 2657–2675.

21. Didiot,M.C., Subramanian,M., Flatter,E., Mandel,J.L. and Moine,H. (2009) Cells lacking the fragile X mental retardation protein (FMRP) have normal RISC activity but exhibit altered stress granule assembly. *Mol. Biol. Cell*, **20**, 428–437.

22. Ralser,M., Albrecht,M., Nonhoff,U., Lengauer,T., Lehrach,H. and Krobitsch,S. (2005) An integrative approach to gain insights into the cellular function of human ataxin-2. *J. Mol. Biol.*, **346**, 203–214.

23. Kwon,S., Zhang,Y. and Matthias,P. (2007) The deacetylase HDAC6 is a novel critical component of stress granules involved in the stress response. *Genes Dev.*, **21**, 3381–3394.

24. Kedersha,N., Stoecklin,G., Ayodele,M., Yacono,P., Lykke-Andersen,J., Fritzler,M.J., Scheuner,D., Kaufman,R.J., Golan,D.E. and Anderson,P. (2005) Stress granules and processing bodies are dynamically linked sites of mRNP remodeling. *J. Cell. Biol.*, **169**, 871–884.

25. Castets,M., Schaeffer,C., Bechara,E., Schenck,A., Khandjian,E.W., Luche,S., Moine,H., Rabilloud,T., Mandel,J.L. and Bardoni,B. (2005) FMRP interferes with the Rac1 pathway and controls actin cytoskeleton dynamics in murine fibroblasts. *Hum. Mol. Genet.*, **14**, 835–844.

26. Zykova,T.A., Zhu,F., Zhang,Y., Bode,A.M. and Dong,Z. (2007) Involvement of ERKs, RSK2 and PKR in UVA-induced signal transduction toward phosphorylation of eIF2alpha (Ser(51)). *Carcinogenesis*, **28**, 1543–1551.

27. Patel,C.V., Handy,I., Goldsmith,T. and Patel,R.C. (2000) PACT, a stress-modulated cellular activator of interferon-induced double-stranded RNA-activated protein kinase, PKR. *J. Biol. Chem.*, **275**, 37993–37998.

28. Matsuki,H., Takahashi,M., Higuchi,M., Makokha,G.N., Oie,M. and Fujii,M. (2013) Both G3BP1 and G3BP2 contribute to stress granule formation. *Genes Cells*, **18**, 135–146.

29. Mollet,S., Cougot,N., Wilczynska,A., Dautry,F., Kress,M., Bertrand,E. and Weil,D. (2008) Translationally repressed mRNA transiently cycles through stress granules during stress. *Mol. Biol. Cell*, **19**, 4469–4479.

30. Atlas,R., Behar,L., Elliott,E. and Ginzburg,I. (2004) The insulin-like growth factor mRNA binding-protein IMP-1 and the Ras-regulatory protein G3BP associate with tau mRNA and HuD protein in differentiated P19 neuronal cells. *J. Neurochem.*, **89**, 613–626.

31. Martin,S., Zekri,L., Metz,A., Maurice,T., Chebli,K., Vignes,M. and Tazi,J. (2013) Deficiency of G3BP1, the stress granules assembly factor, results in abnormal synaptic plasticity and calcium homeostasis in neurons. *J. Neurochem.*, **125**, 175–184.

32. Jonson,L., Vikesaa,J., Krogh,A., Nielsen,L.K., Hansen,T., Borup,R., Johnsen,A.H., Christiansen,J. and Nielsen,F.C. (2007) Molecular composition of IMP1 ribonucleoprotein granules. *Mol. Cell. Proteomics*, **6**, 798–811.

33. Weidensdorfer,D., Stohr,N., Baude,A., Lederer,M., Kohn,M., Schierhorn,A., Buchmeier,S., Wahle,E. and Huttelmaier,S. (2009) Control of c-myc mRNA stability by IGF2BP1-associated cytoplasmic RNPs. *RNA*, **15**, 104–115.

34. Decker,C.J. and Parker,R. (2012) P-bodies and stress granules: possible roles in the control of translation and mRNA degradation. *Cold Spring Harb. Perspect. Biol.*, **4**, a012286.

35. Buchan,J.R., Muhlrad,D. and Parker,R. (2008) P bodies promote stress granule assembly in Saccharomyces cerevisiae. *J. Cell. Biol.*, **183**, 441–455.

36. Wolozin,B. (2012) Regulated protein aggregation: stress granules and neurodegeneration. *Mol. Neurodegenerat.*, **7**, 56.

37. Ash,P.E., Vanderweyde,T.E., Youmans,K.L., Apicco,D.J. and Wolozin,B. (2014) Pathological stress granules in Alzheimer's disease. *Brain Res.*, **1584**, 52–58.

38. Li,Y.R., King,O.D., Shorter,J. and Gitler,A.D. (2013) Stress granules as crucibles of ALS pathogenesis. *J. Cell Biol.*, **201**, 361–372.

39. Arimoto,K., Fukuda,H., Imajoh-Ohmi,S., Saito,H. and Takekawa,M. (2008) Formation of stress granules inhibits apoptosis by suppressing stress-responsive MAPK pathways. *Nat. Cell Biol.*, **10**, 1324–1332.

40. Takahara,T. and Maeda,T. (2012) Transient sequestration of TORC1 into stress granules during heat stress. *Mol. Cell*, **47**, 242–252.

41. Buchan,J.R., Capaldi,A.P. and Parker,R. (2012) TOR-tured yeast find a new way to stand the heat. *Mol. Cell*, **47**, 155–157.

Chapter 3

Object Segmentation

Breakthrough Technologies

Check for updates

PaCeQuant: A Tool for High-Throughput Quantification of Pavement Cell Shape Characteristics[1][OPEN]

Birgit Möller,[a,2] Yvonne Poeschl,[a,b] Romina Plötner,[c] and Katharina Bürstenbinder[c,2]

[a]Institute of Computer Science, Martin Luther University Halle-Wittenberg, 06120 Halle (Saale), Germany
[b]German Integrative Research Center for Biodiversity (iDiv) Halle-Jena-Leipzig, 04103 Leipzig, Germany
[c]Department of Molecular Signal Processing, Leibniz Institute of Plant Biochemistry, 06120 Halle (Saale), Germany

ORCID IDs: 0000-0002-7146-043X (B.M.); 0000-0002-3493-4800 (K.B.).

Pavement cells (PCs) are the most frequently occurring cell type in the leaf epidermis and play important roles in leaf growth and function. In many plant species, PCs form highly complex jigsaw-puzzle-shaped cells with interlocking lobes. Understanding of their development is of high interest for plant science research because of their importance for leaf growth and hence for plant fitness and crop yield. Studies of PC development, however, are limited, because robust methods are lacking that enable automatic segmentation and quantification of PC shape parameters suitable to reflect their cellular complexity. Here, we present our new ImageJ-based tool, PaCeQuant, which provides a fully automatic image analysis workflow for PC shape quantification. PaCeQuant automatically detects cell boundaries of PCs from confocal input images and enables manual correction of automatic segmentation results or direct import of manually segmented cells. PaCeQuant simultaneously extracts 27 shape features that include global, contour-based, skeleton-based, and PC-specific object descriptors. In addition, we included a method for classification and analysis of lobes at two-cell junctions and three-cell junctions, respectively. We provide an R script for graphical visualization and statistical analysis. We validated PaCeQuant by extensive comparative analysis to manual segmentation and existing quantification tools and demonstrated its usability to analyze PC shape characteristics during development and between different genotypes. PaCeQuant thus provides a platform for robust, efficient, and reproducible quantitative analysis of PC shape characteristics that can easily be applied to study PC development in large data sets.

Leaves are the major sites of photosynthesis in most plants and play central roles in carbon fixation and energy supply. In addition, leaves control gas exchange and transport of water and nutrients from roots to shoots (Kalve et al., 2014). From a morphological perspective, leaves are remarkably diverse structures. The diversity is reflected in numerous species-specific shapes, which are reliable traits for taxonomic identification and classification of species (Viscosi and Cardini, 2011; Tsukaya, 2014). Leaf size and shape are not solely determined by genetic variability but also change during development and adapt to environmental conditions (Sultan, 1995, 2000; Cho et al., 2007; Bar and Ori, 2014). Phenotypic plasticity of leaf morphology helps plants to optimize sunlight harvesting, CO_2 gas exchange, and acclimatization to changing ambient temperatures (Nicotra et al., 2010; de Casas et al., 2011). Hence, understanding the cellular and molecular mechanisms of growth regulation is of central importance to improve plant yield, quality, and resource use efficiency.

Because of its high relevance in plant biology, leaf development has been extensively studied in the past decades in many plant species (Bar and Ori, 2014, 2015), including maize (*Zea mays*; Freeling, 1992), tomato (*Solanum lycopersicum*; Gray, 1957), and *Medicago truncatula* (Wang et al., 2008). Genetic and phenotypic analyses, mostly in the model species Arabidopsis (*Arabidopsis thaliana*), provide insights into the dynamics of cellular events that underlie the development from primordia to the final flat and polar organ (Tsukaya, 2010, 2013; Vanhaeren et al., 2015). Growth and development are controlled by complex molecular networks that integrate internal and external signals (Cho et al., 2007; Wolters and Jürgens, 2009). A central role in regulation of expansion is played by the leaf epidermis, which forms a rigid outer coat of the leaf (Savaldi-Goldstein et al., 2007; Kutschera, 2008; Bai et al., 2010; Marcotrigiano, 2010). In several plant species, including Arabidopsis, the leaf epidermis is

composed of three different cell types, which are derived from specialized epidermal progenitor cells: pavement cells (PC), stomatal guard cells, and trichomes, also called leaf hairs (Glover, 2000). Out of these three cell types, PCs are always present in the epidermis and usually the most frequently occurring cell type in the epidermis. PCs can develop highly interlocked, jigsaw-puzzle-like shapes during expansion and provide a structural barrier against mechanical insults. This epidermal barrier is interrupted only by stomata, which form small pores required for gas exchange and transpiration (Glover, 2000).

At the macroscopic level, studies of leaf geometry were one of the first applications for shape analysis in biology (Dale et al., 1971; Ghent, 1973; Kalyoncu and Toygar, 2015). For example, shapes were quantified by use of Fourier-based descriptors (McLellan and Endler, 1998), moments, geometric codes, and margin statistics (Kalyoncu and Toygar, 2015), and leaf serration was quantified using hierarchy analyses (Biot et al., 2016). With improving imaging techniques, several studies addressed the dynamics and changes of PC shape during leaf development and expansion (Iwata and Ukai, 2002; Andriankaja et al., 2012; Elsner et al., 2012; Barbier de Reuille et al., 2015; Vanhaeren et al., 2015). Key regulators of PC interdigitation were identified by mutant phenotyping (Xu et al., 2010; Lin et al., 2012; Li et al., 2013; Guo et al., 2015). In many cases, PC characteristics were described by global descriptors, such as area or circularity, or by skeleton-based approaches (Horiguchi et al., 2006; Staff et al., 2012). These methods oversimplify PC shape and contain no informational value on the number and degree of lobes (Ivakov and Persson, 2013; Wu et al., 2016). To quantify lobe numbers, two main approaches are widely used: skeleton-based detection and manual quantification, which typically underestimate the number of lobes or calculate the number of lobes per area rather than per cell, respectively (Xu et al., 2010; Gao et al., 2015). More recently, a MATLAB-based tool was developed, which aims to provide a platform for objective and robust quantification of lobe number and shape characteristics by analysis of a refined convex hull (Wu et al., 2016). The number of features extracted by these tools, however, is limited to only a few parameters and thus is not capable to cover the shape variability observable in PCs.

In addition, most methods that quantify PC shape rely on manual segmentation of individual cell outlines. The manual preparation has two major disadvantages. First, it is very time consuming and laborious and prevents high-throughput analysis of PC shape. Second, unless segmentation is carried out in blind studies, it is prone to bias by the experimenter (Vanhaeren et al., 2015). Hence, for objective quantification of PC shape characteristics, robust and fully automatic techniques for the segmentation of cell regions, as well as meaningful and clearly defined shape descriptors are required. Powerful descriptors combined with an unbiased automatic segmentation would enable comparative analysis of PC shapes not only within a single set of experiments but also between data from independent studies (Ivakov and Persson, 2013). Ideally, such tool should be user-friendly and run on publicly available open source platforms.

Here, we present our newly developed tool, PaCeQuant, for fully automatic segmentation of individual PCs from confocal input images and simultaneous extraction of 27 different shape parameters. In addition, PaCeQuant optionally offers the analysis of shape characteristics of individual lobes at two-cell and three-cell contact points. The tool is implemented as plugin for the widely used open-source image analysis software ImageJ and publicly available under GPL v3.0. It features a graphical user interface for user-friendly data input of large data sets. We provide a supplemental R script for extended data analysis (including statistical analysis) and comfortable data visualization by box plots or violin plots. Extensive comparative evaluations prove that PaCeQuant is able to produce high-quality segmentations of PCs and to provide robust, reliable, and reproducible quantification of PC shape characteristics suitable for shape analysis during development and for mutant phenotyping.

RESULTS

Overview of PaCeQuant Workflow

The workflow for cell segmentation and feature extraction implemented in PaCeQuant is composed of two parts. In the first part, we offer an optional automatic segmentation of PCs from confocal microscopy input images, which consists of four basic stages covering (I) improvement of the image quality, (II) cell boundary enhancement, (III) binarization and morphological postprocessing, and (IV) region filtering. In the second part, we provide a pipeline for feature extraction from segmented cell regions, which runs on cell regions, either defined by the automatic segmentation, after manual correction of automatic segmentation results or from manually segmented input images (Fig. 1).

We implemented PaCeQuant in a platform-independent fashion in Java as part of the Microscope Image Analysis Toolbox (MiToBo; Möller et al., 2016). Most of the individual algorithms and processing steps that are part of the complete workflow are directly implemented as operators in MiToBo. If external functionality is used, it is explicitly indicated below. The PaCeQuant tool is publically available under GPL License 3.0 as part of the MiToBo distribution at MiToBo's website (http://www.informatik.uni-halle.de/mitobo/) and integrates seamlessly in ImageJ/Fiji (Schindelin et al., 2012) via MiToBo's own update site (see "Installation and Usage of PaCeQuant" in "Materials and Methods" for more information). The source code of MiToBo and the PaCeQuant plugin is available from the MiToBo website or on Github (https://github.com/mitobo-hub/mitobo); detailed documentation can be found

A

1. Contrast Stretching

2. Gaussian Filtering

Noise reduction by convolution with Gaussian kernel

Image Quality Improvement (I)

3. Vesselness Enhancement Filtering

Mexican Hat profile in x, constant profile in y

Boundary Enhancement (II)

4. Niblack Binarization + **5.** Post-Processing

Input image (inverted)   Binary image (inverted)

6. Gap Closing + **7.** Post-Processing

Binarization & Morphological Post-Processing (III)

8. Component Labeling

Group pixels into components based on pixel connectivity.

9. Region Filtering

Remove regions below or above a size threshold or touching the image border.

Region Filtering (IV)

10. Feature Extraction

Global Features	Contour-Based Features	Skeleton-Based Features	Pavement Cell-Specific Features
area, perimeter, length, width, circularity, eccentricity, etc.	margin roughness avg contour concavity sd contour concavity	longest skeleton path branch count avg branch length avg end-point distance	lobe count, avg lobe length, avg baseline lobe width, minimal/maximal core width, non-lobe area, etc.

Feature Extraction (V)

B

Input Image (0.)   **Intermediate Image (7.)**   **Output Image (9.)**
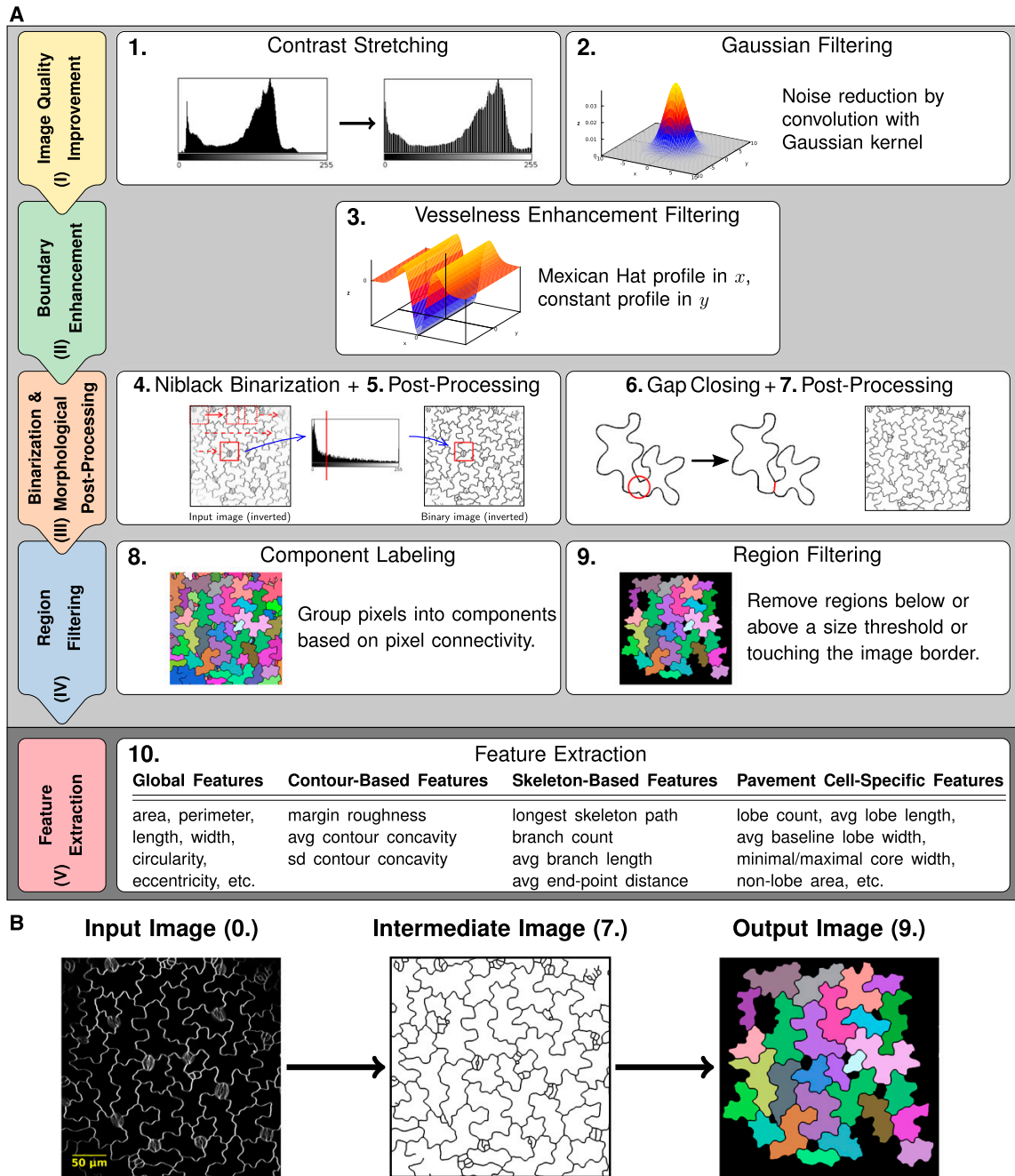
50 µm

Figure 1. Workflow of cell segmentation and feature extraction. A, Workflow implemented in PaCeQuant for automatic detection of cell outlines (part 1, light gray) and extraction of shape features (part 2, dark gray). In the first part, input images are processed in four basic stages (image quality improvement [I] to region filtering [IV], left column) via nine individual processing steps (panels on the right). In the second part, the feature extraction (step 10) is performed. B, Example of an input image (step 0), the processed image after outline extraction (step 7), and an image containing identified and filtered regions (step 9) used for feature extraction. For an overview of all image-processing steps, see Supplemental Figure S1.

85

on PaCeQuant's webpage (http://mitobo.informatik. uni-halle.de/index.php/Applications/PaCeQuant). PaCeQuant takes full benefit of all built-in features of MiToBo and its core library Alida (Posch and Möller, 2017). Alida provides a framework for the development of data analysis applications in a modular fashion and supports the automatic generation of graphical user interfaces for operators (Möller and Posch, 2013). Since Alida also provides command line interfaces for all implemented operators, PaCeQuant can be run optionally without graphical interaction, e.g. remote on a server.

Cell Boundary Segmentation and Region Filtering

To develop an automatic segmentation method, we used confocal images as input (Fig. 1B), which is a common method for cell shape analysis (Xu et al., 2010; Wu et al., 2016). We initially focused our studies on cotyledons of 5-d-old Arabidopsis seedlings. Cotyledons resemble the characteristics of true leaves in many aspects and have therefore been developed into a popular model system to study leaf development (Tsukaya et al., 1994). Cell outlines were visualized by staining of cotyledons with the lipophilic dye FM4-64 and imaged as single optical sections covering groups of adjacent epidermis cells at a resolution of 2.2 to 3.2 pixels/μm.

To extract cell boundaries, images are processed as follows. First, we use a global contrast stretching to enhance the contrast of the input images (step 1), followed by applying a Gaussian filter (with SD $\sigma = 1$) for noise removal (Fig. 1A; Supplemental Fig. S1; step 2). In stage (II), we further enhance the cell structures by vesselness enhancement filtering (step 3), which is used typically to enhance thin elongated structures, such as, e.g. blood vessels, in digital images (Chaudhuri et al., 1989; Zhang et al., 2010; Fraz et al., 2012). Here, we apply a filter kernel, as proposed in Sofka and Stewart (2006). The rectangular kernel mask is composed of a Mexican hat profile in normal and a constant profile in tangential direction. The Mexican hat profile is defined by a second-order Gaussian derivative (Fig. 1A; Supplemental Fig. S1). Each image is convolved with the filter mask in 18 different orientations from 0° to 170°, in successive steps of 10° increase. The resulting 18 filter responses are joined into a single filter response image by selecting for each pixel the maximal response to any of the 18 filters while setting negative maxima to zero.

The filtered image is postprocessed with a local median filter of radius r = 1 pixel to account for locally varying contrasts. To segment cell boundaries, in stage (III) a local Niblack binarization is applied to the filter response images (Niblack, 1986; Fig. 1A; Supplemental Fig. S1; step 4). The Niblack algorithm only yields reliable thresholds if intensity variance is present within the local sliding window. Thus, we introduced an additional test for local variance prior to threshold extraction, and thresholds are only calculated for window positions with sufficient local variance. All other positions are classified as background. To account for small breaks in

the boundaries, the resulting binary image is dilated (r = 3 pixels) and eroded (r = 5 pixels). Very small components (e.g. dots, linear structures, etc.), which most likely refer to noise artifacts, are eliminated and the detected boundary components are thinned to a width of 1 pixel (step 5).

An optional gap-closing step may be subsequently performed on the 1-pixel-boundary images to eliminate larger gaps in the cell boundaries (step 6). For this, two different heuristics are implemented in PaCeQuant, a very simple heuristic based on end-point distances and a second heuristic, relying on a watershed transformation on binary images. In the first case, all boundary end points with an empirically determined maximum distance of d = 20 pixels are linked to close gaps in between. In the second case, we apply a Euclidean distance transformation to the binary image, followed by a watershed transformation on the distance image using the implementation available in ImageJ according to Leymarie and Levine (1992). To remove implausible boundaries resulting from oversegmentation of the watershed segmentation, we apply a combination of different criteria for filtering boundary segments. A segment is preserved only if it existed in the binary image already, if it is shorter than 40 pixels and its end points are located close to previously detected end points, if it extends a skeleton branch with a sufficient length, or if it crosses an image region with sufficiently dark intensity values most likely referring to a cell boundary. In the resulting binary image, individual cell regions are segmented according to the postprocessed watersheds detected in the distance image. To ensure that neighboring cell regions are not merged, boundaries are dilated (radius r = 3 pixels) and holes within regions are filled (step 7). This is achieved by applying a component-labeling algorithm to each detected region, where the region is treated as background and potential holes are identified as foreground components. By adding their pixels to the corresponding region, holes are eliminated. In the final step of stage (III), cell boundaries are thinned again to a width of 1 pixel using the skeletonization algorithm of ImageJ (Zhang and Suen, 1984), short open branches are removed, and boundaries are extended to a final width of three pixels. As a result, the grayscale input image is converted into a binary output image where the cell boundaries are represented by black pixels and the cell regions by white pixels.

Stage (IV) starts with assigning a unique identifier to each detected cell region by applying component labeling to the binary image resulting from the previous stage (Fig. 1B; step 8). Regions with boundaries that exceed the image border are removed since these reflect incomplete cells (Fig. 1; Supplemental Fig. S1; step 9). The remaining regions are filtered for size to remove unspecific regions that likely reflect noise from the data set.

Quantification of Shape Characteristics

To analyze and quantify shape characteristics, we implemented the automatic extraction of 27 features for

each segmented cell in part 2 (or alternatively stage IV) of our PaCeQuant workflow (Figs. 1 and 2; part 2 (or alternatively stage IV), step 10). The features cover four major groups: (A) global features, (B) contour-based features, (C) skeleton-based features, and (D) special features of PCs, e.g. the number and length of lobes (Fig. 2A; Supplemental Table S1). Group A features characterize the cell as a whole with a focus on its global shape (e.g. circularity, degree of elongation). In group B, we quantify the boundary characteristics, such as changes in local contour concavity and tangent orientations, which correlate with the degree of contour folding. The third group of features analyzes shape characteristics relative to the skeleton of the cell, which we extract using the algorithm of Zhang and Suen (1984). The skeleton is defined as the 1-pixel-wide main axis or backbone of a cell region. It consists of a set of branches, end points, and branch points and is frequently used to quantify morphological characteristics of regions in images (Marchand-Maillet and Sharaiha, 1999; Xiong et al., 2010). In the set of group C features, we included the length of the longest path in the skeleton, the number of branches (as a rough estimate for the number of lobes), the average branch length, and the average radius of the lobes associated with detected branches. In group D, we extract specific features for PCs, i.e. their lobe and neck characteristics.

Definition of PC-Specific Features

The automatic extraction and analysis of features suitable to characterize PC shapes requires a unique definition of relevant components of PCs involved in the calculation of shape features, e.g. of lobes or neck regions. So far, however, most PC shape characteristics lack such clear definitions, and only vague ideas are reported in the existing literature that discuss how lobes are localized along a contour or how their exact dimensions are quantified. Similarly, a clear definition for the neck width of PCs, which is a common measure for the growth restriction at neck regions, is lacking. Instead, the neck width is calculated based on manually selected line segments that appear reasonable to the individual user (Li et al., 2003; Bannigan and Baskin, 2005; Fu et al., 2005). This quantification is highly biased since the selection and definition of the line segments is highly variable. Thus, we first provide clear definitions of all relevant components of a PC, which form the fundament for the set of PC-specific features automatically extracted by PaCeQuant.

We define lobes and necks along the cell contour based on local curvature orientation (Fig. 2, B–E). First, the local curvature is estimated for each point along the contour by applying the robust algorithm of Freeman and Davis (1977) available in MiToBo. Resulting curvature values are convolved with a Gaussian filter mask to smooth the profile, which increases the quality of feature extraction from curvature values. Values for positions where no curvature is extracted are

interpolated by a nearest neighbor approach. The resulting contour contains concave and convex segments that are defined by positive and negative curvature values, respectively. Lobes and necks are detected by analyzing the signs of local curvature (Fig. 2B). Points where signs change refer to inflection points connecting the apical parts of lobe contours (Fig. 2C). We define a neck point as the center point of a convex contour segment (Fig. 2E). This definition turned out to be more robust than selecting the point with a maximal curvature value since local curvature values may suffer from noise and numerical instabilities due to discretization. The distance between two adjacent neck points, termed lobe baseline, represents the basal width of an individual lobe. If the lobe baseline intersects with the background of the image (Fig. 2E), we heuristically adapt the positions of the boundary points of a lobe section, resulting in *virtual neck points*. In detail, the neck points defining a lobe section are virtually shifted toward each other along the contour until the baseline connecting the two virtual points no longer intersects with the image background. By shifting a neck point to a new virtual position, the original neck point is expanded to a *neck region*, which is flanked by the original neck point and the new virtual neck point. If shifts for both adjacent lobe regions of a neck point are necessary, two virtual neck points will flank the neck region.

A lobe is defined as the contour segment between two adjacent neck points. To increase detection robustness and to distinguish lobes from unspecific curvature variations due to segmentation inaccuracies, we set the threshold for lobe detection by default to contour segments with a length of ≥8 pixels. If necessary, the threshold can be adjusted by the user. The number of inflection points divided by two refers to the number of lobes. The line connecting the two inflection points of a lobe defines the lobe equator and represents the equator width of a lobe (Fig. 2C). The lobe length is calculated as the maximal distance between the lobe baseline and the lobe contour (Fig. 2D). In addition, we calculate the distances between baseline and equator, and equator and contour as estimates for the basal and apical lobe lengths (Supplemental Table S1).

The area enclosed by the baseline and the contour segments of the lobe represents the lobe region. As a measure for the relative proportion of lobes to the total cell area, we calculate the nonlobe area, which yields a first estimate of the growth restriction/expansion of the core cell region (i.e. the region between the necks). The nonlobe area is the area of a cell enclosed by the lobe baselines and, if neck points were virtually shifted, by additional pixels of the resulting neck regions (Supplemental Table S1). As an additional value for the size of the cellular core region, we quantify the width of the core region. For this, we utilize the parts of the region skeleton not belonging to any branch, i.e. the core skeleton, as an estimate of the cellular core region (Fig. 2F). We exclude the skeleton branches as they usually refer to lobes of a region. For each pixel of the core skeleton, the Euclidean distance to the background
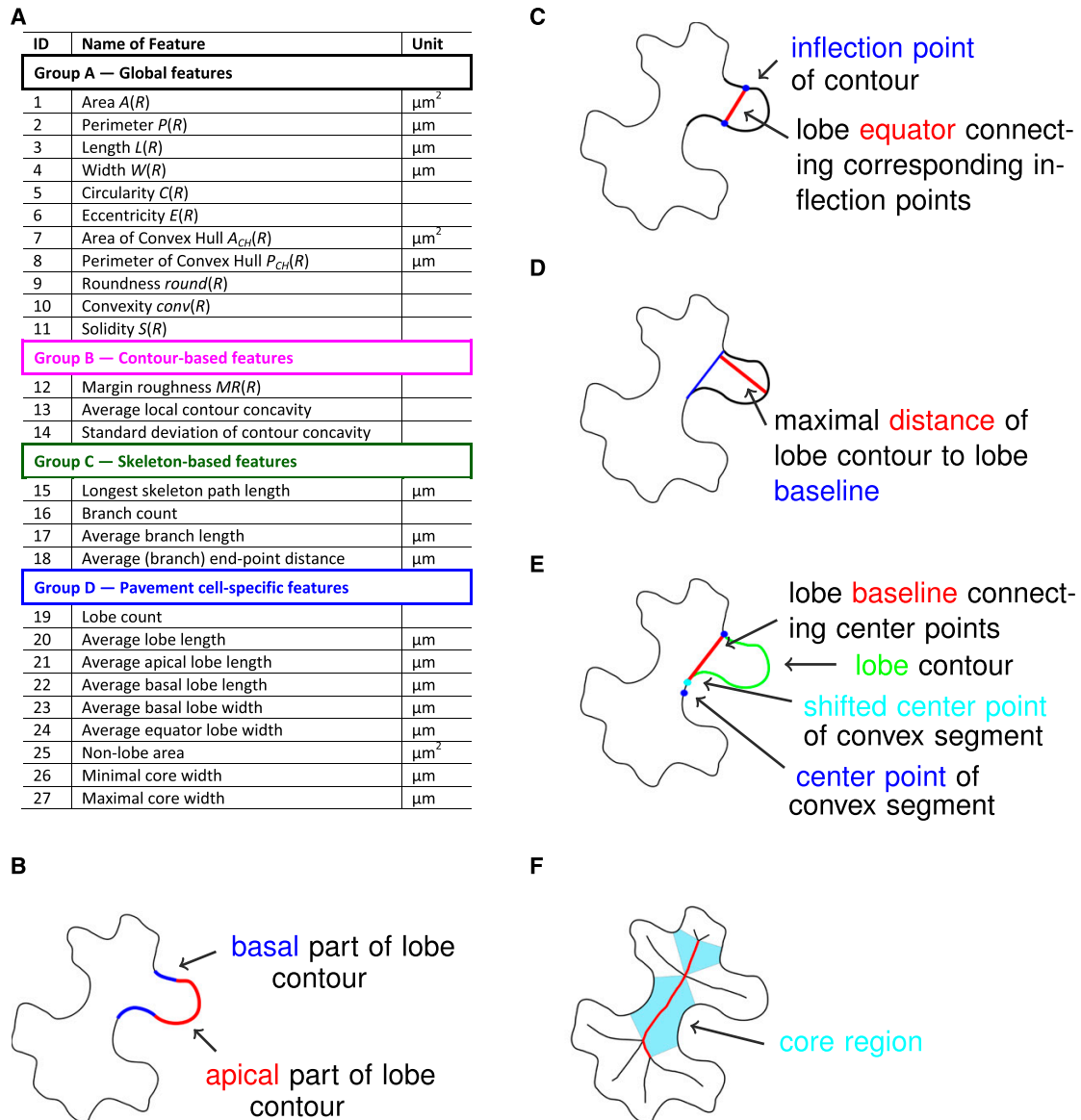
A

ID	Name of Feature	Unit
Group A — Global features		
1	Area $A(R)$	μm^2
2	Perimeter $P(R)$	μm
3	Length $L(R)$	μm
4	Width $W(R)$	μm
5	Circularity $C(R)$	
6	Eccentricity $E(R)$	
7	Area of Convex Hull $A_{CH}(R)$	μm^2
8	Perimeter of Convex Hull $P_{CH}(R)$	μm
9	Roundness $round(R)$	
10	Convexity $conv(R)$	
11	Solidity $S(R)$	
Group B — Contour-based features		
12	Margin roughness $MR(R)$	
13	Average local contour concavity	
14	Standard deviation of contour concavity	
Group C — Skeleton-based features		
15	Longest skeleton path length	μm
16	Branch count	
17	Average branch length	μm
18	Average (branch) end-point distance	μm
Group D — Pavement cell-specific features		
19	Lobe count	
20	Average lobe length	μm
21	Average apical lobe length	μm
22	Average basal lobe length	μm
23	Average basal lobe width	μm
24	Average equator lobe width	μm
25	Non-lobe area	μm^2
26	Minimal core width	μm
27	Maximal core width	μm

C

D

E

B

F

Figure 2. List of PaCeQuant cell shape features and basic definitions of PC-specific features. A, List of cell shape features extracted by PaCeQuant and their units. For a detailed description of feature characteristics, see Supplemental Table S1. B to F, Basic definitions used for the quantification of PC-specific shape features, including (B) apical and basal parts of a lobe, (C) lobe equator, (D) lobe baseline, (E) neck point and neck point correction, and (F) core region dimensions.

is calculated, and the width of the core region is double the Euclidian distance. To characterize the core region more globally, we sort all distances along the core skeleton in ascending order. Since the region skeleton is very sensitive to changes in contour shape, extreme values of this sorted list, i.e. the minimum and maximum distances, often refer to outliers. To increase the robustness of the core width values, we thus extract the first and third quartile entries as

estimates for the minimum and maximum core widths, respectively.

Accuracy of the Automatic Detection

To assess the quality of PaCeQuant's segmentation approach, we compared the results from the fully automatic cell segmentation with manually extracted cell

outlines. For a subset of 15 individual cells from one input image (data set 1) we calculated recall and precision measures of segmented cell areas and distance measures between manually and automatically extracted cell contours (Fig. 3; Supplemental Table S2). The recall quantifies the fraction of the cell area in the manually segmented cell, which has also been extracted by the automatic approach. Vice versa, precision quantifies the ratio of the area detected by automatic segmentation, which is also part of the manually segmented cell. For both measures, a value of 1 represents an optimal fit. To compare the similarity of the two samples, we calculated the F1-score (also known as Sørensen-Dice index), which is the harmonic mean of recall and precision and ranges from 0 to 1, with 1 representing highest similarity. To compare the accuracy of the detected contours, we computed the Hausdorff distance, which measures the (dis)similarity between two sets of points. For each point of one set it extracts the minimal distance to a contour pixel in the second set, and vice versa. Next, the maximum over all minimal distances is identified within the two sets, hence the Hausdorff distance refers to the maximal distance that occurs between two contours at any position.

For the 15 cells in our evaluation set (Fig. 3A) we found an average recall of 0.976 ± 0.0088 (SD) and an average precision of 0.973 ± 0.0121, resulting in an average F1-score of 0.975 ± 0.0079 (Supplemental Table S2). The values are consistent, with a large overlap between manual and automatic segmentation (Fig. 3B, left and middle). For 12 of the 15 test cells, the Hausdorff distance was smaller than 6.5 pixels. For 3 cells, distances of 12.2, 11.7, and 44.2 pixels were found, resulting in contour changes of maximum 1 or 2 μm. These are due to short contour segments where larger deviations between manual and automatic segmentation appear, often caused by low contrasts along the contour or by stomata accidentally classified as part of a cell (Fig. 3B, right).

To further validate the accuracy of the automatic segmentation and to assess the impact of differences in segmentation on feature values, we performed a pairwise comparison of all 27 extracted features between manually and automatically segmented cells (Fig. 3C; Supplemental Fig. S2). For the 11 global features (group A), the values from manual and automatic segmentation are almost identical in all 15 cells, as revealed by straight diagonal lines in the scatterplots. Thus, short sections at which automatically segmented boundaries deviate from manual cell boundaries have only minor effects on this feature class. The three other groups of contour-based (group B), skeleton-based (group C), and
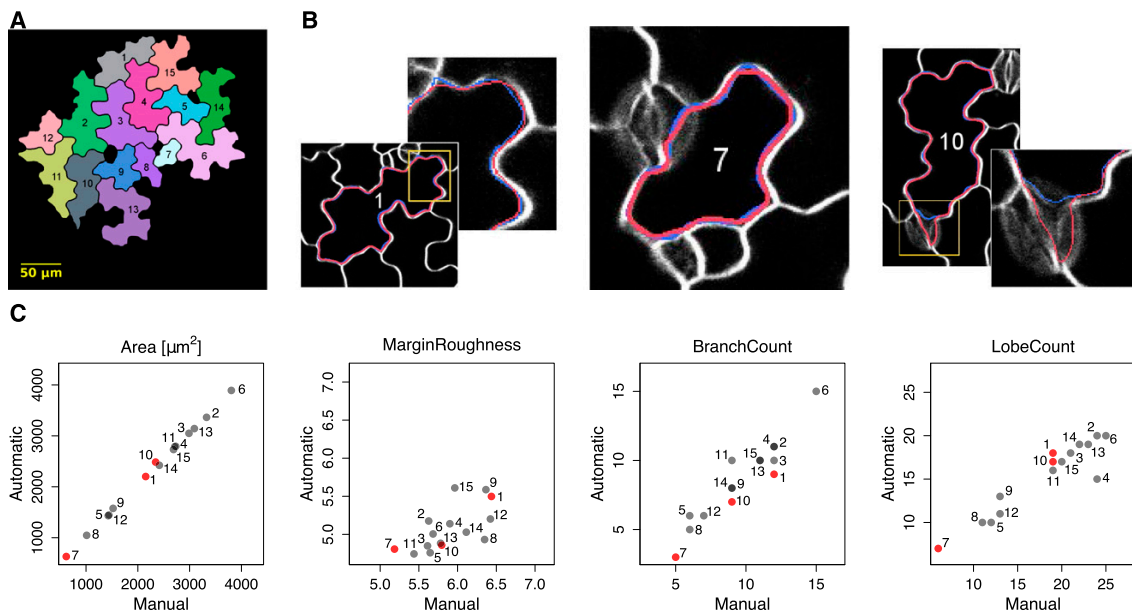
Figure 3. Comparison between automatic and manual segmentation of cells based on PaCeQuant features. A, Sample image containing 15 individual cells, which were segmented fully automatically and manually. Numbers correspond to cell identifiers. B, Overlay of cell outlines detected by automatic (red) and manual (blue) segmentation for sample cells with high congruence and different cell sizes (left, ID 1, large cell; and middle, ID 7, small cell) and with local deviations (right, ID 10). C, Scatterplots of one exemplary feature for each of the four feature groups (see Fig. 2 and Supplemental Table S1) from a pairwise comparison between automatic and manual segmentation of all 15 cells. Cells shown in B are highlighted in red (for a summary of all features, see Supplemental Fig. S2).

PC-specific (group D) features are slightly more sensitive to the different segmentation methods. A tendency for higher values in manually segmented cells is seen for margin roughness (group B), as indicated by an accumulation of points below the diagonal. Variation without a clear tendency toward the segmentation method is seen for "average branch length" and "average end point distance" (group C), and "average lobe length" (group D; Supplemental Fig. S2). Notably, feature values sometimes differ significantly even for cells with a high overlap of cell boundaries between manual and automatic segmentation, as seen e.g. in cell ID 7 for "average branch length" (group C) or "average lobe length" (group D; see Fig. 3B; Supplemental Fig. S2; Supplemental Table S2). The differences are consistent with a general sensitivity of skeleton-based features (group C) to even small changes in region shape (Gonzales and Woods, 1996). Likewise, the quantification of contour-based features (group B) and PC-specific features (group D) relies on analyzing curvatures and concavities of the region boundary and thus is sensitive to small variations in shape contours (e.g. Utcke, 2003).

Variance can be generated by manual segmentation due to bias caused by different persons and even by a single person between segmentations generated at different time points (Vanhaeren et al., 2015). The accuracy of manual segmentation further depends on the sampling density, i.e. the frequency of sampling points along the cell boundary. To reduce the impact of geometrical discontinuities in manual segmentation, extracted contours are commonly smoothed, e.g. using ImageJ's spline fitting (Wu et al., 2016), which additionally affects the curvature of the contour. Automatic segmentation, on the other hand, mostly depends on the image quality. Local regions of low contrast are prone to segmentation inaccuracies, and the optional watershed gap closing potentially leads to over-segmentation of actual cell borders. In a direct comparison, however, bias introduced by PaCeQuant is identical for all images, and PaCeQuant reproducibly detects cell regions from experiments conducted at different times. Thus, our first experiment suggests that PaCeQuant is suitable to detect cell outlines automatically and to quantify cell shape characteristics with a very high quality.

The automatic segmentation step works best with input images with a high image quality, which sometimes is difficult to obtain, e.g. when leaves are curled or wavy. In such cases, results from the automatic segmentation can be exported from PaCeQuant and can be manually corrected in ImageJ to remove local inaccuracies. PaCeQuant supports two formats for exporting segmentation results. Segmented cell regions can be saved as ImageJ ROI files or exported as label images, which can be edited directly in ImageJ via the ROI manager or allow for manual corrections by editing pixel intensities in, e.g. programs like Gimp, respectively. Alternatively, if only a few cells are detectable in the input image or if input images are not compatible

with the automatic segmentation method implemented in PaCeQuant, as is the case, e.g. for agarose imprints of cell outlines, segmentation can be conducted manually in ImageJ. In both cases, manually corrected or manually defined cell regions can be imported into PaCeQuant for automatic feature quantification and data analysis.

Precision of Lobe Detection

The formation of lobes and indentations is a specific property of PCs. Thus, lobes are quantified commonly to describe shape characteristics, mostly by manual counting or by skeleton-based approaches, which are prone to bias and often inaccurate. Recently, the Lobe-Finder program was released, which automatically measures lobe numbers by analysis of a refined convex hull (Wu et al., 2016). However, the LobeFinder tool still relies on manual segmentation of individual cells and tends to underestimate lobes when compared to manual counting (Wu et al., 2016). To investigate the quality and suitability of lobe detection based on sign changes of curvature, which we implemented in PaCeQuant, we compared the efficiency of PaCeQuant's lobe detection with LobeFinder and with manual lobe counting using the sample set of 15 cells (Fig. 4). For manual counting, lobes were quantified in the original input image by four individuals with expertise in PC shape analysis in a blind study. To compare PaCeQuant with LobeFinder, cell boundaries automatically extracted by PaCeQuant were used as input, and all features for which quantification algorithms are implemented in both tools (e.g. area, solidity, and circularity) were quantified and compared.

For all three approaches, a tendency for increased lobe counts was detected with increasing cell size (Fig. 4A). Notably, manual counting resulted in large differences between individual persons in the number of detected lobes for some of the analyzed cells (Fig. 4, A and B). The deviations within the four sets of manually detected lobes increased with increasing size and complexity of the analyzed cells. For some cells (e.g. cell ID 15, ID 14, and ID 6) manual counting resulted in differences of 8 to 11 lobes per cell, which refers to a deviation of up to 100% for single cells. Nevertheless, a significant number of lobes is consistently detected in all four manual annotations and by both tools, as shown for cell ID 2 (Fig. 4B). PaCeQuant and LobeFinder are in the range of lobe numbers counted by the four individuals. In general, PaCeQuant tends to detect more lobes than LobeFinder, and lobe numbers detected by PaCeQuant are in the upper range of manual lobe counts (Fig. 4, A and C). For all other features, such as area, solidity, and convex hull characteristics, we obtained identical results with LobeFinder and PaCeQuant, as expected for identical input contours (Fig. 4C; Supplemental Fig. S3). Thus, we conclude that PaCeQuant is suitable to quantify lobes efficiently. The comparison of the three methods further highlights the subjectivity of manual lobe counting and demonstrates a requirement for automatic
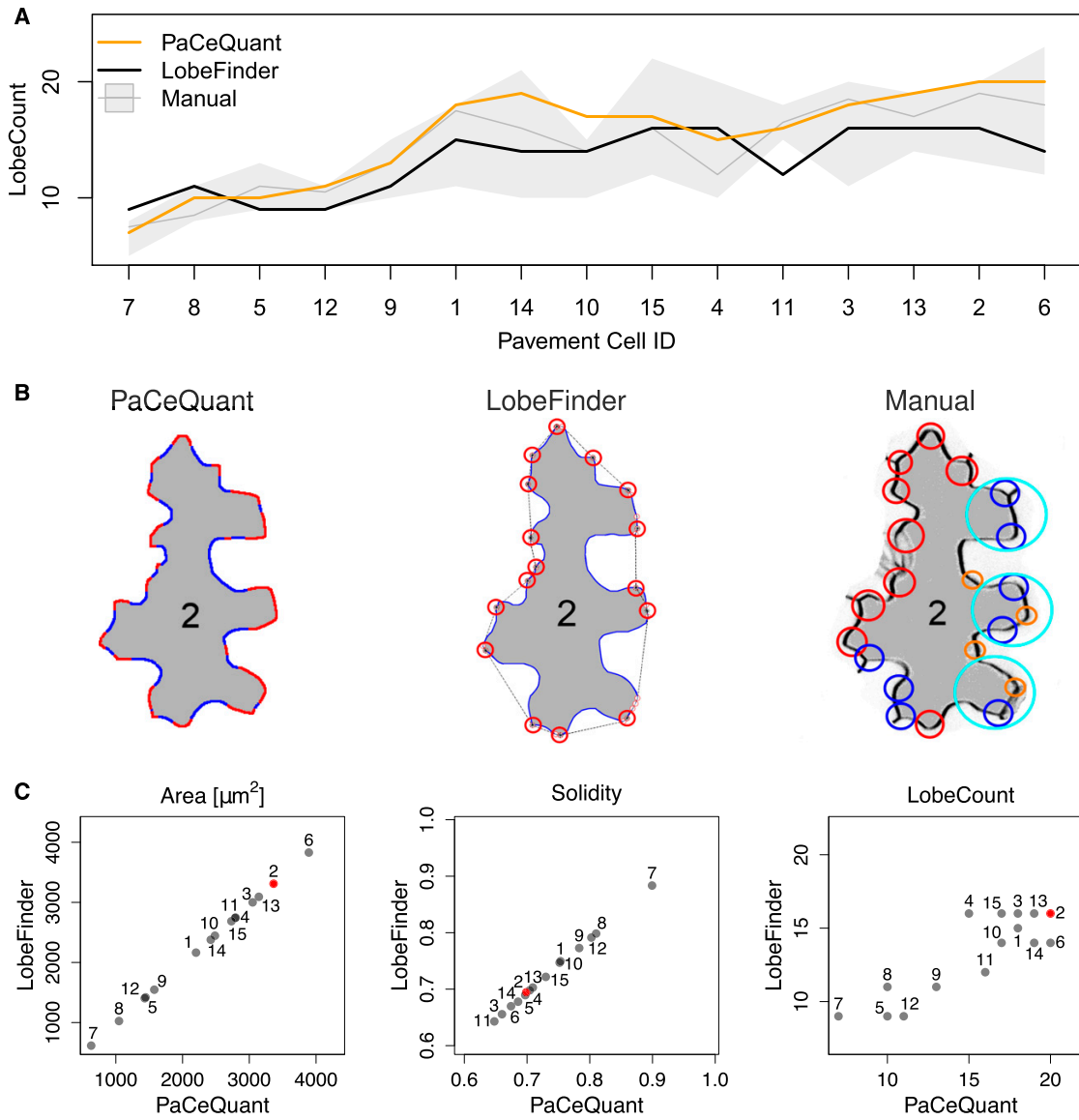
90

Figure 4. Evaluation of lobe detection accuracy by comparison between PaCeQuant results with LobeFinder and manual lobe counting. A, Number of lobes detected in the 15 sample cells (ID 1–ID 15) after automatic segmentation by PaCeQuant (orange), LobeFinder (black), and by manual lobe counting (gray). For manual counting, lobes were analyzed by four independent researchers. The gray line represents the mean lobe number per cell of the four measurements; the gray strip represents the range of the independent measurements. Cells are sorted by their area from small (left) to large (right). B, Lobe count results in one exemplary cell (ID 2) analyzed with PaCeQuant (left, 20 lobes), LobeFinder (middle, 16 lobes), and manually (right, 13–20 lobes). Lobes identified by PaCeQuant or LobeFinder are marked in red. Lobes identified manually are marked in red (nine lobes), blue (eight lobes), pink (four lobes), and turquoise (three lobes) if identified by four, three, two, or at least one person, respectively. C, Pairwise comparison of features computed by PaCeQuant and LobeFinder in the sample set of the 15 automatically segmented cells. Scatterplots are shown for area, solidity, and lobe count (for a summary of all features, see Supplemental Fig. S3). The cell shown in B is highlighted in red.

lobe quantification to generate comparable and reproducible data with a constant bias.

Application of PaCeQuant to Large Data Sets

The main goal of PaCeQuant is to enable high-throughput analysis of PC characteristics. We thus validated the quality of PaCeQuant results on a larger data set (data set 2) consisting of 14 individual images of PCs from cotyledons of 5-d-old wild-type seedlings (Supplemental Fig. S4). In total, 373 cells were identified in the sample set within a time span of 10 min, starting from data input to result output, which refers to 1.6 s per cell. As seen for the initial set of 15 cells (see Fig. 3), some cells, in particular cells adjacent to stomata or with regions of low contrast, were not detected correctly (Supplemental Fig. S4B). To study the impact of such detection inaccuracies on the overall feature quantification we manually filtered the automatic PaCeQuant segmentation results. Thirty-one cells showed locally incorrect cell borders and failed to pass the manual quality control, which represents ≤10% of all detected cells. A pairwise comparison of the distributions for all 27 features between the unfiltered data set, and the manually filtered data set revealed that the manual filtering step has only minor effects on feature distributions and set properties, which are not statistically significant (P values of 0.623–0.997; Supplemental Fig. S4A). Thus, detection errors within individual cells are compensated by the analysis of large data sets, which can easily be generated with PaCeQuant, and time-intense manual filtering is not required for reliable PC shape quantification with PaCeQuant. Compared to manual segmentation, PaCeQuant offers the advantage of a much faster and unbiased segmentation. Thus, PaCeQuant is capable of increasing the amount of quantitative data, which improves the power of statistical analyses and guarantees a larger objectivity and reproducibility of extracted data.

Analysis of Cell Shape Characteristics during Development

To assess the usability of high-throughput cell shape analysis in a biological context, we applied PaCeQuant to a developmental series of Arabidopsis cotyledons (Fig. 5). We analyzed PCs of the adaxial side of wild-type cotyledons at stages of early cell expansion (3 d after germination [DAG]), of rapid expansion (5 DAG), and at a stage with first fully expanded cells (7 DAG; data set 3; Zhang et al., 2011). At 3 DAG, cells range in size between 245 and 2,320 μm^2, with 90% of the cells being smaller than 1,400 μm^2 (Fig. 5, A and B). At 5 DAG, cells span sizes between 245 and 6,367 μm^2. At this stage, 90% of the cells are smaller than 4,042 μm^2, and approximately 50% of all detected cells range in size between 1,400 and 4,042 μm^2. The number of small cells decreased to <40% when compared to 3-d-old

seedlings. At 7 DAG, roughly one-third of the detected cells belongs to the groups of small and medium-sized cells each, and the last third consist of cells with sizes larger than 4,040 μm^2, ranging up to 12,600 μm^2. The detected cell sizes are in the range of previous reports (Zhang et al., 2011), which further demonstrates the accuracy of the PaCeQuant measurements. In leaves, neighboring cells differ largely in their shapes ranging from small and simple-shaped cells to large and highly complex cells (Elsner et al., 2012). Consistent with large differences in cell size and cell differentiation, we observed high variability within the feature values calculated from analysis of the complete set of detected cells as input (Fig. 5, C–E, Supplemental Fig. S5A). The variability increases with increasing age of the analyzed cotyledons (Supplemental Fig. S5) and reflects the increasing diversity of cell shape and size during later stages of cotyledon (Zhang et al., 2011) and leaf growth (Elsner et al., 2012).

We next aimed to analyze PC shape transitions during development of cotyledons in more detail. To quantify shape characteristics in cells at similar stages of cellular expansion, we categorized them into small, medium, and large cells by applying the 90% thresholds from 3 DAG and 5 DAG as small size threshold (t_s) and medium size threshold (t_m), respectively (Fig. 5B). We analyzed the 27 PC shape features in the three sets and in the three size categories and compared the feature distributions by applying the Dunn's test and by adjusting P values with the Benjamini-Hochberg procedure (Fig. 5, C–E; Supplemental Fig. S5). As expected, values of parameters such as perimeter, length, width, convex hull perimeter, lobe count, and lobe width increased significantly ($P < 0.001$) with increasing cell size (Fig. 5C; Supplemental Figure S5, A and B). Consistent with increased lobe formation and growth, inversely related parameters such as solidity, circularity, or convex hull roundness decreased significantly ($P < 0.001$). Other parameters, in particular cellular eccentricity, were unaffected by cell size or by the developmental stage (Fig. 5E). Within the individual developmental stages, we observed highly significant differences between small and medium and small and large cells for almost all analyzed features. Between medium and large cells, significant differences were visible at 7 DAG ($P < 0.001$). Within the 5 DAG set, only a few features differed significantly ($P < 0.01$) between medium and large cells. The tendencies, however, were identical to those measured at 7 DAG.

To analyze geometries of similar-sized cells during leaf growth, we next compared the three cell populations (small, medium, and large) across the three developmental stages. Small cells at 3 DAG differed significantly from small cells at 7 DAG in 23 out of the 27 analyzed parameters ($P < 0.05$). Large differences were observed for PC-specific features such as lobe count and lobe length and for circularity and solidity (Supplemental Fig. S5, A and B). Similarly, medium-sized cells differed between the analyzed developmental stages. Only weak differences were observed
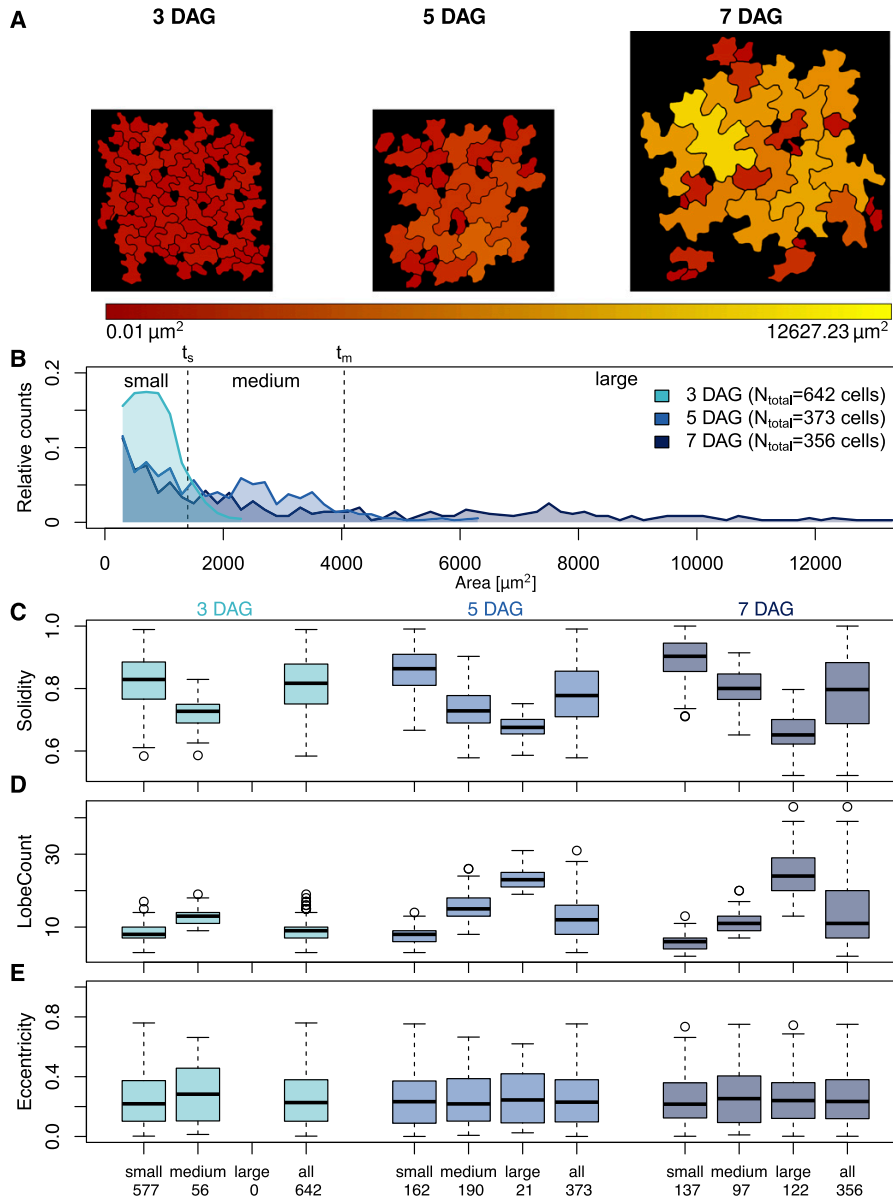
Figure 5. Analysis of pavement cell shape characteristics during development. A, Epidermal pavement cell shape in the adaxial side of cotyledons from wild-type (Col-0) seedlings 3, 5, and 7 DAG. The color gradient represents the area of the detected cells (red, small to yellow, large). B, Relative distribution of cell areas in cotyledons of 3-, 5-, and 7-d-old seedlings. Cells were categorized into small cells (threshold $t_s \leq 1400$ μm^2, which includes 90% of the cells in 3-d-old seedlings), medium-sized cells (threshold $t_m \leq 4042$ μm^2, which includes 90% of the cells in 5-d-old seedlings that exceed t_s), and large cells ($\geq t_m$), which represent the different stages of cell differentiation. C to E, Quantification of cell shape features during differentiation. Cells were grouped according to (B) or treated as a single input set (all). Numbers on the x axis refer to the number of cells analyzed per sample set. Feature values are shown in box plots. Results are medians; boxes range from first to third quartile. For a summary of all features and statistical analysis of feature values, see Supplemental Figure S5. (C) Solidity decreases with increasing cell size and differentiation, which is consistent with (D) an increased number of lobes, while other parameters, as shown for (E) eccentricity are largely unaffected during differentiation.

between the shape characteristics of large cells at 5 DAG and 7 DAG samples, and differences were restricted mostly to PC-specific shape features. These results are consistent with previous studies, which revealed age-dependent differences in cell shape of similar sized cells throughout development (Zhang et al., 2011; Wu et al., 2016). Differences in shape characteristics are more pronounced in expanding cell populations and disappear once cells are fully expanded. Thus, our data provide the first evidence that automatic analysis of large populations of PCs with PaCeQuant in combination with size filtering is a suitable approach to quantify developmental changes of PC shape.

Analysis of Lobe Type Characteristics

Most interdigitations are formed at two-cell contact points between two adjacent cells and form lobes referred to as type I lobes (Wu et al., 2016). In addition, some lobes form at three-cell contact points, referred to as type II lobes. So far, however, analyses of type I and type II lobe characteristics are limited because lobe classification has to be done manually, and methods for the quantification of lobe characteristics are missing. With its automatic segmentation and feature extraction capabilities, PaCeQuant provides a framework for the automatic classification of lobe types and shape quantification. We implemented an optional analysis of shape characteristics of individual lobes within single cells, which allows discriminating between lobes at two-cell contact points (TYPE-1) and three-cell contact points (TYPE-2; Fig. 6).

We classify lobes into TYPE-1 or TYPE-2 based on the analysis of their neighborhood in fields of adjacent cells (Fig. 6A). All lobes for which information on neighboring cells is missing, e.g. at border regions or adjacent to stomata, and which thus cannot be sorted in either of the two groups, are classified as UNDEFINED. For classification, we use the output label images generated during stage (IV) of the image analysis workflow implemented in PaCeQuant (see Supplemental Fig. S1, image 9). In these images, cell boundaries are represented as a black line with a width of 3 pixels that refers to the background of the image, and regions belonging to individual cells are defined by unique labels. In a first step, cell regions are expanded by dilation with a mask size of ≤5 using a built-in MiToBo function. The dilation locally stops as soon as two adjacent cell regions touch, which removes the boundaries between two adjacent cells and prevents the fusion of individual cell labels. Remaining black pixels thus exclusively refer to background regions, which do not belong to a cell boundary nor to any cell. Subsequently, for each individual lobe within each analyzed cell, the total number of different labels in its vicinity, defined by an 11 × 11 pixel-sized neighborhood around each pixel of the lobe, is counted and yields the base for type classification. If two different labels are present, i.e. the label of the analyzed lobe region and one additional neighboring

region, the lobe is classified as TYPE-1 (Fig. 6A). In case of three labels present at the vicinity of a single lobe, it belongs to the TYPE-2 group of lobes. The presence of one or more background pixels at the lobe indicates that it is located close to an undefined image region where no further information is available, and the lobe thus is classified as UNDEFINED.

We applied the lobe classification approach to the developmental series (see Fig. 5) and quantified the number of type I and type II lobes within the small-, medium-, and large-sized cells at 3 DAG, 5 DAG, and 7 DAG (Fig. 6B; Supplemental Figure S6A). We observed only small increases in the number of type II lobes per cell between small (average lobe number of 1 to 2)- and medium-sized cells (average lobe number of 2 to 3.5), and only weak differences between small, medium, and large cells (averaeg lobe number of 1 to 3.8), respectively, during development (Fig. 6B). In contrast, the number of type I lobes increased during cellular expansion from an average lobe number of 2 to an average lobe number of 10 in small- and large-sized cells, respectively. In medium- and large-sized cells, most lobes form at two-cell junctions (Fig. 6B). Thus, consistent with earlier studies, our data suggest that type II lobes are established mostly during cytokinesis, while most of the lobes formed during cellular expansion are type I lobes (Jura et al., 2006; Wu et al., 2016).

To quantify shape characteristics of individual lobes, PaCeQuant extracts five lobe features per lobe, including the equator length, the baseline length, the apical and basal lengths, and the contour length of each lobe (Supplemental Fig. S6B). In addition, for type II lobes, we quantify the distance of the contour segments from the equator to the three-cell-junction point. To estimate the bending of a lobe from these features, we calculated the ratio of the lobe equator length to the total length of the lobe contour (Supplemental Fig. S6, C–E). In a direct comparison, type II lobes displayed a larger contour length relative to the equator than type I lobes, as indicated by log(Equator Length/Total Lobe-Contour Length) values of ≤1 and ≥1 in type II and type I lobes, respectively (Fig. 6C). This points to a higher degree of asymmetric expansion at three-cell junctions. We measured the length of the lobe contour segments corresponding to the contact sides with the two neighboring cells and compared the ratio of the short versus the long contour segments (Fig. 6D; Supplemental Fig. S6, E and F). Our data revealed a tendency for a more similar length of the two lobe contours in small cells when compared to large cells, which suggests that different levels of expansion of the two neighboring cells contributed to these changes in type II lobe characteristics. During development, the difference between the two nonuniformly growing parts of the lobes increased further. Our results suggest that predominant expansion of one the two neighboring cells contributes to the higher degree of asymmetric expansion at type II lobes, which is consistent with previous reports (Jura et al., 2006). Together, we conclude that our novel method robustly detects and
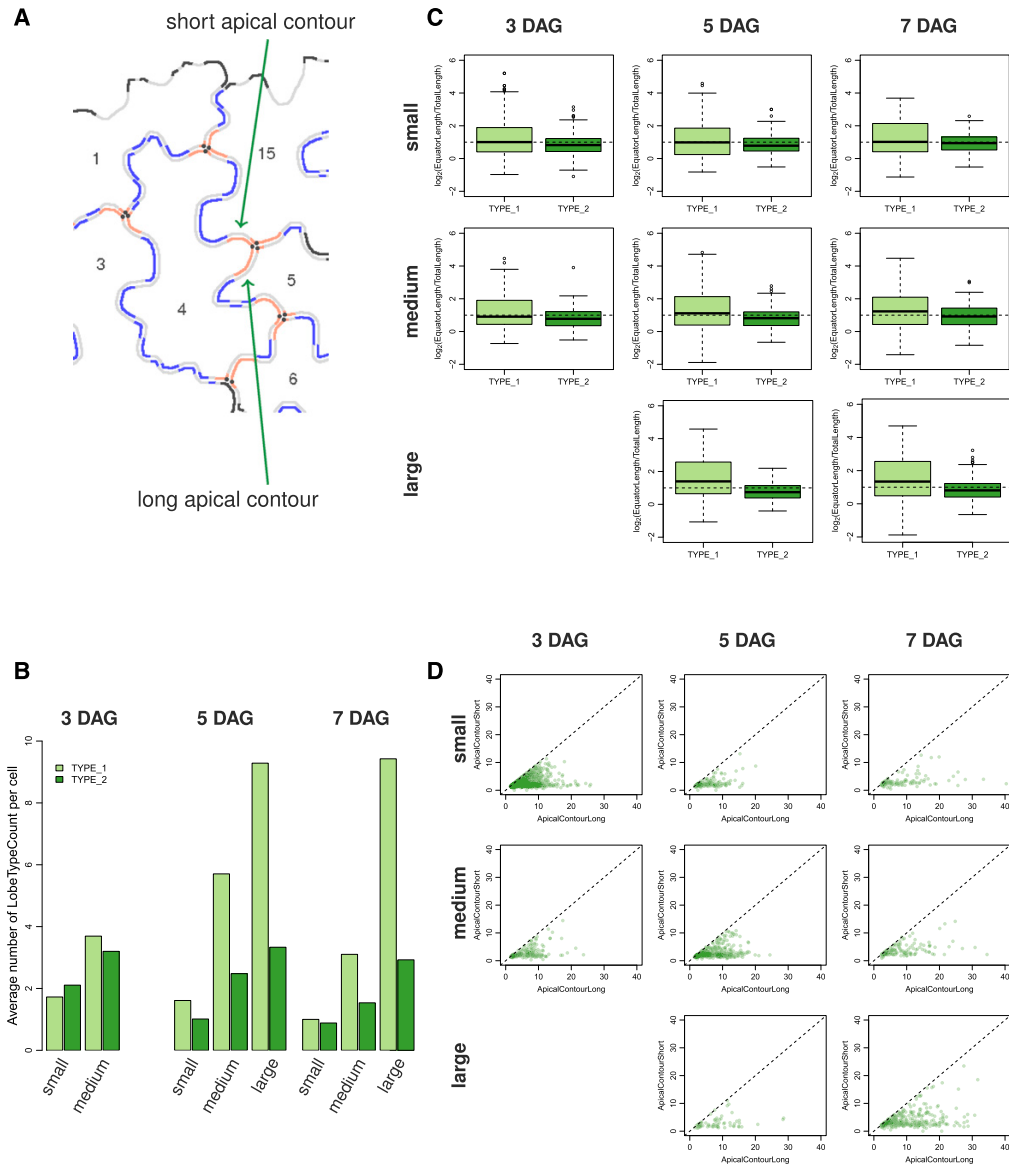
Figure 6. Analysis of type I and type II lobes and quantification of lobe characteristics. A, Image of an exemplary group of adjacent PCs after neighborhood analysis of individual lobes. Apical contours of type I and type II lobes are shown in blue and red, respectively. In type II lobes, three-cell contact points (shown as black dots) separate the lobe contours corresponding to the contact sides with the two neighboring cells (referred to as short and long contour segments). B to D, Analysis of lobe characteristics in PCs from the three developmental time points (3 DAG, 5 DAG, and 7 DAG) and the three size categories (small, medium, large; see Fig. 5). For an overview of quantified lobe features and statistical analysis, see Supplemental Figure S6. B, Bar plots showing the average number of type I (TYPE_1) and type II (TYPE_2) lobes per PC. C, Average ratio of lobe equator length to total contour length in type I and type II lobes in a logarithmic scale. D, Analysis of the length of the two parts of type II lobe contours that span the distances from the lobe equator to the three-cell contact point. Scatterplots of all individual type II lobes compare the short (ApicalContourShort) and long (ApicalContourLong) fragments.

95

classifies type I and type II cells based on the analysis of the lobe neighborhood and quantifies characteristics of individual lobes, which provides a platform to study effects of, e.g. gene functions specifically involved in the formation of type II lobes.

Phenotype Analysis of Cell Shape Mutants

We next applied the PaCeQuant program to analyze shape characteristics in mutants that differ in PC shape compared to wild-type seedlings (data set 4; Fig. 7). We selected a mutant impaired in KATANIN (KTN1) function. *KTN1* encodes a protein with microtubule severing function, and *ktn* mutant lines have extensively been studied with respect to microtubule ordering and PC shape characteristics (Lin et al., 2013; Lindeboom et al., 2013; Zhang et al., 2013). In particular, lobe growth has been shown to be impaired in *ktn1* mutants, which results in a reduced length of lobes (Lin et al., 2013). In addition, we included transgenic *Arabidopsis Pro-35S:IQ67-DOMAIN (IQD)16* lines in our analysis. IQD16 belongs to a novel class of calmodulin-binding proteins with potential roles in cellular calcium signaling (Abel et al., 2005; Abel et al., 2013). Most IQD family members localize to microtubules, and overexpression of *IQD16* alters microtubule organization and induces cell elongation (Bürstenbinder et al., 2017a; Bürstenbinder et al., 2017b). As input, we used 10 images per genotype of cotyledons from 5-d-old seedlings (Fig. 7A). We selected 5-d-old seedlings because at this stage cotyledons are flat, which facilitates fast and easy image acquisition. In addition, a large number of PCs is actively expanding at 5 DAG and has developed pronounced lobes (see Fig. 5), which is a prerequisite for reliable lobe quantification.

We observed comparable distributions of cell sizes in the two mutant lines when compared to the wild-type control (Fig. 7B). Since medium- and large-sized cells are highly similar in cotyledons of Arabidopsis wild-type seedlings at 5 DAG (see Fig. 5) we assumed that removal of only small cells from the sample set is sufficient for size filtering and reliable quantification (see Supplemental Fig. S5B). We thus applied the empirically determined small size threshold t_s from the developmental analysis of wild-type seedlings (see Fig. 5) to the mutant data set and quantified PC shape characteristics in all cells larger than 1,404 μm^2. After size filtering, we retained 161, 156, and 226 cells for wild-type, *ktn1-5*, and *Pro-35S:IQD16*, respectively (Fig. 7B). Thus, approximately 50% of all cells detected by PaCeQuant are included in the subsequent data analysis.

Consistent with previous reports, we detected a significantly reduced average lobe length in *ktn1-5* ($P <$ 0.001; Fig. 7C; Supplemental Fig. S7B). The lobe length automatically measured with our novel algorithm resembles lobe length values measured manually by Lin et al. (2013) and correlates with lobe length values measured in *ktn1* mutants. The total number of lobes is

not altered in *ktn1-5* when compared to wild-type seedlings, which points to roles of KTN1 in lobe growth, but not in lobe initiation. In addition, we observed increases in cell size (area), but not in cellular elongation (eccentricity). Nonlobe area as well as minimum and maximum core width are increased in *ktn1-5* mutants. Thus, our data suggest that loss of *KTN1* promotes isotropic expansion of PCs, possibly by preventing growth restriction at the neck regions. In transgenic *Pro-35:IQD16* lines, lobe length is reduced even stronger than it is in *ktn1-5* ($P \leq$ 0.001). The reduced lobe growth is further reflected by an increased nonlobe area ratio and by a reduced irregularity of the cell contour, as reflected by a decreased margin roughness in *Pro-35S:IQD16 < ktn1-5 <* Col-0 (Fig. 7C; Supplemental Fig. S7, A and B). Compared to wild-type and *ktn1-5* mutants, we observed a significant elongation of individual cells in *Pro-35:IQD16* seedlings ($P <$ 0.001), which is represented by increased length ($P <$ 0.001), reduced width ($P <$ 0.001), and increased eccentricity ($P <$ 0.001) values of cells from *Pro-35:IQD16* lines. The cell sizes, however, are not altered in *Pro-35S:IQD16* lines, which suggests that IQD16 does not promote cellular expansion but alters the direction of cellular growth. Thus, our data suggest distinct functions of KTN1 and IQD16 in regulation of cell growth and cellular expansion and provide first insights into potential roles of IQD16 in establishment of cellular polarity or growth anisotropy. Together, the combined analysis of 27 shape descriptors provides a platform for comparative and quantitative analyses with statistical support suitable for mutant phenotyping and developmental analyses.

DISCUSSION

In this study, we present PaCeQuant, a novel ImageJ-based tool for automatic segmentation of leaf epidermal PCs and simultaneous quantification of PC shape characteristics. The fully automatic segmentation of individual cells by PaCeQuant is a major advance because currently all measurements of PCs require manual segmentation. Manual segmentation is very time consuming and prone to bias introduced by the subjectivity of sample choice and contour labeling (Vanhaeren et al., 2015; Wu et al., 2016). PaCeQuant efficiently detects cell outlines in confocal input images using a combination of contrast and boundary enhancement, analysis of skeletons in binary images and watershed-based gap closing (Fig. 1).

We validated the accuracy of the automatic segmentation implemented in PaCeQuant by comparison to results from manually segmented cells (Fig. 3; Supplemental Fig. S2; Supplemental Table S2). In few cases, PaCeQuant locally determined cell contours with low accuracy, mostly at regions of lower contrast (Fig. 3; Supplemental Fig. S4). Local regions of deviation from manually segmented contours, however, only had weak effects on calculated shape properties of
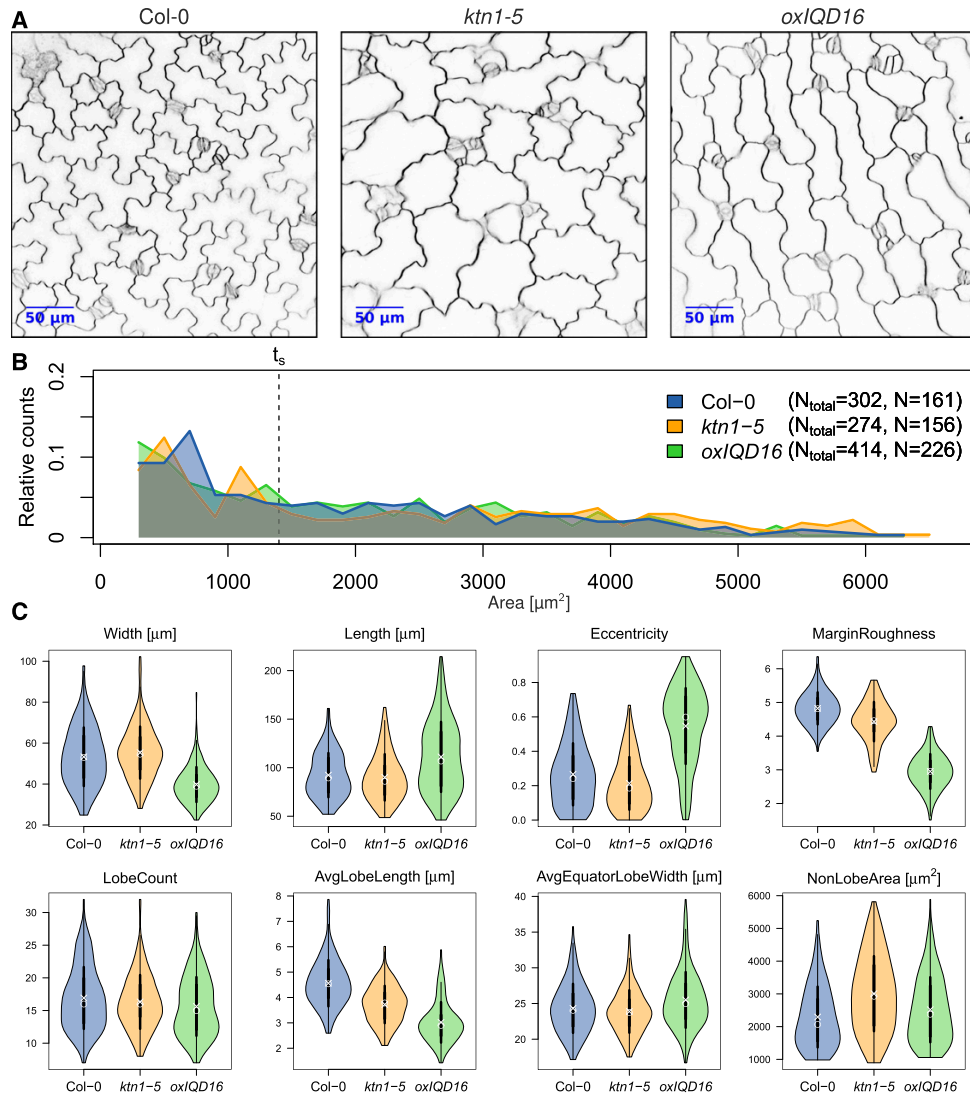
96

Figure 7. Phenotypic analysis of pavement cell shape mutants. PC shape analysis in cotyledons of 5-d-old seedlings from wild type (Col-0) and two mutants, *ktn1-5* and transgenic *Pro-35S:IQD16* (*oxIQD16*) plants. A, Inverted confocal images of wild type and the two mutants stained with FM4-64. B, Relative distribution of cell areas in the three data sets. Numbers in the legend refer to the total number of cells (N_{total}) from 13 images of the wild type, 13 of *ktn1-5*, and 17 of *oxIQD16*. Cells larger than size threshold $t_s = 1,400 \, \mu m^2$ (N) were used for further analysis (see Fig. 5). C, Violin plots of value distributions for four global (top) and four PC-specific features (bottom). Circles and crosses refer to medians and means; the vertical black lines in each category represent the sd (thick lines) and the 95% confidence intervals (thin lines). The width of each violin box represents the local distribution of feature values along the y axes. For a summary of all features and a statistical analysis, see Supplemental Figure S7.

individual cells (Fig. 3; Supplemental Fig. S2). When applied to a larger data set of 373 cells, PaCeQuant partially failed to segment the contours of less than 10 percent of the cells (Supplemental Fig. S4). Manual filtering did not affect the calculated feature values significantly (Supplemental Fig. S4), which suggests that the large number of analyzed cells generates data sets robust enough to compensate for minor detection errors. When compared to manual segmentation, PaCeQuant reduces the variance in segmentation results due to its high reproducibility, and bias induced by PaCeQuant is constant irrespective of when or by whom segmentations are performed. Thus, PaCeQuant provides a robust platform for cell segmentation, which can be used for shape quantification without manual postprocessing. To account for input data sets with

medium to low quality of input images and to increase the compatibility of PaCeQuant with independent imaging techniques such as agarose imprints or scanning electron microscopy images, we additionally implemented an optional semiautomatic or manual segmentation mode, respectively, into PaCeQuant.

The automatic segmentation of individual PCs implemented in PaCeQuant provides the basis for high-throughput shape quantification. To study developmental shape transitions and molecular and genetic modes of shape establishment, imaging must be combined with the quantification of PC shape characteristics. A combination of advanced imaging, automatic quantification of shape features, and computational modeling has been used to study the development of embryos (De Rybel et al., 2014; Yoshida et al., 2014), of root meristems (Campilho et al., 2006), of shoots apical meristems (Reddy et al., 2004; Kierzkowski et al., 2012; Serrano-Mislata et al., 2015), and of flower organs (Tauriello et al., 2015). In these studies, the direction of cell division and cell expansion is tracked over time from semi- and fully automatic 3D segmentations of individual cells using the MorphoGraphX software (de Reuille et al., 2014; Barbier de Reuille et al., 2015). In combination with molecular and developmental data imaging-based analyses enabled the generation of comprehensive models of the underlying regulatory processes (Roeder et al., 2011; Sozzani et al., 2014). All cell types analyzed by these approaches, however, are simple shaped cubic or cylindrical cells, whereas PCs form much more complex and diverse cell shapes during expansion. The algorithms used to track and measure growth in these model systems cannot directly be applied to quantify PC characteristics, as they oversimplify PC shape geometry, and comparable quantification methods suitable for robust analysis of PCs are lacking. In most existing studies that analyzed PC characteristics, differences in PC shape between mutants and wild type were quantified by the analysis of single shape parameters only (Tisné et al., 2008; Xu et al., 2010; Li et al., 2013) or by a combination of a few select shape parameters (Fu et al., 2005; Guo et al., 2015). Although sufficient for direct comparison of phenotypes, the limited shape information reduces the informational value of the individual measurements and hampers a comprehensive comparison of phenotypes identified in different studies and from different laboratories (Ivakov and Persson, 2013; Wu et al., 2016).

To provide a standardized platform for PC shape phenotyping, we included the simultaneous quantification of 27 different shape parameters in PaCeQuant. The 27 features include global geometric and morphological features, such as area, perimeter, and circularity (Fig. 2; Supplemental Table S1). In addition, we quantify the number, length, and width of lobes, which are specific characteristics of PCs, by analysis of the local curvature along the cell contour. In a direct comparison with the recently released LobeFinder tool (Wu et al., 2016), PaCeQuant quantifies more shape features than LobeFinder (27 versus 8, respectively), including the

length of lobes and the degree of cellular elongation. PaCeQuant detects in average one to three lobes more per cell than LobeFinder (Fig. 4). The lobe numbers detected by PaCeQuant are in the range of lobes detected by individuals (Fig. 4; Supplemental Fig. S3). The large differences in the lobe numbers measured by four independent individuals further demonstrate the high variability of manual image analysis and highlight the requirement for automated phenotyping platforms. Another important parameter in PC shape development is the degree of growth restriction at neck regions, sometimes estimated by the quantification of the neck width (Li et al., 2003; Bannigan and Baskin, 2005; Fu et al., 2005). The definition of the neck width, however, is not clear and renders objective measurements difficult. We thus developed novel objective and well-defined algorithms to quantify the extent of the cellular core region. These two features, the minimal and maximal core width (Fig. 2; Supplemental Table S1), quantify the widths of narrow and wide parts of the core region of PCs, defined by distance of the cell contour to the central skeleton branch. Small values represent a small neck width, while large values correspond to a larger neck width. In combination with the analysis of the nonlobe area and the ratio of nonlobe area to total area, these values provide information on the growth restriction at neck regions.

Finally, PaCeQuant is the first tool that not only analyzes and quantifies shape features per cell but also extracts data per lobe. It provides length and width measurements of various parts of an individual lobe, e.g. of the overall lobe length as well as of baseline and equator lengths. PaCeQuant additionally includes an automatic identification of type I and type II lobes at two-cell and three-cell contact points, respectively, for cells with sufficient neighborhood information (Fig. 6; Supplemental Fig. S6). This allows detailed analyses of the characteristics in particular of three-cell contacts, e.g. at different time points during development, with regard to lobe symmetry. By applying PaCeQuant on a developmental time series of Arabidopsis cotyledon PCs, we demonstrate that PaCeQuant accurately quantifies shape features over a wide range of cellular sizes and differing shapes (Fig. 5). In our sample sets, PaCeQuant automatically detected between 21 and 34 cells per image and simultaneously quantified the 27 shape features with an average processing speed of 1.6 s per cell. The supplemental R script enables fast and easy data visualization, statistical analysis and data processing, and supports high-throughput data analysis. Thus, PaCeQuant is suitable to generate large data sets in a short period of time, which are free from biased selection of individual cells. The data sets contain information on the number and distribution of cells from different stages of cellular expansion. To quantify and compare shape geometries in defined subsets of cells during development we applied size filters and analyzed shape features in small-, medium-, and large-sized cells. Our data reveal age-dependent differences in cell shape between populations of similar sized cells,

which likely reflect different stages of differentiation (Fig. 5). The results are in agreement with previous studies, which reported changes of shape in actively expanding PCs during development (Zhang et al., 2011; Elsner et al., 2012).

Analysis of lobe types and feature characteristics of individual lobes indicated that, in a direct comparison, type II lobes are less regular shaped (Fig. 6). These differences might be caused by higher growth rates of one of the two neighboring cells, which is consistent with earlier reports (Jura et al., 2006). The optional analysis of lobe types implemented in PaCeQuant thus provides the first tool to study effects of gene functions specifically involved in the formation of three-cell junctions and to measure shape characteristics of individual lobes. For the phenotypic analysis of mutants that, e.g. affect PC shape globally it, however, is not necessary to distinguish between these two lobe types. If desired, the user thus can optionally activate lobe type classification and analysis in the graphical user interface.

We additionally validated the usability of PaCeQuant for PC shape quantification in mutant phenotyping as a first proof of concept (Fig. 7). Consistent with previous reports, we observed a reduced length of lobes in *ktn1-5* lines (Lin et al., 2013), which confirms that the analysis of contour curvature implemented in PaCeQuant is suitable for measurements of lobe characteristics. We provide evidence that the combined analysis of 27 shape features with PaCeQuant generates information on multiple aspects of PC shape and growth, including the size of the cellular core region and the number and characteristics of lobes. The analyzed features reflect the geometrical complexity of PCs. Thus, PaCeQuant enables comparative analyses between mutants or during development even across different laboratories with a more complex description of PC shapes than other existing tools.

We will establish a future database where PaCeQuant results can be uploaded and directly compared to existing data sets. Such a database will enable the identification of, e.g. mutants with highly similar defects in shape formation, which might point to similar cellular functions and could assist the molecular and physiological characterization. In the long term, we aim to implement the collected data in systems biology approaches to model the underlying molecular and genetic pathways.

In conclusion, our data demonstrate that PaCeQuant is suitable for automatic segmentation of PCs, which builds a prerequisite for the generation of objective and reproducible data. Together with the implemented feature analysis, PaCeQuant offers the possibility of high-throughput PC shape analysis. It thus provides a user-friendly platform for large-scale shape quantification with potential application in studies of cell shape changes in response to external stimuli, during development or in mutant phenotyping.

MATERIALS AND METHODS

Plant Material and Growth Conditions

Arabidopsis (*Arabidopsis thaliana*; ecotype Col-0) seeds were originally obtained from the Arabidopsis Biological Resource Center. Seeds of the *ktn1-5* T-DNA line (SAIL_343_D12) were obtained from the Nottingham Arabidopsis Stock Centre (NASC). Transgenic *Pro-35S:IQD16* lines were generated by *Agrobacterium tumefaciens*-mediated transformation using the floral-dip method (Clough and Bent, 1998), as described in Bürstenbinder et al. (2017b). Arabidopsis seeds were surface-sterilized with chlorine gas, and after 2 d of stratification at 4°C grown vertically on square plates containing Arabidopsis salt (ATS) medium and 0.5% (w/v) agar (Lincoln et al., 1990) at 20°C with cycles of 16 h light and 8 h dark.

Confocal Microscopy

For analysis of epidermal pavement cell shape, seedlings were incubated 10 to 30 min in 50 μM FM4-64 dye (Synapto-Red, Sigma). Imaging was performed with an LSM 700 inverted laser scanning microscope (Carl Zeiss) using a 20× plan neofluar objective. The laser line 555 nm was used for FM4-64 excitation, and fluorescence emission was detected between 560 and 590 nm. Single optical sections of FM4-64-stained cotyledons were acquired with pixel dwell times between 0.8 and 1.3 μs and an averaging of 4 to increase the signal-to-noise ratio. For each developmental stage and line, 10 to 17 cotyledons from 5 to 10 individual seedlings were scanned. Groups of adjacent pavement cells were imaged with a resolution of 3.19 pixels/μm in 3-d-old and 5-d-old seedlings and with a resolution of 2.24 pixels/μm in 7-d-old seedlings.

Data Sets

For the evaluation of PaCeQuant, we used four data sets, which were comprised as follows:

Data set 1: 15 individual cells from one input image of 5-d-old Arabidopsis Col-0 seedlings.

Data set 2: 373 cells identified from 14 input images of cotyledons from 5-d-old Arabidopsis Col-0 seedlings. After manual postprocessing, 342 cells were retained in the curated data set.

Data set 3: 642 cells from 10 input images of Arabidopsis Col-0 seedlings 3 DAG, 373 cells from 14 input images of seedlings 5 DAG, and 356 cells from 13 images of seedlings 7 DAG.

Data set 4: 302 cells from 13 input images of Col-0 wild type, 274 cells from 13 input images of *ktn1-5* mutants, and 414 cells from 17 input images of transgenic *Pro-35S:IQD16* lines from 5-d-old seedlings. After size filtering for removal of small cells ($\leq$1400 μm^2), the data set contains 161, 156, and 226 cells from Col-0, *ktn1-5*, and *Pro-35S:IQD16*, respectively.

Installation and Usage of PaCeQuant

PaCeQuant is part of the Microscope Image Analysis ToolBox MiToBo. The PaCeQuant plugin can be added to ImageJ/Fiji by activation of MiToBo's update site via the update site manager of Fiji. Alternatively, binary packages of MiToBo, including the PaCeQuant plugin for direct use with ImageJ/Fiji, can be downloaded from MiToBo's website. A detailed description of PaCeQuant, including installation instructions can be found at PaCeQuant's website (http://mitobo.informatik.uni-halle.de/index.php/Applications/PaCeQuant).

After installation, PaCeQuant can be found in ImageJ's plugin menu. After selection of Plugins > MiToBo > PaCeQuant, a graphical user interface opens, which allows easy configuration and execution of the tool (Supplemental Fig. S8).

First, the general workflow and data input format have to be configured. For this purpose, PaCeQuant offers three options:

SEGMENTATION_AND_FEATURES

In this mode, PaCeQuant uses confocal input images for automatic cell segmentation and feature quantification. As output, images showing the segmentation results, ImageJ ROI files containing the region data and for each image

a text file including all extracted feature values per analyzed cell are exported and saved in the result directory.

SEGMENTATION_ONLY

By choosing this mode, PaCeQuant performs automatic segmentation of individual cells on confocal input images. As output, images and ImageJ ROI files are exported and saved in the result directory, which subsequently can be reimported to ImageJ for manual correction of segmentation inaccuracies.

FEATURES_ONLY

In this mode, either manually corrected segmentation data or input ROIs generated by, e.g. manual segmentation of individual cells, can be imported, and for each image or ROI set a text file including all extracted feature values per analyzed cell are exported and saved in the result directory.

For feature analysis, PaCeQuant accepts the following input formats:

BINARY_IMAGE: a binary image where the background is marked by intensity value 0 and the foreground by intensity value 255.
LABEL_IMAGE: a grayscale image where each cell region is marked by a unique label.
IMAGEJ_ROIs: ImageJ ROI files with region data, e.g. exported from ImageJ's ROI manager

PaCeQuant supports two basic working modes:

INTERACTIVE

Only the image or ROI set currently selected in ImageJ/Fiji is analyzed.

BATCH

All images or region files in a selected directory, and all its subdirectories are analyzed.

Depending on the chosen option for image analysis and the selected working mode, the configuration of the GUI dynamically changes, and mode-specific entry fields appear. For segmentation, the user has to specify how the cells in the images are labeled, i.e. if the boundaries are darker than the background or vice versa. Note that these are the only settings that are required to run PaCeQuant. Optionally, PaCeQuant offers to activate a gap-closing heuristic to account for incomplete cell boundaries due to low contrast and to define thresholds on the size of valid regions, which we recommend to set properly. For expert users, further configuration options are available via the parameter "Morphology Analyzer Operator." It allows for changing thresholds and parameter values applied in lobe detection and enables deactivation of subsets of features. Users, however, should be aware that any change in these parameters might hamper comparative evaluations between experiments and thus is not recommended. Finally, the analysis of lobe types (TYPE-1, TYPE-2, UNDEFINED) and the quantification of shape features of individual lobes can be activated optionally. These results will be exported to a separate table per cell and can be analyzed separately. At the bottom of the configuration, window options for generating supplemental results are provided, e.g. for an image stack including additional intermediate result images and for a stack with additional images visualizing the feature values of each cell by mapping them to the intensity values of the cell in the corresponding image. Lastly, additional verbose output to console can be activated.

Clicking the button "Run" starts the processing. If a single image is analyzed, the results will directly pop up, while in batch mode, the results are written to file. For each input image in each processed folder, several files with segmentation data and an output text file are generated and saved to a new subfolder "results." The text files contain row-wise for each cell the set of feature values extracted by PaCeQuant. If the optional lobe type classification is activated, additional text files are generated for each individual cell per image. These files contain information on the lobe type and on lobe-specific feature values for each lobe of the particular cell. Note that a complete documentation of all configuration options and all general information about PaCeQuant can be found at MiToBo's Web site in the documentation section:
http://mitobo.informatik.uni-halle.de/index.php/Applications/PaCeQuant.

Automatic Cell Shape Analysis and Data Visualization with R

For analysis of the data output, we provide an R script that visualizes the features and performs basic statistical evaluations. As data input the path to the

results directory of the PaCeQuant output has to be provided by the user. Each subdirectory within the results directory defines a sample set. First, all detected cells are sorted according to size, and a histogram of the size distribution is generated for each individual sample set. In the next step, size filtering can be applied to exclude small cells below a manually defined size threshold. Values between 0 and 1 are considered as percentages of the total number of detected cells. Values larger than 1 are considered as absolute PC areas in μm^2. The size filtering can be performed on the global level, where the same threshold is applied to all analyzed sample sets. Alternatively, local thresholds can be set separately for individual data sets. Since size reflects the developmental stage of PCs and thus has a large impact on most analyzed features, we recommend comparing the size distributions of the analyzed sample sets before selecting a threshold for further analysis. We recommend using global thresholds if all samples have comparable size distributions. For comparison of, e.g. mutants with generally reduced cell sizes compared to wild-type seedlings, locally defined thresholds might be useful to compare cells at similar stages of cell differentiation. Lastly, we suggest using absolute sizes as threshold, which, however, requires careful analysis of size distributions and of size impacts prior to threshold selection. For comparative analyses in Arabidopsis (also across different laboratories), we recommend to use developmental stages and thresholds introduced in this work ($t_s = 1400$ μm^2). To visualize the data, the results of all individual features are plotted as histograms, boxplots, or violin plots.

Statistical Analysis

Analysis of the data sets with the Kolmogorov-Smirnov test revealed nonnormal distribution for most sample sets and features. For statistical analysis, we thus applied the nonparametric Kruskal-Wallis test provided by the R package stats, which is the equivalent of the parametric one-way ANOVA (Kruskal and Wallis, 1952). The Kruskal-Wallis test is intended for testing whether samples originate from the same distribution. Specifically, it tests for the null hypothesis that the location parameters (medians) of the distributions of the given samples are the same in each group. The alternative is that they differ in at least one group. In the latter case, we applied the Dunn's test (Dunn, 1964) provided by the R package dunn.test as post hoc test to analyze each pair of samples separately.

Accession Numbers

Sequence data for *KTN1* and *IQD16* can be found in the EMBL/GenBank data libraries under accession numbers At1g80350 and At4g10640, respectively.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Effects of the individual processing steps implemented in PaCeQuant on the input image.

Supplemental Figure S2. Comparison of PaCeQuant features between manually and automatically segmented cells.

Supplemental Figure S3. Pairwise comparison of features computed by PaCeQuant and LobeFinder.

Supplemental Figure S4. Comparison of feature values between Raw PaCeQuant output data and after cell removal by manual filtering.

Supplemental Figure S5. Quantification and statistical analysis of pavement cell shape features during development.

Supplemental Figure S6. Classification and quantification of type I and type II lobe characteristics.

Supplemental Figure S7. Quantification of pavement cell shape features in wild-type and mutant lines.

Supplemental Figure S8. Graphical user interface of PaCeQuant.

Supplemental Table S1. Detailed definitions of features analyzed by PaCeQuant.

Supplemental Table S2. Quantitative evaluation of segmentation quality.

LITERATURE CITED

Abel S, Bürstenbinder K, Müller J (2013) The emerging function of IQD proteins as scaffolds in cellular signaling and trafficking. Plant Signal Behav **8:** e24369

Abel S, Savchenko T, Levy M (2005) Genome-wide comparative analysis of the *IQD* gene families in *Arabidopsis thaliana* and *Oryza sativa.* BMC Evol Biol **5:** 72

Andriankaja M, Dhondt S, De Bodt S, Vanhaeren H, Coppens F, De Milde L, Mühlenbock P, Skirycz A, Gonzalez N, Beemster GTS, et al (2012) Exit from proliferation during leaf development in *Arabidopsis thaliana*: a not-so-gradual process. Dev Cell **22:** 64–78

Bai Y, Falk S, Schnittger A, Jakoby MJ, Hülskamp M (2010) Tissue layer specific regulation of leaf length and width in Arabidopsis as revealed by the cell autonomous action of ANGUSTIFOLIA. Plant J **61:** 191–199

Bannigan A, Baskin TI (2005) Directional cell expansion—turning toward actin. Curr Opin Plant Biol **8:** 619–624

Bar M, Ori N (2014) Leaf development and morphogenesis. Development **141:** 4219–4230

Bar M, Ori N (2015) Compound leaf development in model plant species. Curr Opin Plant Biol **23:** 61–69

Barbier de Reuille P, Routier-Kierzkowska AL, Kierzkowski D, Bassel GW, 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:** 05864

Biot E, Cortizo M, Burguet J, Kiss A, Oughou M, Maugarny-Calès A, Gonçalves B, Adroher B, Andrey P, Boudaoud A, et al (2016) Multi-scale quantification of morphodynamics: MorphoLeaf software for 2D shape analysis. Development **143:** 3417–3428

Bürstenbinder K, Mitra D, Quegwer J (2017a) Functions of IQD proteins as hubs in cellular calcium and auxin signaling: A toolbox for shape formation and tissue-specification in plants? Plant Signal Behav **12:** e1331198

Bürstenbinder K, Möller B, Plötner R, Stamm G, Hause G, Mitra D, Abel S (2017b) The IQD family of calmodulin-binding proteins links calcium signaling to microtubules, membrane subdomains, and the nucleus. Plant Physiol **173:** 1692–1708

Campilho A, Garcia B, Toorn HV, Wijk HV, Campilho A, Scheres B (2006) Time-lapse analysis of stem-cell divisions in the *Arabidopsis thaliana* root meristem. Plant J **48:** 619–627

Chaudhuri S, Chatterjee S, Katz N, Nelson M, Goldbaum M (1989) Detection of blood vessels in retinal images using two-dimensional matched filters. IEEE Trans Med Imaging **8:** 263–269

Cho KH, Jun SE, Lee YK, Jeong SJ, Kim GT (2007) Developmental processes of leaf morphogenesis in Arabidopsis. J Plant Biol **50:** 282–290

Clough SJ, Bent AF (1998) Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana.* Plant J **16:** 735–743

Dale MB, Groves RH, Hull VJ, O'Callaghan JF (1971) A new method for describing leaf shape. New Phytol **70:** 437–442

de Casas RR, Vargas P, Perez-Corona E, Manrique E, Garcia-Verdugo C, Balaguer L (2011) Sun and shade leaves of *Olea europaea* respond differently to plant size, light availability and genetic variation. Funct Ecol **25:** 802–812

de Reuille PB, Robinson S, Smith RS (2014) Quantifying cell shape and gene expression in the shoot apical meristem using MorphoGraphX. Methods Mol Biol **1080:** 121–134

De Rybel B, Adibi M, Breda AS, Wendrich JR, Smit ME, Novák O, Yamaguchi N, Yoshida S, Van Isterdael G, Palovaara J, et al (2014) Plant development. Integration of growth and patterning during vascular tissue formation in Arabidopsis. Science **345:** 1255215

Dunn OJ (1964) Multiple comparisons using rank sums. Technometrics **6:** 241–252

Elsner J, Michalski M, Kwiatkowska D (2012) Spatiotemporal variation of leaf epidermal cell growth: a quantitative analysis of *Arabidopsis thaliana* wild-type and triple cyclinD3 mutant plants. Ann Bot (Lond) **109:** 897–910

Fraz MM, Remagnino P, Hoppe A, Uyyanonvara B, Rudnicka AR, Owen CG, Barman SA (2012) Blood vessel segmentation methodologies in retinal images—a survey. Comput Methods Programs Biomed **108:** 407–433

Freeling M (1992) A conceptual framework for maize leaf development. Dev Biol **153:** 44–58

Freeman H, Davis LS (1977) Corner-finding algorithm for chain-coded curves. IEEE Trans Comput **26:** 297–303

Fu Y, Gu Y, Zheng Z, Wasteneys G, Yang Z (2005) Arabidopsis interdigitating cell growth requires two antagonistic pathways with opposing action on cell morphogenesis. Cell **120:** 687–700

Gao Y, Zhang Y, Zhang D, Dai X, Estelle M, Zhao Y (2015) Auxin binding protein 1 (ABP1) is not required for either auxin signaling or Arabidopsis development. Proc Natl Acad Sci USA **112:** 2275–2280

Ghent AW (1973) Gravity and the distribution of leaf shape in the trees of *Sassafras albidum.* New Phytol **72:** 1141–1158

Glover BJ (2000) Differentiation in plant epidermal cells. J Exp Bot **51:** 497–505

Gonzales RC, Woods RE (1996) Digital Image Processing. Addison-Wesley, Amsterdam, pp 491

Gray RA (1957) Alteration of leaf size and shape and other changes caused by gibberellins in plants. Am J Bot **44:** 674–682

Guo X, Qin Q, Yan J, Niu Y, Huang B, Guan L, Li Y, Ren D, Li J, Hou S (2015) TYPE-ONE PROTEIN PHOSPHATASE4 regulates pavement cell interdigitation by modulating PIN-FORMED1 polarity and trafficking in Arabidopsis. Plant Physiol **167:** 1058–1075

Horiguchi G, Fujikura U, Ferjani A, Ishikawa N, Tsukaya H (2006) Large-scale histological analysis of leaf mutants using two simple leaf observation methods: identification of novel genetic pathways governing the size and shape of leaves. Plant J **48:** 638–644

Ivakov A, Persson S (2013) Plant cell shape: Modulators and measurements. Front Plant Sci **4:** 439

Iwata H, Ukai Y (2002) SHAPE: A computer program package for quantitative evaluation of biological shapes based on elliptic Fourier descriptors. J Hered **93:** 384–385

Jura J, Kojs P, Iqbal M, Szymanowska-Pulka J, Wloch W (2006) Apical intrusive growth of cambial fusiform initials along the tangential walls of adjacent fusiform initials: Evidence for a new concept. Aust J Bot **54:** 493–504

Kalve S, De Vos D, Beemster GT (2014) Leaf development: A cellular perspective. Front Plant Sci **5:** 362

Kalyoncu C, Toygar Ö (2015) Geometric leaf classification. Comput Vis Image Underst **133:** 102–109

Kierzkowski D, Nakayama N, Routier-Kierzkowska AL, Weber A, Bayer E, Schorderet M, Reinhardt D, Kuhlemeier C, Smith RS (2012) Elastic domains regulate growth and organogenesis in the plant shoot apical meristem. Science **335:** 1096–1099

Kruskal WH, Wallis A (1952) Use of ranks in one-criterion variance analysis. J Am Stat Assoc **47:** 583–621

Kutschera U (2008) The growing outer epidermal wall: design and physiological role of a composite structure. Ann Bot (Lond) **101:** 615–621

Leymarie F, Levine MD (1992) Fast raster scan distance propagation on the discrete rectangular lattice. CVGIP Image Understanding **55:** 84–94

Li S, Blanchoin L, Yang Z, Lord EM (2003) The putative Arabidopsis arp2/3 complex controls leaf cell morphogenesis. Plant Physiol **132:** 2034–2044

Li H, Xu T, Lin D, Wen M, Xie M, Duclercq J, Bielach A, Kim J, Reddy GV, Zuo J, et al (2013) Cytokinin signaling regulates pavement cell morphogenesis in Arabidopsis. Cell Res **23:** 290–299

Lin D, Cao L, Zhou Z, Zhu L, Ehrhardt D, Yang Z, Fu Y (2013) Rho GTPase signaling activates microtubule severing to promote microtubule ordering in Arabidopsis. Curr Biol **23:** 290–297

Lin D, Nagawa S, Chen J, Cao L, Chen X, Xu T, Li H, Dhonukshe P, Yamamuro C, Friml J, et al (2012) A ROP GTPase-dependent auxin signaling pathway regulates the subcellular distribution of PIN2 in Arabidopsis roots. Curr Biol **22:** 1319–1325

Lincoln C, Britton JH, Estelle M (1990) Growth and development of the *axr1* mutants of Arabidopsis. Plant Cell **2:** 1071–1080

Lindeboom JJ, Nakamura M, Hibbel A, Shundyak K, Gutierrez R, Ketelaar T, Emons AM, Mulder BM, Kirik V, Ehrhardt DW (2013) A mechanism for reorientation of cortical microtubule arrays driven by microtubule severing. Science **342:** 1245533

Marchand-Maillet S, Sharaiha YM (1999) Binary Digital Image Processing: A Discrete Approach. Academic Press, London

Marcotrigiano M (2010) A role for leaf epidermis in the control of leaf size and the rate and extent of mesophyll cell division. Am J Bot **97:** 224–233

McLellan T, Endler JA (1998) The relative success of some methods for measuring and describing the shape of complex objects. Syst Biol **47:** 264–281

Möller B, Glaß M, Misiak D, Posch S (2016) MiToBo - A toolbox for image processing and analysis. J Open Res Softw **4:** e17

Möller B, Posch S (2013) A framework unifying the development of image analysis algorithms and associated user interfaces. Proc. of 13th Int. Conf. on Machine Vision Appl 447–450

Niblack W (1986) An Introduction to Digital Image Processing. Prentice Hall, Englewood Cliffs, NJ

Nicotra AB, Atkin OK, Bonser SP, Davidson AM, Finnegan EJ, Mathesius U, Poot P, Purugganan MD, Richards CL, Valladares F, et al (2010) Plant phenotypic plasticity in a changing climate. Trends Plant Sci **15:** 684–692

Posch S, Möller B (2017) Alida - Advanced library for integrated development of data analysis applications. J Open Res Softw **5:** 7

Reddy GV, Heisler MG, Ehrhardt DW, Meyerowitz EM (2004) Real-time lineage analysis reveals oriented cell divisions associated with morphogenesis at the shoot apex of *Arabidopsis thaliana*. Development **131:** 4225–4237

Roeder AHK, Tarr PT, Tobin C, Zhang X, Chickarmane V, Cunha A, Meyerowitz EM (2011) Computational morphodynamics of plants: Integrating development over space and time. Nat Rev Mol Cell Biol **12:** 265–273

Savaldi-Goldstein S, Peto C, Chory J (2007) The epidermis both drives and restricts plant shoot growth. Nature **446:** 199–202

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al (2012) Fiji: An open-source platform for biological-image analysis. Nat Methods **9:** 676–682

Serrano-Mislata A, Schiessl K, Sablowski R (2015) Active control of cell size generates spatial detail during plant organogenesis. Curr Biol **25:** 2991–2996

Sofka M, Stewart CV (2006) Retinal vessel centerline extraction using multiscale matched filters, confidence and edge measures. IEEE Trans Med Imaging **25:** 1531–1546

Sozzani R, Busch W, Spalding EP, Benfey PN (2014) Advanced imaging techniques for the study of plant growth and development. Trends Plant Sci **19:** 304–310

Staff L, Hurd P, Reale L, Seoighe C, Rockwood A, Gehring C (2012) The hidden geometries of the *Arabidopsis thaliana* epidermis. PLoS One **7:** e43546

Sultan SE (1995) Phenotypic plasticity and plant adaptation. Acta Bot Neerl **44:** 363–383

Sultan SE (2000) Phenotypic plasticity for plant development, function and life history. Trends Plant Sci **5:** 537–542

Tauriello G, Meyer HM, Smith RS, Koumoutsakos P, Roeder AHK (2015) Variability and constancy in cellular growth of Arabidopsis sepals. Plant Physiol **169:** 2342–2358

Tisné S, Reymond M, Vile D, Fabre J, Dauzat M, Koornneef M, Granier C (2008) Combined genetic and modeling approaches reveal that epidermal cell area and number in leaves are controlled by leaf and plant developmental processes in Arabidopsis. Plant Physiol **148:** 1117–1127

Tsukaya H (2010) Leaf development and evolution. J Plant Res **123:** 3–6

Tsukaya H (2013) Leaf development. Arabidopsis Book **11:** e0163

Tsukaya H (2014) Comparative leaf development in angiosperms. Curr Opin Plant Biol **17:** 103–109

Tsukaya H, Tsuge T, Uchimiya H (1994) The cotyledon: A superior system for studies of leaf development. Planta **195:** 309–312

Utcke S (2003) Error-bounds on curvature estimation. *In* LD Griffin, M Lillholm, eds, Scale Space Methods in Computer Vision. Scale-Space-2003. Lecture Notes in Computer Science, **Vol 2695**. pp 657–666

Vanhaeren H, Gonzalez N, Inzé D (2015) A journey through a leaf: Phenomics analysis of leaf growth in *Arabidopsis thaliana*. Arabidopsis Book **13:** e0181

Viscosi V, Cardini A (2011) Leaf morphology, taxonomy and geometric morphometrics: A simplified protocol for beginners. PLoS One **6:** e25630

Wang H, Chen J, Wen J, Tadege M, Li G, Liu Y, Mysore KS, Ratet P, Chen R (2008) Control of compound leaf development by FLORICAULA/LEAFY ortholog SINGLE LEAFLET1 in Medicago truncatula. Plant Physiol **146:** 1759–1772

Wolters H, Jürgens G (2009) Survival of the flexible: Hormonal growth control and adaptation in plant development. Nat Rev Genet **10:** 305–317

Wu TC, Belteton SA, Pack J, Szymanski DB, Umulis DM (2016) Lobe-Finder: A convex hull-based method for quantitative boundary analyses of lobed plant cells. Plant Physiol **171:** 2331–2342

Xiong Y, Kabacoff C, Franca-Koh J, Devreotes PN, Robinson DN, Iglesias PA (2010) Automated characterization of cell shape changes during amoeboid motility by skeletonization. BMC Syst Biol **4:** 33

Xu T, Wen M, Nagawa S, Fu Y, Chen JG, Wu MJ, Perrot-Rechenmann C, Friml J, Jones AM, Yang Z (2010) Cell surface- and rho GTPase-based auxin signaling controls cellular interdigitation in Arabidopsis. Cell **143:** 99–110

Yoshida S, Barbier de Reuille P, Lane B, Bassel GW, Prusinkiewicz P, Smith RS, Weijers D (2014) Genetic control of plant development by overriding a geometric division rule. Dev Cell **29:** 75–87

Zhang Q, Fishel E, Bertroche T, Dixit R (2013) Microtubule severing at crossover sites by katanin generates ordered cortical microtubule arrays in Arabidopsis. Curr Biol **23:** 2191–2195

Zhang C, Halsey LE, Szymanski DB (2011) The development and geometry of shape change in *Arabidopsis thaliana* cotyledon pavement cells. BMC Plant Biol **11:** 27

Zhang TY, Suen CY (1984) A fast parallel algorithm for thinning digital patterns. Commun ACM **27:** 236–239

Zhang B, Zhang L, Zhang L, Karray F (2010) Retinal vessel extraction by matched filter with first-order derivative of Gaussian. Comput Biol Med **40:** 438–445

102

3.2 Morphological Analysis of Leaf Epidermis Pavement Cells with PaCeQuant

Original Publication:

Abstract:

Morphological analysis of cell shapes requires segmentation of cell contours from input images and subsequent extraction of meaningful shape descriptors that provide the basis for qualitative and quantitative assessment of shape characteristics. Here, we describe the publicly available ImageJ plugin PaCeQuant and its associated R package PaCeQuantAna, which provides a pipeline for fully automatic segmentation, feature extraction, statistical analysis, and graphical visualization of cell shape properties. PaCeQuant is specifically well suited for analysis of jigsaw puzzle-like leaf epidermis pavement cells from 2D input images and supports the quantification of global, contour-based, skeleton-based, and pavement cell-specific shape descriptors.

Link:

Note on the following publication:

SEMI-AUTOMATIC CELL SEGMENTATION FROM NOISY IMAGE DATA FOR QUANTIFICATION OF MICROTUBULE ORGANIZATION ON SINGLE CELL LEVEL

Birgit Möller [*] *and Katharina Bürstenbinder* [†]

[*]Inst. of Computer Science, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany
[†]Dept. of Molecular Signal Processing, Leibniz Institute of Plant Biochemistry, Halle (Saale), Germany

ABSTRACT

The structure of the microtubule cytoskeleton provides valuable information related to morphogenesis of cells. The cytoskeleton organizes into diverse patterns that vary in cells of different types and tissues, but also within a single tissue. To assess differences in cytoskeleton organization methods are needed that quantify cytoskeleton patterns within a complete cell and which are suitable for large data sets. A major bottleneck in most approaches, however, is a lack of techniques for automatic extraction of cell contours. Here, we present a semi-automatic pipeline for cell segmentation and quantification of microtubule organization. Automatic methods are applied to extract major parts of the contours and a handy image editor is provided to manually add missing information efficiently. Experimental results prove that our approach yields high-quality contour data with minimal user intervention and serves a suitable basis for subsequent quantitative studies.

Index Terms— cytoskeleton, cell segmentation, vesselness, gap closing, label editor, texture, clustering, ImageJ

1. INTRODUCTION

Cell morphogenesis in plant cells is largely driven by cytoskeleton organization. The structure of the cytoskeleton is highly dynamic and undergoes significant changes, e.g., during growth or in response to environmental stimuli [1]. Quantitative analysis of cytoskeleton properties and structural characteristics helps to understand developmental events, including the cellular processes underlying morphogenesis [2].

The challenges for robust and reliable quantitative analysis of cytoskeleton organization are manifold. Large variations in cytoskeleton structure are observable between different cell types and tissues, but sometimes also among cells in a single tissue. This demands for sufficiently large and representative data sets in quantitative analysis to cover the full range of natural variation. Moreover, characteristics of the complete cytoskeleton of an individual cell need to be considered rather than global properties of complete images or samples taken from selected image positions only.
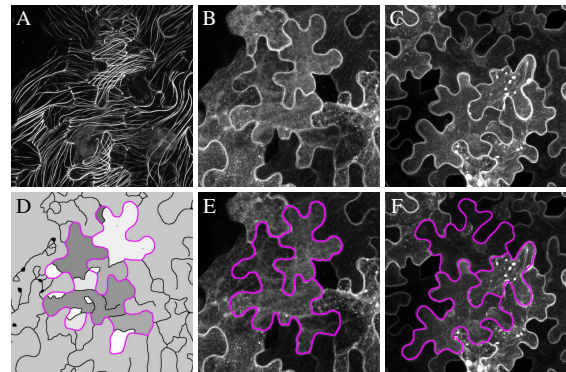
Fig. 1. Top, (contrast enhanced) images of FP-labeled microtubules (A) and cell membranes (B,C). Bottom, intermediate (D) and final (E,F) results of contour extraction.

Recent advances in cell biology allow for imaging of cytoskeleton structures at high spatial and temporal resolution [3], and for simultaneous fluorescent protein (FP)-labeling of cell membranes. Extracting cell contours, however, is still challenging due to noisy images of locally varying contrast (Fig. 1, 2). Thus, unbiased and largely automated methods for cell segmentation and cytoskeleton analysis are still rare.

Here, we present a new semi-automatic workflow for analyzing cytoskeleton organization on the single cell level. Compared to manual annotation, which often is the standard for extracting contour data, our approach reduces the need for user invention and allows to include larger sets of cells in comparative studies. The method builds on two key components, a new efficient semi-automatic procedure for contour extraction and an unsupervised structure analysis technique adopted from [4]. The contour extraction is based on vesselness enhancement filters and gap closing. These methods usually extract most parts of the contours, but cannot always guarantee complete contours due to the difficulty of the task. We, thus, provide an image editor for manual post-processing that allows to extract high-quality contour data even from large sets of noisy images in a time-efficient way.

Quantitative assessment of cytoskeleton structures often relies on the analysis of individual fibers combined with manual selection of sample regions in cells of interest. To avoid

fiber segmentation we adopt an unsupervised approach based on texture analysis and clustering using local binary patterns as features. The method has proven suitable for analysis of cytoskeleton organization in mammalian and plant cells [4, 5]. Here, we show results for cells from *Nicotiana benthamiana*.

All components of the analysis pipeline and the label image editor are implemented in Java as part of the open source image analysis toolbox MiToBo [6, 7] and are available as operator plugins for ImageJ/Fiji via MiToBo's update site.

2. RELATED WORK

The analysis of cytoskeleton organization in cells using image analysis methods has been of interest for several years. Many approaches focus on segmenting single fibers and quantify their orientation and length to characterize cytoskeleton properties [8]. Often the analysis relies on either complete images as samples for a certain cell type [9] or on selected regions only. This is prone to bias and does not cover the full variety of cells in an image or tissue. To overcome this lack Cui et al. [10] propose to couple actin analysis and cell segmentation. They apply classification methods, but rely on relatively strict assumptions about actin properties not easily generalizable.

In many cases, contour data cannot be extracted from images of FP-labeled microtubules. In our data, even for biological experts, it is challenging to locally trace the cell contours accurately solely based on this data as no explicit contour information is present (Fig. 1A). Thus, contour segmentation is completely decoupled from cytoskeleton quantification.

The complexity of contour extraction highly depends on the type, quality and resolution of the input data. Cell properties also have a notable impact on the segmentation. The collection of potential methods for cell segmentation is manifold [11] and often methods are specially dedicated to a certain cell type and acquisition scenario. For largely homogeneous and regularly shaped cells binarization methods combined with morphological operations or a watershed transformation are common [12]. More advanced techniques subsume active contours. They are, e.g., well-suited if nuclei can be used as initialization from which the cell region is expanded or for incorporating assumptions about cell shape [13]. If appropriate training data is available also machine learning techniques have proven suitable [14]. Methods aiming to extract contours directly by exploiting boundary information are rare as these are often sensitive to noise. Due to the large region inhomogeneity in our data we use a boundary-based approach.

3. METHODS

Our pipeline consists of a new semi-automatic cell segmentation stage (Sec. 3.1) and subsequent extraction and clustering of texture features (Sec. 3.2, for details refer to [4]) from two-channel microscopy image z-stacks with FP-labeled cell contours and microtubules, respectively.

3.1. Cell Contour Segmentation

To extract cell contours we use maximum projections of the membrane channel from given z-stacks (Fig. 1B,C). Pixels of the anticlinal cell contour are generally brighter than pixels of the periclinal contour (i.e., the cell surface), cytoplasm or background. FP-labeling still might locally vary, and image quality may suffer due to spurious signals from beyond the membranes. This results in a high level of noise and a lack of contrast rendering exact localization of contours challenging.

3.1.1. Vesselness Enhancement Filtering and Binarization

Anticlinal cell contours usually form piecewise thin and linear structures. To localize them we use vesselness enhancement filters [15]. After convolving the image with a Gaussian kernel ($\sigma = 1$) we apply a vesselness filter to quantify for each pixel the similarity of the local intensity landscape to vessel-like structures typical for cell contours. The filter mask sized 15×15 px has a cross-sectional Mexican hat profile with $\sigma = 2$ to account for blurred boundaries and is applied in 18 orientations from $0°$ to $170°$ in $10°$ steps. For each pixel the maximal response to any of the 18 convolutions yields the result value stored in a vesselness map. Negative map entries are set to zero before the map is binarized using an extended Niblack operator [16]. Natively the Niblack operator extracts a threshold for each pixel from a local neighborhood NB of size n_v irrespective of the local grayscale distribution. To prevent it from binarizing homogeneous regions we compute for each pixel the grayscale variance σ_v in NB. Only if σ_v exceeds a threshold θ_v the Niblack is applied to determine a local threshold, otherwise the pixel is classified as background. This procedure yields a binary image with potential contour fragments marked as foreground regions.

To increase robustness against noise we delete regions smaller than 200 px. For the remaining ones the skeletons are extracted using the algorithm of Zhang and Suen [17] to ease further processing. Skeleton branches with a length below 40 px are pruned as they are most likely related to spurious FP-labeling signals. At this point, in image sections with sufficient local contrast and striking intensity profiles cell contours are properly localized, but may still contain small gaps.

3.1.2. Automatic Gap Closing

Gaps in contours mainly occur in image parts where a lack of local contrast results in small vesselness responses during vesselness enhancement filtering. Still, contour pixels have a higher response than the local background and shortest path techniques like the Dijkstra algorithm [18] can be applied to find a proper continuation of the local contour.

To close a gap between two contour pixels we transform the gap closing problem into a shortest path problem on a weighted and directed graph G. First, we define a region of

relevant pixels (ROP) around both pixels that most likely contains all candidate pixels for the optimal path (details see below). Then for each pixel in the ROP a node is added to G.

Edges in G connect pixel nodes according to 8-neighborhood. Each edge gets a weight that grades the edge's suitability to be included in the optimal path. This path is supposed to pass by pixels with high vesselness responses, i.e., edges connecting nodes with high values get small weights. As it makes a difference if the path traverses two nodes from the pixel with low vesselness response towards the one with a higher response or vice versa, two edges with different weights are inserted between every two neighboring nodes. The weights are set anti-proportional to the vesselness response ν_t of the edge's target pixel, i.e., are calculated by subtracting ν_t from the maximal response in the ROP. Solving the gap closing problem is then identical to finding a path in G with minimal weight sum which can be tackled with the Dijkstra algorithm.

The gap closing procedure is split-up in two phases. First, we try to connect branch end points of the contour skeleton with a maximal distance of 45 px according to maximum norm. Second, we address more complex situations where a gap is not obviously marked by two end points. We iterate over all remaining skeleton end points, search in the vicinity of each end point for skeleton pixels on other branches as potential partners for gap closing, and extract the shortest path to each partner. The path with lowest weight sum is finally chosen as candidate to close a gap. In both phases the ROPs for end points are chosen as squared regions around the pixels sized 91×91 and 61×61 px, respectively. Paths are only valid if the average weight per edge does not exceed a threshold. Finally, remaining branches shorter than 100 px are removed.

3.1.3. Manual Post-Processing with the Label Image Editor

The fraction of contours properly segmented highly depends on the quality of the input data, but even in noisy data usually most parts are robustly extracted. In images with lots of spurious signals the approach tends to over-segmentation. If desired contours are still localized correctly this can easily be corrected by merging sub-regions. A more serious source of inaccuracies are image regions with ambiguous or completely missing contour information. This results in large gaps which cannot be closed automatically and require user input. To ease manual post-processing we provide an editor for label images. Segmented regions can be merged or removed and the image be relabeled. The editor supports filling holes and closing small gaps by drawing free-hand lines into an image. As most gaps are small this allows to complete contours efficiently.

3.2. Quantification of Cytoskeleton Structures with LBPs

Extracted cell contours yield the basis for a quantitative comparison of cytoskeleton organization on the single cell level. To quantify structural properties of the cytoskeleton we follow the ideas proposed in [4] utilizing maximum projections of the cytoskeleton images. The area of each cell is divided into 16×16 px non-overlapping patches. For each patch a 68-dimensional feature vector of rotation invariant uniform local binary patterns (LBP) [19] is extracted compiled from 4 histograms of LBP codes for radii of $1, 1.5, 2$ and 3 around each pixel. All vectors of all cells of a data set are commonly clustered using k-means ($k = 6$) to identify groups of similar patterns. Each group represents a certain cytoskeletal structure jointly appearing in the cells. From the group indices of all patches in each cell a characteristic distribution of patterns per cell can be derived which serves as a structural fingerprint and provides a suitable basis for subsequent comparative analysis.

4. EXPERIMENTAL RESULTS

The semi-automatic segmentation pipeline was experimentally evaluated on a set of 38 images (Fig. 2) acquired from leaf epidermis cells of *Nicotiana benthamiana* upon transient overexpression of 5 different isoforms of the microtubule-associated protein 65 (MAP65) family[1]. MAP65 proteins are microtubule binding and bundling proteins with important roles in plant development [20]. We restrict the evaluation to cells located completely inside the image area to only include cytoskeletons of complete cells resulting in 75 cells in total.

Vesselness filtering and automatic gap closing successfully extract contours from input images with sufficiently high local contrast and few distracting signals (Fig. 1E,F). The algorithm provides smooth boundaries hardly to acquire manually in reasonable time as can be seen from most parts of the contours in Fig. 2, G-L. Inaccuracies in contour localization mainly result from local ambiguities (Fig. 2I) or point signals (Fig. 2H) which yield significant vesselness responses and may also disturb the gap closing (Fig. 2L, green circle).

Manual annotation of cell contours that could not properly be segmented automatically is time-consuming and tedious. If too many contours are returned, however, this is a less serious problem as long as the correct cell contour is still included. Our supplemental label image editor facilitates fast and easy elimination of surplus contours by eliminating or merging corresponding sub-regions with few mouse-clicks yielding high-quality segmentation results. Thus, to not exclude contour parts too early in the process the parameters for the Niblack in local binarization were empirically chosen following a rather liberal strategy ($k = $-1.0, $n_v = 21$, $\theta_v = 3.0$).

As result of the initial automatic segmentation, cell contours are often completely segmented (Fig. 1E,F / Fig. 2 J-L, magenta lines), but cell areas are partitioned into several sub-regions (Fig. 1D, sub-regions shown in different gray values; Fig. 2G). Yet sometimes contour extraction locally fails due to ambiguous intensity structures (Fig. 2 H,I). While most of the remaining gaps are small and can properly be closed by our automatic gap closing heuristic (Fig. 2 J,K, blue parts), larger
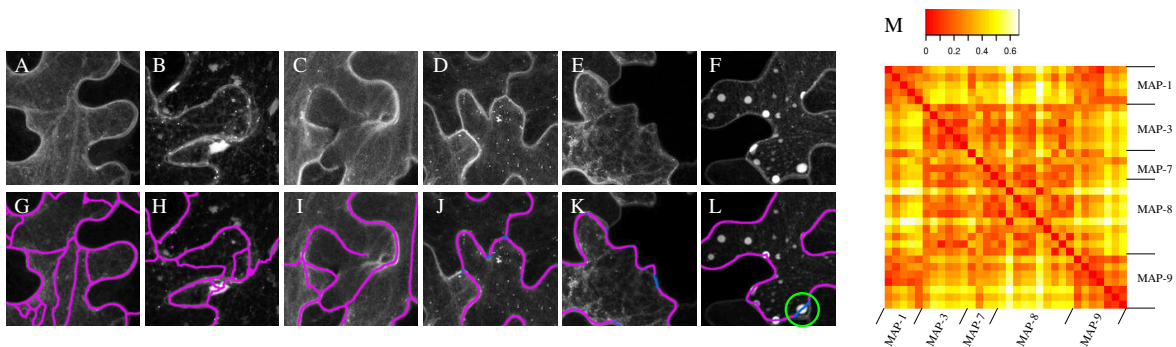

Fig. 2. Left, clips from the 38 test images: top, maximum projections of membranes (A-F), bottom, segmentation results before (G-I) and after (J-L) manual post-processing. Automatically extracted contours are shown in magenta, gaps automatically closed in blue. (M) Heatmap of pairwise distances between fingerprint vectors for 32 cells, grouped by MAP65 isoform.

gaps require user input either using our label editor or more flexible image editing tools like ImageJ, Gimp or Photoshop.

Fig. 3 shows for all 38 images the recalls and precisions of the automatic contour segmentation and gap closing steps, where we use the final, manually post-processed contour segmentation result of each image as reference for the evaluation. The black line shows the fractions of contour pixels automatically extracted in the initial contour segmentation step, and the red line refers to the fraction of pixels extracted after automatic gap closing. Vesselness enhancement and binarization already extract for 36 images more than 85% of the contour pixels, and for 29 of them even more than 90%. After automatic gap closing these numbers further increase to a fraction of more than 0.95 for most images with a minimum of 0.79 for one image with extremely low contrast. Hence, most of the time a maximum of 5% of the contours require manual tracing, while for 5 images there is no need for manually filling-in gaps at all. The fractions of true contour pixels among all extracted ones are shown in blue and vary between ≈ 0.6 and 1.0. For 32 images the precision exceeds 0.75, while for 6 images it is notably smaller due to small holes included in the evaluation. As filling these holes is a simple one-click operation in our editor like region merging or deletion, even such fair precision rates are still acceptable when taking the complete semi-automatic workflow into account.

Extracted contours are used to quantify cytoskeleton organization on the level of single cells. Usually not all segmented cells are suitable for analysis due to differences in FP expression and fluorescence intensities. A suitable subset can be chosen with our editor during post-processing. We show quantification results for a subset of 32 cells. In Fig. 2M a heatmap of pairwise distances of structural fingerprint vectors for the cells is shown. Cells belonging to the same MAP65 isoform tend to be more similar to each other than to cells of other isoforms. Yet also among cells of a single isoform variation is observable stressing the importance for single cell
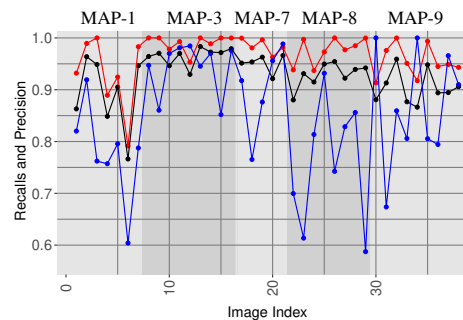
Fig. 3. Recalls of contour pixels before (black line) and after (red line) automatic gap closing for 38 images covering five MAP65 isoforms as indicated by the background shading. Precisions after gap closing are shown by the blue line.

analysis based on accurate cell contour segmentation.

5. CONCLUSION

Unbiased segmentation of cell contours is best achieved applying fully automatic methods. However, in cytoskeleton analysis data is often very noisy and not treatable with fully-automatic methods. Our new approach combining robust automatic contour extraction with efficient manual post-processing enables time-efficient and comfortable contour extraction even from such data. Future work will target at further reducing the fraction of manual user intervention towards a fully automatic pipeline for cell contour extraction.

6. REFERENCES

[1] X. Chen, S. Wu, et al., "Environmental and endogenous control of cortical microtubule orientation," *Trends in Cell Biology*, vol. 26, no. 6, pp. 409–419, 2016.

[2] H.G. Yevick and A.C. Martin, "Quantitative analysis of cell shape and the cytoskeleton in developmental biology," *Wiley Interdisciplinary Reviews: Developmental Biology*, p. e333, 2018.

[3] J. Dyachok, A. Paez-Garcia, et al., "Fluorescence imaging of the cytoskeleton in plant roots," in *Cytoskeleton Methods and Protocols*, pp. 139–153. Springer, 2016.

[4] B. Möller, E. Piltz, and N. Bley, "Quantification of actin structures using unsupervised pattern analysis techniques," in *Proc. of Int. Conf. on Pattern Recognition (ICPR)*, Stockholm, 2014, pp. 3251–3256.

[5] K. Bürstenbinder, B. Möller, et al., "The IQD family of calmodulin-binding proteins links calcium signaling to microtubules, membrane subdomains, and the nucleus," *Plant Physiology*, vol. 173, no. 3, pp. 1692–1708, 2017.

[6] B. Möller, M. Glaß, et al., "MiToBo - a toolbox for image processing and analysis," *Journal of Open Research Software*, vol. 4, no. 1, pp. e17, 2016.

[7] MiToBo Development Team, "MiToBo - A microscope image analysis toolbox," https://mitobo.informatik.uni-halle.de, 10/1/2019.

[8] K. Akita, T. Higaki, et al., "Quantitative analysis of microtubule orientation in interdigitated leaf pavement cells," *Plant Signaling & Behavior*, vol. 10, no. 5, pp. e1024396, 2015.

[9] C. Matschegewski, S. Staehlke, et al., "Automatic actin filament quantification of osteoblasts and their morphometric analysis on microtextured silicon-titanium arrays," *Materials*, vol. 5, no. 7, pp. 1176–1195, 2012.

[10] C. Cui, J. JaJa, et al., "Quantifying the astrocytoma cell response to candidate pharmaceutical from f-actin image analysis," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009, pp. 5768–5771.

[11] F. Xing and L. Yang, "Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: A comprehensive review," *IEEE Reviews in Biomedical Engineering*, vol. 9, pp. 234–263, 2016.

[12] E. Meijering, "Cell segmentation: 50 years down the road," *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 140–145, 2012.

[13] C. Molnar, I.H. Jermyn, et al., "Accurate morphology preserving segmentation of overlapping cells based on active contours," *Scientific Reports*, vol. 6, pp. 32412, 2016.

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[15] M.M. Fraz, P. Remagnino, et al., "Blood vessel segmentation methodologies in retinal images - a survey," *Computer Methods and Programs in Biomedicine*, vol. 108, no. 1, pp. 407–433, 2012.

[16] W. Niblack, *An Introduction to Digital Image Processing*, Prentice Hall, 1986.

[17] T. Y. Zhang and C. Y. Suen, "A fast parallel algorithm for thinning digital patterns," *Commun. ACM*, vol. 27, no. 3, pp. 236–239, Mar. 1984.

[18] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271, Dec 1959.

[19] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, 2002.

[20] C. Mollinari, J.-P. Kleman, et al., "PRC1 is a microtubule binding and bundling protein essential to maintain the mitotic spindle midzone," *The Journal of Cell Biology*, vol. 157, no. 7, pp. 1175–1186, 2002.

Note on the following publication:

COMPARING ACTIVE CONTOURS FOR THE SEGMENTATION OF BIOMEDICAL IMAGES

Birgit Möller and Stefan Posch

Institute of Computer Science, Martin Luther University Halle-Wittenberg
Von-Seckendorff-Platz 1, 06120 Halle (Saale), Germany

ABSTRACT

Application of active contours for image segmentation raises the question of contour representation, i.e. whether to use snakes or level sets. The representation directly affects issues like topology-preservation and energy optimization. In this paper we aim to contribute to the understanding of specific characteristics of contour representations with a detailed comparison of snakes vs. non-PDE level sets. Based on the same energy functional and applied to different kinds of real-world data our experiments show minor differences in segmentation quality, but outline important distinctions regarding implementation, parameter settings and computational effort.

Index Terms— Snakes, non-PDE level sets, comparison

1. INTRODUCTION

Active contours have become a popular technique for image segmentation in various domains over the last two decades since the publication of the seminal papers [1] and [2]. The underlying idea of these methods is to minimize an appropriate energy functional. This energy comprises a data term, also called external energy, and typically a regularizer or internal energy which depends exclusively on the shape of the contour(s). The choice of this energy depends heavily on the images considered and on the aim of the segmentation task.

Active contours are distinguished into snake and level set methods. Snakes explicitly represent a contour by a parametric curve, where as level sets use an implicit representation as the zero level of an embedding function. Despite the common idea of both techniques to minimize a given energy functional, they differ considerably with regard to the capacity to handle multiple objects, the ability to change or preserve topology, the optimization techniques including numerical stability, as well as computation time and implementational effort. Consequently to efficiently apply such techniques to segmentation problems from a specific domain it is fundamental to develop understanding and experience for the characteristics of different active contour approaches and their behaviors.

The main intention of this paper is to contribute to this understanding by directly comparing one snake and one non-PDE level set approach for two segmentation problems in microscopy images from the biomedical domain [3, 4]. In con-

trast to [5] for both techniques the same energy functional based on the Chan-Vese energy [6] is used, which is region-based in contrast to edge-based energies considered in [5]. To provide practically meaningful results original images from biomedical applications are used in the experiments, and besides qualitative and quantitative results also implementation issues and computation times are analyzed in detail.

2. RELATED WORK

Active contours date back to the pioneering work of [1] where snakes where introduced for energy-based image segmentation. The inherent nature of snakes impedes topological changes during segmentation. Although in [7] an approach to overcome this limit was presented, snakes are typically used in topology-preserving contexts [8, 9]. Regarding biomedical applications often multiple objects per image have to be segmented. To this end energy functionals hampering snake overlap have been developed [10]. Also geometric post-processing of snakes to eliminate self-overlap is often required. The model of explicit snake segmentation defines a PDE problem which is usually solved by iterative gradient-descent [1]. Due to numerical issues that sometimes occur , greedy optimization techniques were proposed [11], and level sets have gained increasing importance and wide-spread use due to their advantageous numerical properties.

As for snakes the prevailing approach for optimization of level sets is to use Euler-Lagrange equations to derive an iterative solution of the underlying variational problem (e.g. [2, 6]). The associated questions of time constants, re-initialization, and computation times can be bypassed by the direct non-PDE method proposed in [12] to minimize the energy. Pursuing a greedy approach, for each pixel the phase is changed if this decreases the global energy. Especially if this energy difference can be evaluated locally as for the Chan-Vese (CV) fitting term, this optimization strategy often yields a very fast scheme. Level set representations inherently allow to change the topology, e.g. to split or merge contours. However, this is of disadvantage in applications where the number and approximate position of objects to segment is given or determined by an initialization phase. A topology preserving scheme based on topological numbers was proposed in [13] which we adopted to the non-PDE
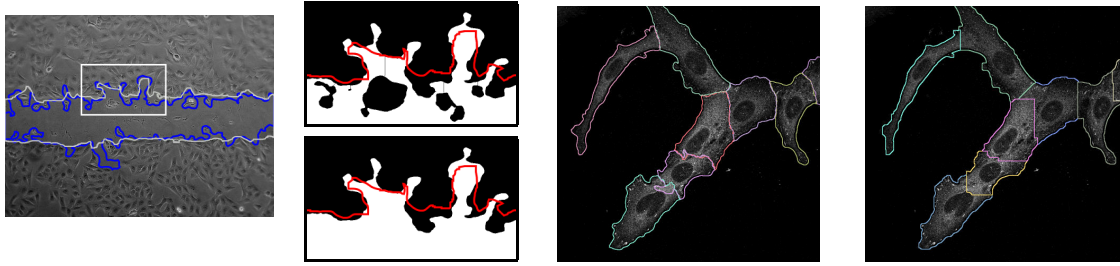
131

Fig. 1. From left to right images (contrast enhanced) and segmentation results are shown: (i) A scratch assay image with the final snake contour in dark grey (coloured version: blue) and the expert ground-truth in light gray; (ii) the level set result for a region of the left image (cf. to white rectangle) before (top) and after (bottom) post-processing, overlaid with the corresponding ground-truth contour in light grey (coloured version: red); (iii) snake and (iv) level set results for a sample cell image.

method in previous work [3]. Application of PDE level sets include the segmentation of leukocyte in microscopic images where both the CV fitting energy and an edge-based term are employed [14]. Cell segmentation in time-lapse fluorescence microscopy images in [15, 16] uses the CV energy with both the non-PDE and the PDE level set optimization.

3. METHODS

For our experiments we employ the commonly used CV energy [6]. In a level set framework it is given as

$$E(\phi, c_{fg}, c_{bg}) = \lambda_f \int_\Omega |I(\vec{x}) - c_f|^2 H(\phi(\vec{x})) d\vec{x} +$$

$$\lambda_b \int_\Omega |I(\vec{x}) - c_b|^2 (1 - H(\phi(\vec{x}))) d\vec{x} + \mu \int_\Omega |\nabla H(\phi(\vec{x})| d\vec{x},$$

where ϕ denotes the level set function and H the heaviside function. c_{fg} and c_{bg} are the mean intensities of foreground and background according to the current state of ϕ, and $\lambda_f, \lambda_b, \mu$ are weights of the three energy terms. Both explicit and implicit active contours were implemented in Java in terms of MiToBo[1] operators and ImageJ plugins.

The snake implementation follows the original work of Kass et al. [1], i.e. adopts a combined implicit and explicit gradient-descent technique for solving the PDE, resulting in a linear system of equations to be solved in each iteration. The basic parameters for snake segmentation to be chosen properly are a step size γ for gradient-descent and the preferred length s of each snake segment which is related to the number of snake points used for contour approximation. The segment length is updated in each iteration by contour resampling. Although snakes are not able to change topology during optimization non-simple polygons might occur from self-overlap which are explicitly simplified by polygon tracing. This modifies the segmentation, e.g., intrusions might be removed.

If multiple objects are present in an image, snake segmentation of all objects is done in parallel by concurrently optimizing one snake for each object. To prevent snakes from overlapping a coupling term is added to the energy functional which penalizes snake overlap, however, does not guarantee non-overlapping snakes (see [10] for details).

The implementation of the non-PDE level set optimizer follows [12] with the sequential update scheme, i.e. Gauss-Seidel iteration. If the length term of the CV energy is used. forward differences are used to approximate the gradient of the heaviside function. To preserve topology we adopt the method proposed in [13] for PDE-based minimization which is based on topological numbers. However, we enforce to preserve topology only for the foreground phases (see also [3]).

4. RESULTS

For comparing the two active contour techniques they were each applied to two sets of images originating from different biomedical applications. Below we outline the data and evaluation methods, and provide a comprehensive discussion of qualitative as well as quantitative results.

Test Data   The first test set $D1$ consists of microscopy images of scratch assays which are often used for cell motility assessment [3]. Each assay contains a mono layer of U2OS cells in which a scratch is mechanically induced. Subsequently this "wound" is closed by cell migration, which is monitored at different time points to evaluate cell motility. For $D1$, images of 10 cell populations were taken at 7 points in time yielding in total 70 images with scratches showing a large variety in size and shape. The second set $D2$ consists of 8 epifluorescence microscopy images of 57 cells in total with fluorescently labeled focal contacts. For biomedical investigations the number of focal contacts per cell is of interest requiring an accurate extraction of the cell boundaries. Comparison of segmentation results is done qualitatively and quantitatively, given manual ground-truth labelings for both sets. Scratch assays were labeled by two biologist experts and the intersection of their labelings was used as ground-truth, while labelings of one person were available for images of $D2$. Example images from $D1$ and $D2$ are given in Fig. 1.
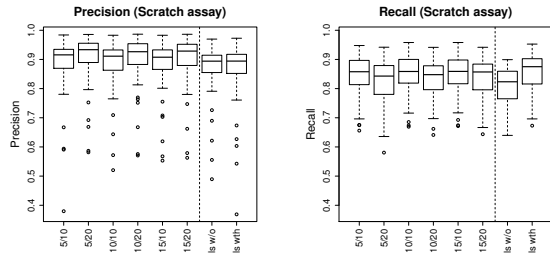
---

[1]http://www.informatik.uni-halle.de/mitobo

Fig. 2. $\mathcal{R}$ and $\mathcal{P}$ for dataset $D1$. The first six boxes in each plot refer to different snake configurations, the final two refer to level set segmentation without and with post-processing.

Performance Measures For quantitative evaluation we provide recall $\mathcal{R}$ and precision $\mathcal{P}$, where $\mathcal{R} = \frac{TP}{TP+FN}$ and $\mathcal{P} = \frac{TP}{TP+FP}$. TP and FP denote the number of pixels correctly resp. incorrectly segmented as foreground (i.e. scratch or cell area), while FN is the number of ground truth foreground pixels segmented as background. In some experiments parameters were chosen using F-scores defined as $2 \cdot \frac{\mathcal{P} \cdot \mathcal{R}}{\mathcal{P}+\mathcal{R}}$ which is proportional to the Jaccard index used in [5].

Scratch Assays Images For segmenting the images of $D1$ the common CV energy functional [6] with $\lambda_f = \lambda_b = 1$ and $\mu = 0$ is used, i.e. only the CV fitting term is employed. It is, however, applied to an image of local entropies (cf. to [3]). Initially each assay image is smoothed by a Gaussian with $\sigma = 5$, and then for each pixel the local entropy in a window sized 31×31 is calculated. As each image contains one horizontal scratch, the scratch is initialized as a rectangle in the middle of the image spanning 95% of its width and 10% of its height. In addition, as each image contains exactly one scratch topology-preservation is enforced.

In Fig. 1, second column, clips of a prototypical segmentation result for level sets are shown. The scratch boundary is accurately extracted, however, due to topology-preservation intrusions remain which are linked to the cell area by corridors of one pixel width (top). To eliminate these artifacts morphological post-processing by closing with a squared mask of size 3 followed by hole filling is applied (see Fig. 1).

In Fig. 2 box plots of $\mathcal{R}$ and $\mathcal{P}$ for both segmentations are shown. Level set segmentation results in a median $\mathcal{R}$ of 0.8234 and a median $\mathcal{P}$ of 0.8941. As expected, removing intrusions yields a distinctly higher median $\mathcal{R}$ of 0.8751, where median $\mathcal{P}$ is barley affected as scratch boundaries are not expanded and, thus, no new FP are introduced.

Applying snakes suitable configuration parameters are required, i.e. segment length s and step size γ, and a suitable termination criterion needs to be applied. In all cases optimization was stopped at iteration t if the smoothed relative change in snake area between $t-5$ and $t-15$ dropped below 0.001. To find suitable s and γ, segmentation was accomplished with different combinations of parameters, i.e. $\gamma \in \{5, 10, 15\}$ and $s \in \{10, 20\}$. The left six boxes in each plot

of Fig. 2 summarize the results. On the one hand larger s, which result in a coarser contour approximation, increase $\mathcal{P}$ about 0.0180 on average, and lead to a slight decrease in $\mathcal{R}$ of -0.0092 on average. On the other hand the step size γ does not appear to have significant influence as the median $\mathcal{P}$ for all segment lengths lie in ranges of 0.9117 ± 0.0040 and 0.9299 ± 0.0035, respectively. Likewise the median $\mathcal{R}$ for constant segment length are in the ranges of 0.8584 ± 0.0006 and 0.8500 ± 0.0070. For subsequent comparisons the parameters $\gamma = 5$ and $s = 10$ resulting in the largest median F-score were used. Exemplary results are given in Fig. 1 (left). Post-processing is not required for snakes due to polygon simplification which removes intrusions as explicitly done via post-processing for level sets (cf. Sec. 3).

The quantitative results for snake segmentation suggest to favor smaller snake segments to achieve higher $\mathcal{R}$. This is particularly true if objects with very ragged contours are to be segmented, like the scratch contours. There are numerous examples in the test images where only the use of smaller segments allows for concavities to be properly segmented. Note, however, that computational costs increase with decrease of s as this results in larger numbers of polygon points.
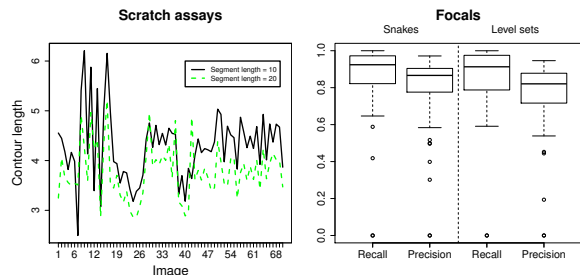

Fig. 3. Left: Lengths of extracted snake contours for all images of data set $D1$ are for $s = 10$ and $s = 20$, respectively. Right: Box plots of $\mathcal{R}$ and $\mathcal{P}$ the snake and level set segmentations of the cell images in data set $D2$.

Another interesting observation results from correlating the segment length s with the overall length of final contours. As can be seen in Fig. 3, left, there is an obvious tendency of snakes with larger segments to yield shorter object contours. Accordingly, by varying segment length an implicit length regularizer is introduced into snake segmentation.

A direct comparison of the F-scores for snake (0.8828) and level set segmentation (0.8868) does not induce a clear tendency towards either of the two techniques. Level sets tend to yield higher $\mathcal{R}$ while snakes result in larger $\mathcal{P}$. But, there is a significant difference regarding the computational effort of both approaches. While the median processing time for level sets is $118s$ with a maximum of $241s$, the median for the snakes is $540s$ with a maximum of $49min$. In each snake iteration the new set of snake points is calculated by solving a linear system of equations containing twice as many variables as snake points, thus, the computation time is directly related to the number of snake points and increases with growing

133

contour length and decreasing s. Also smaller γ often coincide with higher computation times. For level sets the effort of each iteration depends on the number of contour pixels due to topology preservation which is directly related to the compactness of the object region. Elongated structures require more iterations than compact ones.

Cell Images To segment the borders of individual cells in the images of $D2$ again the CV fitting energy is used. For preprocessing the images are smoothed by Gaussian with $\sigma = 4$ which blurs the focal contacts' intensities.

Adopting snakes for segmentation a parameter ratio of $\frac{\lambda_f}{\lambda_b} = \frac{1}{150}$ was chosen for the CV energy, and in addition a snake overlap penalty was applied with a weight of 0.6 relative to the CV term. As termination criterion a maximum of 700 iterations was used, or alternatively a relative change in area between two iterations of not more than 0.001. Considering again F-scores a segment length of $s = 5$ and a step size of $\gamma = 2$ turned out to be favorable, yielding a median $\mathcal{R}$ of 0.9241 and a median $\mathcal{P}$ of 0.8663.

Using level sets with the same CV fitting term proper segmentation was impossible due to missing regularization. This reflects the sensitivity to noise as also observed, e.g., in [15]. It seems that the implicit length regularizer hypothesized above allows snakes to successfully segment objects in absence of an explicit regularizer. These findings show that the same energy does not necessarily lead to comparable results on the same data when used with implicit active contours compared to explicit ones as might have been expected.

These shortcoming of the level sets were overcome including the length penalty term of the CV energy into the functional, i.e. setting $\mu = 2000$. Box plots of quantitative segmentation results are shown in Fig. 3. For level sets a median $\mathcal{R}$ of 0.9130 and a median $\mathcal{P}$ of 0.8207 is observed, resulting in a median F-score of 0.8502 compared to 0.8775 for snake segmentation. However, as can also be seen from the qualitative segmentation results in Fig. 1 (right), the differences are marginal. Again it is difficult to favor one of the techniques, provided parameters are properly chosen for each technique. The main difference between snakes and level sets is again computational effort. The median time for level sets is $30s$ and $1014s$ for snake segmentation.

5. CONCLUSION

Using the same energy functional, snakes and non-PDE level sets achieve comparable results of high-quality. But, regarding implementation details, necessity for data pre- and postprocessing, and computational effort there are distinct differences. Particularly our findings support the conjecture that snakes incorporate an implicit length penalty varying with the segment length. Depending on the application each of both representations may have advantages, and knowing their specifics is essential for chosing the adaquate representation.

6. REFERENCES

[1] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *IJCV*, vol. 1, pp. 321–331, 1988.

[2] S. Osher and J.A. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations," *Journal of Computational physics*, vol. 79, no. 1, pp. 12–49, 1988.

[3] M. Glaß et al., "Scratch assay analysis with topology-preserving level sets and texture measures," in *Proc. IbPRIA*, Gran Canaria, 2011, pp. 100–108.

[4] B. Möller et al., "MiCA - easy cell image analysis with normalized snakes," in *MIAAB*, K. Rohr, Ed., 2011.

[5] L. He et al., "A comparative study of deformable contour methods on medical image segmentation," *Image and Vision Comp.*, vol. 26, no. 2, pp. 141 – 163, 2008.

[6] T.F. Chan and L.A. Vese, "Active contours without edges," *IEEE T-IP*, vol. 10, no. 2, pp. 266–277, 2001.

[7] T. McInerney et al., "T-snakes: Topology adaptive snakes," *Med. Image Anal.*, vol. 4(2), pp. 73–91, 2000.

[8] N. Ray et al., "Active contours for cell tracking," in *Proc. of 5th IEEE Southw. Symp. on Image Analysis and Interpretation*, Washington, DC, USA, 2002, p. 274 ff.

[9] H. Shen et al., "Automatic tracking of biological cells and compartments using particle filters and active contours," *Chemometrics and Intelligent Laboratory Systems*, vol. 82, no. 1-2, pp. 276–282, May 2006.

[10] C. Zimmer et al., "Coupled parametric active contours," *IEEE T-PAMI*, vol. 27, no. 11, pp. 1838–1842, 2005.

[11] D.J. Williams and M. Shah, "A fast algorithm for active contours and curvature estimation," *CVGIP: Image Underst.*, vol. 55, pp. 14–26, January 1992.

[12] B. Song and T. Chan, "A fast algorithm for level set based optimization," *UCLA Cam Report 02-68*, 2002.

[13] X. Han, C. Xu, and J.L. Prince, "A topology preserving level set method for geometric deformable models," *IEEE T-PAMI*, vol. 25, no. 6, pp. 755–768, 2003.

[14] D.P. Mukherjee, N. Ray, and S.T. Acton, "Level set analysis for leukocyte detection and tracking," *IEEE T-IP*, vol. 13, no. 4, pp. 562–572, 2004.

[15] O. Dzyubachyk et al., "Advanced level-set-based cell tracking in time-lapse fluorescence microscopy," *IEEE T-MI*, vol. 29, no. 3, pp. 852 –867, march 2010.

[16] A. Dufour et al., "Segmenting and tracking fluorescent cells in dynamic 3-D microscopy with coupled active surfaces," *IEEE T-IP*, vol. 14(9), pp. 1396–1410, 2005.

3.5 Adaptive Segmentation of Particles and Cells for Fluorescent Microscope Imaging

Abstract:

Analysis of biomolecules in cells essentially relies on fluorescence microscopy. In combination with fully automatic image analysis it allows for insights into biological processes on the sub-cellular level and thus provides valuable information for systems biology studies. In this paper we present two new techniques for automatic segmentation of cell areas and included sub-cellular particles. A new cascaded and intensity-adaptive segmentation scheme based on coupled active contours is used to segment cell areas. Structures on the sub-cellular level, i.e. stress granules and processing bodies, are detected applying a scale-adaptive wavelet-based detection technique. Combining these results yields fully automated analyses of biological processes, and allows for new insights into interactions between different cellular structures and their distributions among different cells. We present an experimental evaluation based on ground-truth data that confirms the high-quality of our segmentation results regarding these aims and opens perspectives towards deeper insights into biological systems for other problems from systems biology.

Link:

https://link.springer.com/chapter/10.1007/978-3-642-25382-9_11

Note on the following publication:

154

Cascaded Segmentation of Grained Cell Tissue with Active Contour Models

Birgit Möller*, Nadine Stöhr†, Stefan Hüttelmaier†, Stefan Posch*

*Institute of Computer Science, Martin Luther University Halle-Wittenberg, Halle, Germany

†ZAMED, Faculty of Medicine, Martin Luther University Halle-Wittenberg, Halle, Germany

Birgit.Moeller@informatik.uni-halle.de

Abstract

Cell tissue in microscope images is often grained and its intensities do not well agree with Gaussian distribution assumptions widely used in many segmentation approaches. We present a new cascaded segmentation scheme for inhomogeneous cell tissue based on active contour models. Cell regions are iteratively expanded from initial nuclei regions applying a data-dependent number of optimization levels. Experimental results on a set of microscope images from a human hepatoma cell line prove high quality of the results with regard to the cell segmentation task and biomedical investigations.

1. Introduction

One important part of the cellular response to stress is the rapid adaption of the translation processes. Stress granules (SGs) are dense aggregations in the cytosol emerging during cellular stress conditions and are assumed to be essential for mRNA storage during stress [11]. Processing bodies (PBs) on the other hand are suggested to be the places of mRNA degradation in eucaryotic cells [5]. As some SGs transiently associate with PBs this suggests that RNAs can be redirected to PBs for decay. To scrutinize these hypotheses it is essential to monitor occurrence and distribution of PBs and SGs in individual cells.

In fluorescence microscopy this is achieved by labeling certain SG/PB-components with different fluorochromes coupled to specific antibodies. Additionally, it is essential to detect the boundary and the nucleus of each cell to assign the structures to a certain cell. This allows the quantification of alterations in number, size and localization of SGs/PBs due to cell manipulations, e.g. viral infections [10]. Cell boundaries can be labeled by either antibody stainings or commercial dyes binding to membranes. Both strategies are strongly restricted as the number of antibodies reliably used in one

experiment is limited and dyes (e.g. HCS CellMask) stain not only the outer membrane but also intra-cellular parts. Hence in this work, we aim at detecting the cell boundary on existing fluorescent images without need of further labeling, where we use the fluorescence channel for PBs in the following,

For cell segmentation we apply active contour models as the basic segmentation technique. The main contribution of this paper is a new cascaded segmentation scheme for active contours which allows to cope with inhomogeneous non-Gaussian intensity distributions of target objects like the grained cell tissue in our application. The overall approach is similar to [3], however, our new cascaded segmentation scheme overcomes the need for an additional dye to explicitly label cell tissue. In cascaded segmentation we make use of the fact that intensity distributions follow a coarse spatial pattern as the local average intensity of the PB-channel decreases towards the cell boundary. Snakes are optimized in a cascaded fashion over several levels whereas contours from the former level are used for initialization. Results on a set of test images prove the high performance of the proposed method with regard to the cell segmentation task and also to biomedical evaluation.

2. Related Work

Several segmentation techniques have been adapted to the special needs of cell segmentation in fluorescent microscopy images as a means for subsequent biological analysis. A level-set based approach for segmentation and tracking of HeLa cells using various fluorescent labelings is described in [4]. For segmentation of cells the fitting term [1] is replaced with a Gaussian likelihood for the intensities with unknown variance. In the rare case of lumped cells, these are separated using the watershed transform and subsequent region merging. A similar problem is treated in [9]. While for tracking of cells a particle filter is employed, snakes are used as a post processing step based on the snake energy [7]

where the external energy is derived from the gradient vector flow field. In the integrated approach [3] for cell area segmentation and sub-cellular particle detection, cell segmentation is based on labeling with HCS Cellmask. Initial contours are heuristically extracted based on k-means clustering and subsequently adapted using snakes. The snake energy incorporates a region fitting and gradient-based data term, combined with a regularization term, a shape prior to enforce elongated cells and a coupling term for multiple cells.

In [2] cell segmentation is modelled as statistical inference based on a graphical model where a random variable is assigned to each pixel. Potential functions are defined from a DAPI channel representing DNA content and cell boundaries detected in a differential interference contrast image. Final segmentation is derived via belief propagation to estimate the joint likelihood.

3 Cell Segmentation

Active contour models have proven suitable for object segmentation as their energy functionals allow easy integration of contour- and region-related energy terms [12]. In our scenario the number of target cells per image is detected in the DAPI channel, hence we take advantage of the topology preserving characteristics of parametric active contours. Each target cell i is modeled by a parametric contour function $c_i : [0, 1] \to \mathbb{R}^2$.

3.1 Coupled Active Contours

To segment multiple cells in a single image the individual snakes are not allowed to overlap, as overlapping cells do not occur in the biomedical setup. We adopt the approach of [12] for integrated optimization of N snake contours with a common energy functional:

$$E(c_1, .., c_N) = \sum_{i=1}^{N} E_s(c_i) + \rho \sum_{i=1}^{N} \sum_{j=i+1}^{N} \int_{O_{i,j}} 1 d\Omega \quad (1)$$

$E_s(c_i)$ denotes the individual energy of a single snake, the second term penalizes for each pair of snakes c_i, c_j the overlap $O_{i,j}$ of their interior regions.

In our scenario the main clue to segment cells is region intensity. Thus we adopt the region-based snake energy [1]. Intensities of the interior of each cell and the background are modeled by different constant intensities and quadratic error terms. This is approximately equivalent to a Gaussian model with means $c_i^{in}, i = 1 \ldots N$, and c^{bg} and with variances λ_{in} and λ_{bg}:

$$E_r(c_i) = \lambda_{in} \int_{R_{in}(c_i)} \left(I(x, y) - c_i^{in} \right)^2 d\Omega + \lambda_{bg} \int_{\Omega \setminus (\bigcup_{i=1}^{N} R_{in}(c_i))} \left(I(x, y) - c^{bg} \right)^2 d\Omega, \quad (2)$$
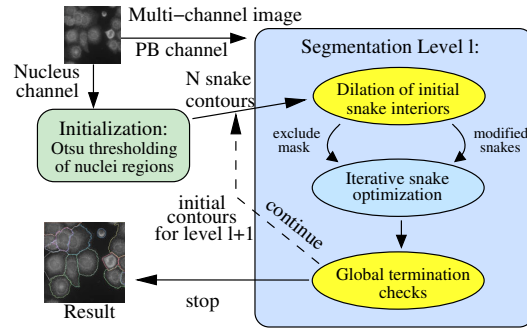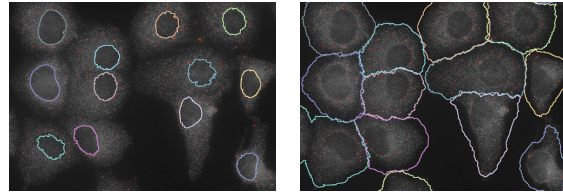

Figure 1. Cascaded snake segmentation.

Figure 2. Clip of snake contours: initial (left) and final (right), detected PBs (red).

where $R_{in}(c_i)$ denotes the interior of snake c_i. Variances λ_{in} and λ_{bg} are specified as parameters, and mean intensities are optimally estimated after each gradient descent step as the means of the corresponding regions. For the final energy of a single snake a regularization term is added penalizing curvature:

$$E_s(c_i) = E_r(c_i) + \frac{1}{2} \int_0^1 \beta \cdot \| c_i''(s) \|^2 ds \quad (3)$$

As the region intensity distributions of cells in our scenario do not agree well with the intensity model this homogenity criterion is quite restrictive. Thus we propose to embed it in a new cascaded segmentation scheme.

3.2 Cascaded Segmentation Scheme

Rather than following a Gaussian distribution the pixel intensities of the cells in our application tend to decrease monotonically with increasing distance from the nucleus region (cf. Fig. 2). Here we propose to segment such cells in a cascaded fashion by sequentially adding new cell fractions to the cell area. Fig. 1 shows an overview of the approach.

The basic idea of our approach is to replace the single optimization level commonly used with snake techniques by an iterative procedure with data-dependent numbers of levels. The state to which a snake converges at the end of one optimization level l is the basis for the initialization of the subsequent optimization level $l + 1$. In detail, the resulting snake region from

level l is dilated by 10 pixels and its contour yields the initial snake contour of level $l + 1$ to segment the adjacent cell area of darker intensities. In addition the area already segmented as part of the cell during level l is masked and thus excluded from further computations. Subsequently, c_i^{in} and c^{bg} are re-estimated from this initialization and are most likely smaller than prior values due to the intensity structure of the cells. Consequently, initializing the optimization procedure with dilated snake contours drives the snakes to further expand towards the background.

One important ingredient for such an iterative expansion scheme is a proper termination criterion. The segmentation should stop as soon as pixels are enclosed that more likely belong to the background than to the cells. To this end we consider the standard deviation $\sigma_{i,l}^{in}$ of intensity values within areas added to a cell in level l and the standard deviation within the background area σ^{bg}. As cell tissue staining is coarse-grained compared to the homogeneous background, iterative segmentation for a single cell terminates if the standard deviation of the interior intensity values of the recently added cell area falls below a threshold θ_b:

$$|\sigma_{i,l}^{in} - \sigma^{bg}| < \theta_b. \qquad (4)$$

In addition, region growth is analyzed and very small snake expansion between two levels $l - 1$ and l, i.e.

$$|1 - area(c_{i,l-1})/area(c_{i,l})| < \theta_a, \qquad (5)$$

also results in termination of the optimization for the snake in question. The overall cascaded segmentation terminates if all individual snakes have terminated according to one of both criteria or if a maximum of 9 levels is reached.

Cell segmentation is initialized with snake contours as extracted from cell nuclei regions segmented by Otsu thresholding of the DAPI nuclei channel and morphological post-processing. For the majority of cells the nucleus is detected correctly, however, sometimes nuclei of neighboring cells are merged. To separate these we apply a simple yet efficient procedure adapted from [8]. For each resulting connected component a cell is hypothesized and its contour yields an initial snake.

4. Experimental Results

For experimental evaluation of the new segmentation scheme 8 images from epifluorescence mioscropy are used. Each image consists of three channels, containing fluorescently labeled nuclei, SGs and, PBs respectively. SGs are labeled by immunostaining of TIAR (a protein localized in SGs), while PBs are labeled by immunostaining of DCP1a (decapping enzyme localized

in PBs). The nuclei are labeled by DAPI. For evaluation a manual cell labeling is available as ground truth for all images. Three of the images belong to a control sample, the remaining five are infected with a virus as mentioned in the introduction. One control and two virus images containing 47 cells in total were used as training dataset, i.e. for estimating values for λ_{in}, λ_{bg} and θ_b (Tab. 1). The other parameters were determined empirically. The remaining five images with 87 cells form the test dataset.

Level l	λ_{in}	λ_{bg}	ρ	β	θ_a	θ_b
1&2	45	250	10^5	0.75	0.013+	100
≥ 3				1.25	$0.002 \cdot l$	

Table 1. Optimization parameters.

In Fig. 3 (a) box plots of recall and precision of the cell area for all 87 cells of the test set are shown over a maximum number of 9 levels. As the cell boundary segmentation starts with the nuclei contours (Fig. 2, left) initially large fractions of the cells are missing yielding low recall values. During cascaded segmentation further parts of the cells are included (Fig. 2, right)[1] resulting in an increase of the recall from 0.505 after the 1st level to 0.902. In contrast, the precision is initially high with 0.998 as almost no non-cell pixels are included, however over time decreases to a still very satisfying value of 0.865. In some cases low precision for single cells results from inclusion of other cells' tissue, mainly at image boundaries where tissue is visible, but no corresponding nucleus exists. Note that 90% of the snakes are optimized for at least 3 levels, while more than 85% of the snakes terminate after a maximum of 5 levels.

From a biological point of view accurate cell boundary detection is required to correctly estimate number and size of SGs and PBs per cell. Both SGs and PBs are detected applying the scale-adaptive wavelet-based segmentation approach published in [6]. Some PB detection results are shown in Fig. 2. In Fig. 3 quantitative detection and evaluation results for the test dataset are shown. Fig. 3 (b) shows scatter plots of the number of SGs and PBs detected per cell in ground truth labeled vs. automatically segmented cells for the test set. Results after the first level are shown in blue, final results in red. Spearman correlation coefficients improve from 0.877 for SGs and 0.835 for PBs after the first level to very good 0.925 and 0.932 after termination of the cascaded approach, proving the high quality of segmentation.

In Fig. 3 (c) the area fractions of PBs and SGs for the 31 cells of the control set and the 56 cells of the virus-infected test cells are depicted. Area fraction is defined

---

[1]The cell at the lower border stems from a nucleus outside the clip.

(a) Evolving recall and precision of cell area. (b) Number of SGs and PBs per cell. (c) Area fraction.
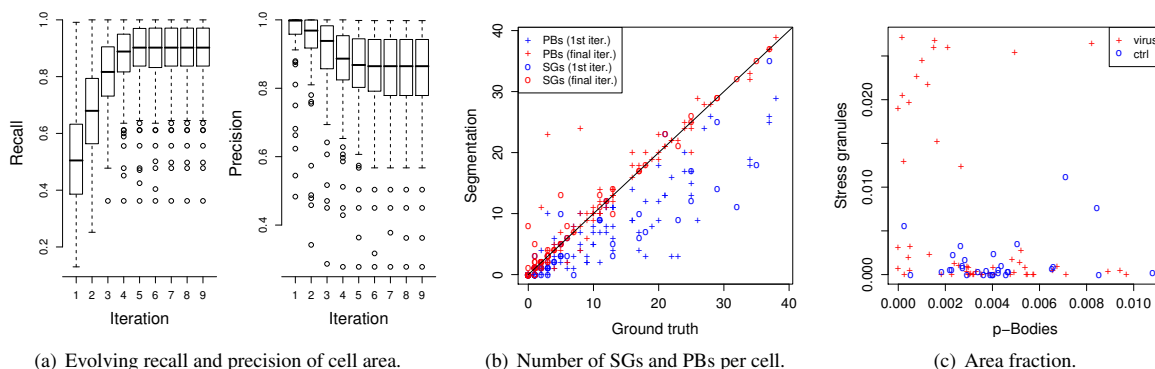
Figure 3. (a) Recalls and precisions of cell area for test set; (b) scatter plot for number of SGs and PBs per cell in ground truth and automatically segmented cells after 1st (blue) and final level (red); (c) scatter plot of area fraction for both types of particles and both cell groups.

as the ratio between the area of a cell which is populated by structures (i.e. PBs or SGs) and complete cell area. The values clearly show that control cells (blue) contain a certain number of large PBs while with a few exceptions SGs remain absent. In contrast, upon viral transfection (red) SG-formation is induced in a portion of the cells, which shows the cells presumably as indeed transfected. This correlates with a reduced PB number and size. The results clearly outline that not every cell in the image undergoes such alterations and, thus, widely-used average quantification over complete images is biased and does not reflect the situation correctly.

5. Conclusion

The paper presents a new cascaded segmentation technique based on coupled active contours which is particularly well-suited to segment objects with non-homogeneous and non-Gaussian intensity distributions. The segmentation of target cells is iteratively expanded yielding very satisfying recall and precision of cell area on the test dataset. Also comparing the numbers of structures detected per cell with ground truth data shows the high quality of segmentation results. Summarizing, the proposed automated segmentation of fluorescent images shows as a valid basis for biological analysis and interpretation of cellular processes.

References

[1] T. Chan and L. Vese. Active contours without edges. *IEEE Trans. on Image Proc.*, 10(2):266–277, 2001.

[2] S. Chen, T. Zhao, G. Gordon, and R. Murphy. A novel graphical model approach to segmenting cell images. In *Proc. IEEE Symp. on Comp. Intelligence in Bioinformatics and Computational Biology*, pages 1–8, 2006.

[3] A. Dufour, V. Meas-Yedid, A. Grassart, and J. Olivo-Marin. Automated quantification of cell endocytosis using active contours and wavelets. In *Proc. Int. Conf. on Pattern Recognition*, 2008.

[4] O. Dzyubachyk, W. Niessen, and E. Meijering. Advanced level-set based multiple-cell segmentation and tracking in time-lapse fluorescence microscopy images. In *IEEE International Symposium on Biomedical Imaging*, pages 185–188. IEEE, Piscataway, 2008.

[5] A. Eulalio, I. Behm-Ansmant, and E. Izaurralde. P bodies: at the crossroads of post-transcriptional pathways. *Nature Reviews Molecular Cell Biology*, 8(1):9–22, 2007.

[6] O. Greß, B. Möller, N. Stöhr, S. Hüttelmaier, and S. Posch. Scale-adaptive wavelet-based particle detection in microscopy images. In *Proceedings of the Workshop "Bildverarbeitung für die Medizin" (BVM)*, pages 266–270. Springer, 2010.

[7] M. Kass, A. P. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.

[8] L. Opitz, A. Schliep, and S. Posch. Joint analysis of in-situ hybridization and gene expression data. In *Advances in Data Analysis: Proc. of the 30th Ann. Conf. of the Gesellschaft für Klassifikation e.V.*, pages 577–584. Springer, 2007.

[9] H. Shen, G. Nelson, S. Kennedy, D. Nelson, J. Johnson, D. Spiller, M. White, and D. Kell. Automatic tracking of biological cells and compartments using particle filters and active contours. *Chemometrics and Intelligent Laboratory Systems*, 82(1-2):276–282, 2006.

[10] J. P. White, A. M. Cardenas, W. E. Marissen, and R. E. Lloyd. Inhibition of cytoplasmic mRNA stress granule formation by a viral proteinase. *Cell Host Microbe*, 2:295–305, 2007.

[11] S. Yamasaki and P. Anderson. Reprogramming mRNA translation during stress. *Curr. Opin. Cell Biol.*, 20:222–226, 2008.

[12] C. Zimmer and J. Olivo-Marin. Coupled parametric active contours. *IEEE PAMI*, 27(11):1838–1842, 2005.

Extraction of protein profiles from primary neurons using active contour models and wavelets

Danny Misiak [a,*], Stefan Posch [b], Marcell Lederer [a], Claudia Reinke [a], Stefan Hüttelmaier [a], Birgit Möller [b]

[a] Institute of Molecular Medicine, Martin Luther University Halle-Wittenberg, Heinrich-Damerow-Str. 1, 06120 Halle, Germany
[b] Institute of Computer Science, Martin Luther University Halle-Wittenberg, Von-Seckendorff-Platz 1, 06099 Halle, Germany

HIGHLIGHTS

- Extraction of neuron areas using active contours with a new distance based energy.
- Location of structural neuron parts by a wavelet based approach.
- Automatic extraction of spatial and quantitative data about distributions of proteins.
- Extraction of various morphological parameters, in a fully automated manner.
- Extraction of a distinctive profile for the Zipcode binding protein (ZBP1/IGF2BP1).

ARTICLE INFO

ABSTRACT

The function of complex networks in the nervous system relies on the proper formation of neuronal contacts and their remodeling. To decipher the molecular mechanisms underlying these processes, it is essential to establish unbiased automated tools allowing the correlation of neurite morphology and the subcellular distribution of molecules by quantitative means.

We developed NeuronAnalyzer2D, a plugin for ImageJ, which allows the extraction of neuronal cell morphologies from two dimensional high resolution images, and in particular their correlation with protein profiles determined by indirect immunostaining of primary neurons. The prominent feature of our approach is the ability to extract subcellular distributions of distinct biomolecules along neurites. To extract the complete areas of neurons, required for this analysis, we employ active contours with a new distance based energy. For locating the structural parts of neurons and various morphological parameters we adopt a wavelet based approach. The presented approach is able to extract distinctive profiles of several proteins and reports detailed morphology measurements on neurites.

We compare the detected neurons from NeuronAnalyzer2D with those obtained by NeuriteTracer and Vaa3D-Neuron, two popular tools for automatic neurite tracing. The distinctive profiles extracted for several proteins, for example, of the mRNA binding protein ZBP1, and a comparative evaluation of the neuron segmentation results proves the high quality of the quantitative data and proves its practical utility for biomedical analyses.

1. Introduction

Neurons are essential components of higher organisms. They form simple up to extraordinary complex networks via their axonal and dendritic cell extensions. During early development, neurons extend elongated protrusions of cytoplasm, called neurites, which finally differentiate into functional axons or dendrites. The morphology of single neurites essentially facilitates the function of neuronal networks, and defects in neurite architecture frequently correlate with severe brain disorders and neurological defects. Thus, the analysis of neurite morphology and underlying molecular

* Corresponding author at: Institute of Molecular Medicine, Sect. for Molecular Cell Biology, Martin Luther University Halle-Wittenberg, Heinrich-Damerow-Str. 1, 06120 Halle, Germany. Phone: +49 345 5522862; fax: +49 345 5522894.
E-mail address: danny.misiak@medizin.uni-halle.de (D. Misiak).

regulatory mechanisms is important to understand the basis of both neuronal function and neurological diseases.

During development or in response to environmental constraints, neurons like most cells modulate their gene expression program resulting in the synthesis and subsequent subcellular sorting of mRNAs or the final product, the proteins. These underlying regulatory mechanisms modulate intrinsic cell functions, control cell migration, and direct the morphology of cellular protrusions including neurites. To understand the function of these regulatory processes, it is essential to reveal the localization of mRNAs and proteins in relation to cell morphology by quantitative means. To this end, fully automated quantitative evaluation of molecular profiles remains a major obstacle.

A fully automated quantitative evaluation of molecular profiles requires an unbiased assessment of images acquired by high-resolution fluorescence microscopy subsequent to the specific labeling of molecules by fluorescent dyes. Although these procedures are well established, the quantitative and high-throughput evaluation of acquired images remains rudimentary and currently the ability to extract and analyze the subcellular distributions of distinct biomolecules in a high-throughput manner does not exist.

The extraction of distinctive molecular profiles based on fluorescence intensities requires an exact segmentation of neurons. The vast majority of published neurite detection approaches are limited to tracing the centerlines of neurites and evaluating the global cell morphology, without a complete segmentation of cell areas. Moreover, various approaches require manual or semiautomatic segmentation of neurites and, thus, still require *manual* intervention.

The goal of our work is to spatially and quantitatively assess distinct biomolecules along neurites. To achieve this, molecular profiles are extracted upon fluorescence labeling of relevant molecules, for example, proteins. Additionally, morphological and biomolecular changes in neurons induced by stress or influence of molecular stimuli are to be characterized in a fully automated manner by correlating the neuronal cell morphology with the extracted protein profiles. This requires an automatic extraction of complete neuron and neurite areas, as well as an automatic extraction of protein profiles along the identified neurites. To detect the neuron areas, we apply energy based active contour models (Kass et al., 1988), where we propose a new distance based energy to detect low contrast object boundaries. To identify the neurites within the detected neuronal cells the individual structural parts of the neurons are located. In general a developing (primary) neuron can be divided into soma (cell body) and neurites, where a complete neurite consists of two structural parts, the neurite shaft and the growth cone. For the identification of neurites a wavelet based approach is used (Mallat, 1998), derived from an algorithm to detect features in mass spectrometry data (Tautenhahn et al., 2008). It relies on the morphology of the cell region, in particular the width of structural neuron parts. The wavelet based approach separates the neuron into the three types of structures: soma, shafts and growth cones. Subsequently, the specific labeling of biomolecules is used to extract the intensity distributions along the identified neurites enabling a spatial and quantitative assessment of the proteins by the extracted profiles.

The algorithm is implemented in the `NeuriteAnalyzer2D` plugin for the Java image processing software *ImageJ*[1], and is available in the open-source package *MiToBo*[2]. The plugin is capable of batch analysis of a large number of 2D images enabling the quantitative high-throughput evaluation of acquired images.

`NeuriteAnalyzer2D` is able to extract distinctive profiles of several proteins, demonstrating its ability to extract fluorescence intensity distributions of labeled molecules along the localized neurites. In particular, a distinctive profile for the Zipcode binding protein (ZBP1/IGF2BP1) was extracted for the first time. In addition to basic measurements of neuronal morphology, like centerlines, end points or branching patterns, our approach reports detailed morphology measurements on neurites. For example, average neurite width, number of filopodia-like protrusions (Mattila and Lappalainen, 2008) and size of growth cone areas are extracted. The presented method to extract profiles of labeled molecules is validated by comparing the profiles of F-actin and α-tubulin to their published distributions in neuronal cells. To validate our fully automated extraction of neuron areas we compare the detected neurons from our approach and two other automatic tracing approaches to neurons manually analyzed by a biomedical expert.

2. Related work

The segmentation of elongated tube-like structures, for example, blood vessels (Kirbas and Quek, 2004; Läthén et al., 2010), plant roots (Erz et al., 2005) or neuronal cells (Capowski, 1983; Ramm et al., 2003), has been a challenge in computer vision for a long period of time. Efforts towards computer-aided segmentation of neurons and the analysis of neuronal morphology reach back 45 years (Glaser and Van Der Loos, 1965). Numerous findings in the fields of neurite tracing, quantitative methods of analysis and morphology extraction have been investigated.

However, to the authors' knowledge, no tools for spatial and quantitative assessment of distinct biomolecules along neurites are currently available to correlate biomolecular and morphological changes in neurons in a fully automated manner.

Schmitz et al. present an automatic image analysis program called *SynD* (Synapse Detector) to calculate the synaptic recruitment of proteins of interest (Schmitz et al., 2011). The calculation of the recruitment is only given as the ratio between synaptic and somatic intensity. Furthermore, the program reports dendritic and synaptic characteristics, like dendrite length as well as number of branches and synapses. To measure the dendritic branching the *Sholl analysis* is applied (Sholl, 1953). This analysis is widely used but offers a limited sensitivity to detect differences between groups of neurons and disregards orientation as well as topology of the dendritic tree (Uylings and van Pelt, 2002).

Since the extraction of neuronal morphology is still a very challenging task (Meijering, 2010), various approaches have been published aiming at the morphology extraction of neuronal cells, but disregarding a quantitative evaluation of molecular profiles. In most cases, trace lines of neurites are extracted and overall cell morphology is evaluated, while numerous algorithms do not overcome the need of manual intervention, which is a major bottleneck in efficient high-content image analysis. Even if the subcellular distribution of distinct biomolecules is disregarded, a few approaches should be mentioned, as the detection of neurons is important to extract protein profiles along neurites.

NeuronJ[3] (Meijering et al., 2004) and the *Simple Neurite Tracer*[4] are two popular tools, which provide semiautomatic tracing and quantification of neurites in 2D or 3D images, initialized by a manual setting of seed points. Zhang et al. (2007) present a novel tracing algorithm with automatic seed point detection. Without any user interaction the soma areas can be detected and centerline points of neurites are estimated. However, the complete neurite areas are not located. Most tools aim to extract neurite morphology, such as

[1] http://rsbweb.nih.gov/ij/
[2] http://www.informatik.uni-halle.de/mitobo/
[3] http://www.imagescience.org/meijering/software/neuronj/
[4] http://fiji.sc/wiki/index.php/Simple_Neurite_Tracer

data about length of neurites, number of branches and spines (small membranous protrusions), for example, Dehmelt et al. (2011), Ho et al. (2011), Meijering et al. (2004) and Pool et al. (2008). Only a few tools offer further measurements of neurons and neurites, such as the number of branching layers (layers of arborizations from primary neurites into secondary or tertiary branches), spine and growth cone shape or intensity statistics (maximum, mean and integrated intensity) (Wang et al., 2010; Xu and Wong, 2006).

The DIADEM Challenge[5] (2009–2010) plays an important role in reconstructing three dimensional neurons. The main goal of the competition was to devise algorithmic methods for automated neuronal tracing and digital reconstructions of neurons in 3D. Final algorithms work well on 3D images (Wang et al., 2011; Chothani et al., 2011; Zhao et al., 2011; Türetken et al., 2011; Bas and Erdogmus, 2011), but do not work properly with 2D images without workarounds, for example, creating artificial 3D images from 2D data.

Another promising tool for neuron tracing and reconstruction is Vaa3D-Neuron[6] (Peng et al., 2010). The tool is designed for 3D images, but can also handle 2D data. In addition to manual and semi-automatic tracing algorithms, Vaa3D-Neuron contains automatic methods to trace the structure of a neuron (Peng et al., 2011). As a result, traces of the neuron structure and the complete area of the neuron can be obtained. However, the tool does not perform well with our 2D image data, especially at overexposed soma regions.

Despite the diversity of tools, most common approaches evaluate overall cell morphology, without analyzing the spatial arrangement of distinct biomolecules within the cells. Thus we conclude that fully automated algorithms allowing the spatial and quantitative assessment of molecular profiles and neurite morphologies in a correlative manner are currently not available.

3. Biological background

The asymmetric distribution of protein factors in polarized cells like neurons are essential to define and maintain cellular polarity and polarized functions (Gavis et al., 2007). In matured neuronal systems as well as developing neurons, the subcellular sorting of proteins is often accomplished by the spatial arrangement of corresponding mRNAs and their subsequent local translation (Mikl et al., 2010; Tübing et al., 2010). This regulatory constraint is essential to prevent promiscuous protein activity and, thus, restricts protein function to specific subcellular sites. Thereby protein function or assembly of multi-protein complexes can be guided, for example, to synapses in matured neurons or growth cones of developing neurites. This allows for a spatiotemporal fine-tuning of gene expression governing essential neuronal functions like growth-cone guidance during embryogenesis as well as synaptic function in matured neurons, for instance during memory consolidation. One of the most prominent examples for such a regulatory mechanism in the neuronal system is the spatially restricted translation of the β-actin (ACTB) mRNA in developing neurons (Mikl et al., 2011). Transport and translational silencing of this mRNA are essentially facilitated by members of the Zipcode binding proteins. For ZBP1 it was shown that it associates with a specific region of 54 nucleotides in the ACTB-3′UTR whereby it inhibits ACTB protein synthesis and allows directed transport of this RNA to the growth cone of developing neurons (Ross et al., 1997; Zhang et al., 2001). Inhibition of ACTB mRNA translation by ZBP1 is abrogated upon phosphorylation of the protein by the protein-tyrosine kinase Src. This activation of ACTB protein synthesis is presumed to allow a

spatially restricted increase in ACTB monomer concentration which promotes F-actin (filamentous actin) polymerization, the driving force of cell protrusion (Ananthakrishnan and Ehrlicher, 2007). To analyze these processes it is essential to have high-resolution fluorescence microscopy and labeling techniques based on fluorescently labeled antibodies directed against specific proteins or fluorescently labeled antisense nucleotide probes hybridizing to endogenous mRNAs.

4. Workflow

The four phases of our approach are illustrated in the data flow diagram in Fig. 1. Initially a multichannel microscopy image is given, where the nuclei and different proteins or neuronal structures of interest are labeled. To yield a suitable basis for neuron segmentation, a maximum intensity projection (MIP) of the multichannel input image, except the nuclei channel, is computed, representing a 2D projection of the neurons. Due to low contrast of neuron boundaries, the extraction of the neuron contour is a challenging task. To solve such problems, several approaches exist, which require an initialization near the target objects. For neuron segmentation, we applied an active contour model, which also uses an initialization in the neighborhood of the object. This initialization is obtained from the extraction of a coarse neuron contour within the first phase (Section 5.1). Subsequently, this contour is refined by an active contour model in the second phase to extract an accurate neuron region contour (Section 5.2). This extraction completes the automatic neuron detection.

Given this neuron cell region, during the third phase the neuron is first segmented into approximate soma and neurite regions. Further analysis with a wavelet based approach localizes the origin of each neurite and detects the different structural parts of the neuron with higher accuracy (Section 6.1). This allows the categorization of the neuronal structure into soma, neurite shafts and growth cones. Subsequently, in the fourth phase the protein profiles are extracted, based on the intensity distributions of the fluorescence labeled proteins along the previously detected neurites (Section 6.2).

5. Neuron detection

The automatic detection of neurons is accomplished in the first two phases (cf. Fig. 1). The MIP contains at least one complete neuron, but also further neurons or fragments of not completely visible neurons can be present. Neurons in the image are localized by binarization and their regions are identified by component labeling. Subsequently, active contours are applied to improve the contours of these neurons.

5.1. Coarse neuron contour

The goal of the first phase is to localize the nuclei and neuron regions, aiming at the extraction of coarse contours of structural intact neuron cells. In our application, fixed cells are analyzed. Thus, structurally intact as well as dead cells are present in the fluorescence images. To exclude the later ones, we exploit that a dead cell does not contain an intact nucleus.

The fluorescently labeled nuclei are first detected in the nucleus channel by applying a global Otsu thresholding (Otsu, 1979). For neuron localization, the MIP image is binarized using the Niblack thresholding method, estimating mean and standard deviation in a neighborhood centered at (x, y) (Niblack, 1986). In initial studies, the complete image was identified as adequate neighborhood, rendering the Niblack thresholding as a global threshold operation. Following binarization, dead and not completely visible neurons are excluded by subsequently considering only neuron regions including at least 90% of the area of exactly one nucleus region.

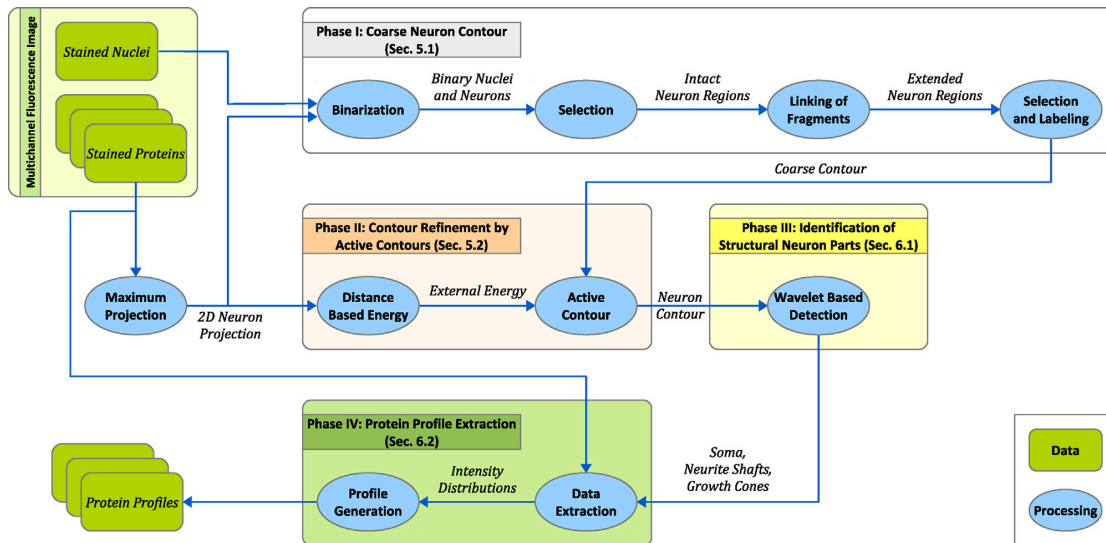
Fig. 1. Data flow diagram of the automatic profile extraction approach. The four phases are sequentially executed, resulting in a complete neuron segmentation and automatic protein profile extraction.

Neurite fragments of the Niblack binarization are linked to the located neuron regions, considering elongated regions as candidates for linking. Such candidates do not overlap with nuclei regions and are located within a maximum distance of the neuron (cf. Fig. 2). Subsequently, neurons that are considered as intact and complete are selected for further analysis. The contour of each neuron is extracted, which provides a coarse localization of the complete neuron region. This approximation is not sufficient to extract morphology features and an accurate protein quantification. However, it is well-suited as initialization of an *active contour model* and is to be improved in the second phase.

5.2. Contour refinement by an active contour model

Active contours are deformable models, influenced by external constraint and image forces. The basic idea is to match a deformable model to an image by means of the iterative minimization of an energy functional. Such models are widely used, for example, for
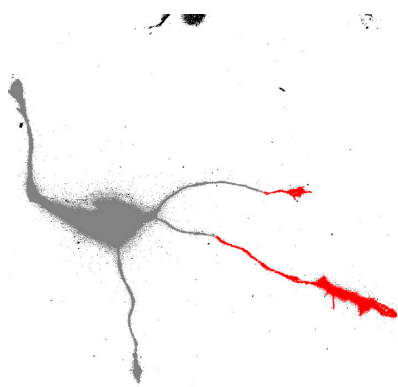
Fig. 2. Binarized image of a neuronal cell. Components not belonging to the neuron are shown in **black**, cell fragments due to binarization are shown in **red**. The identified candidate regions (**red**) are linked to the neuron region (**gray**) based on distance criteria.

image segmentation, edge detection, and object tracking (Liang et al., 2006; Shen et al., 2006; Unde, 2012). They can be represented as an explicit model, the traditional snakes (Kass et al., 1988), or implicitly using level set methods (Chan and Vese, 2001; Osher and Fedkiw, 2002). In our approach *snakes* are used, representing the contour as a parametric curve $v(s) = (x(s), y(s))$, normalized to unit length (cf. Kass et al., 1988). The energy functional, which has to be minimized, subsumes internal and external energy terms and controls snake motion within the image domain. The internal energy $E_{int}(v(s))$ considers, for example, length and curvature, to control smoothness of the snake. Whereas an external energy $E_{ext}(v(s))$ is defined over the image domain to guide the snake towards desired features such as object boundaries. In general, the energy functional of the snake is defined as sum of both energies,

$$E_{snake}(v(s)) = \int_0^1 E_{int}(v(s)) + E_{ext}(v(s)) \ ds \quad . \tag{1}$$

Using the calculus of variations, the energy functional (1) is minimized with regard to $v(s)$ by an iterative gradient descent method. A numerical solution can be derived by discretizing and subsequently solving the resulting system of linear equations (Kass et al., 1988).

External energies are frequently based on gradients of grayvalue images, like the negative gradient magnitude of a (Gaussian blurred) image. A major limitation of these energies is the small capture range and the poor convergence to boundary concavities (Xu and Prince, 1998). In previous work (Misiak et al., 2009), we used *gradient vector flow* (GVF) field as an alternative (Xu and Prince, 1998). This external force is based on the diffusion of gradient vectors of an edge map. It is able to guide the active contour to boundary concavities and allows for initialization far away from the object boundary. For tubular objects like neurites, however, the gradient is propagated from the boundary towards the center of the tube (Bauer and Bischof, 2008; Cabuk et al., 2010; Chang et al., 2003). Thus, the active contour converges within the neurites, due to the contraction of the left and right boundaries to a single line (cf. Fig. 3A). In consequence for our data we get incorrect results along neurite regions. To overcome this problem, we propose an energy
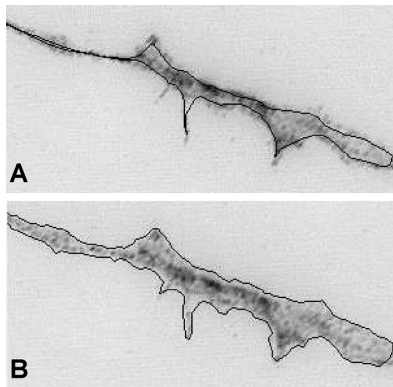
Fig. 3. Inverted image showing details of a detected neurite region using different external energies. (A) The gradient vector flow field as external energy partially causes the snake to move across the boundary inside the neurite and causes the snake to contract to a single line. (B) An accurate localization of the neurite boundary is achieved by using the distance based external energy.

based on a distance transformation that ensures a large capture range as well as a correct localization of the object boundary.

The purpose of the new energy is to define a map of distances by calculating the distance of each pixel in the image to all foreground regions. Hence, this energy term drives the snake along the distances of the transformation towards the object boundary. No smoothing of object boundaries occurs, but an exact localization of the contour is an indispensable prerequisite.

The MIP gray-value image provides the basis for the distance map calculation. The MIP image is first binarized as follows:

1. a median filter (5×5) is applied,
2. the gradient magnitude is computed (central differences),
3. a median filter (5×5) is applied to the gradient magnitude,
4. the filtered gradient magnitude image is binarized by Otsu thresholding,
5. and the binary regions are improved by morphological closing (5×5).

In contrast to the first phase, where we used Niblack thresholding for image binarization, here we apply an Otsu thresholding. The Niblack method generates a more complete neuron binarization with less fragmentation suitable to extract a coarse contour enclosing the entire neuron. In contrast to this, the Otsu thresholding results in more fragments of the neuron, but yields better localization of the neuron boundary which is important in this phase of neuron detection. Potential disconnections of neurites are compensated by the snake, because the neuron contour will be interpolated at these positions. Hence, Otsu gives an advantage over Niblack threshold to generate a more accurate external energy.

For the computation of the distance map the fast *Chamfer distance transformation* is used (Borgefors, 1986). Using an Euclidean distance metric, the distance transformation $D_{\mathrm{Euc}}(x, y) \in \mathbb{R}$ of a pixel (x, y) is defined as

$$D_{\mathrm{Euc}}(x, y) \quad = \quad \min_{(x', y') \in \mathrm{FG}(I)} \sqrt{(x - x')^2 \; + \; (y - y')^2} \; , \tag{2}$$

where FG(I) denotes the foreground of the binary image I, i.e. the neuron region. Inside the neuron region, all distance values are zero. Hence the snake is moved towards the neuron contour, but tends to not move across the boundary to the interior of the neuron region (cf. Fig. 3B).

The distance map calculated from the binary image and normalized to a range of [0, 1] can directly be used as external energy in the active contour model (1).

According to Kass et al. (1988), the internal energy $E_{\mathrm{int}}(v(s))$ subsumes sums over the first and second derivatives ($v'(s)$ and $v''(s)$) of the parametric contour $v(s)$, weighted by $\alpha(s)$ and $\beta(s)$. The first derivatives refer to the elasticity of the contour, where the second derivatives refer to the contour's rigidity. The weights $\alpha(s)$ and $\beta(s)$ rate elasticity against rigidity and balance the internal energy against the external energy $D_{\mathrm{Euc}}(v(s))$.

For all experiments we set $\alpha(s) \equiv \alpha : = 0.2$ and $\beta(s) \equiv \beta : = 0.0$. The step size γ in gradient descent is initially set to 0.05 for all snake control points. During minimization the value of γ is individually adapted for each point. The underlying function is defined as the square root of the external energy D_{Euc} at the current point position (x, y),

$$\gamma(x, y) \quad = \quad \sqrt{D_{\mathrm{Euc}}(x, y)}. \tag{3}$$

This adaptation of the step size γ makes the contour behave more dynamic in contrast to a fixed value of γ. Each control point is moved with an energy based step size. A large external energy value at the current control point corresponds to a larger distance to the neuron boundary. This results in a larger value of γ, whereas a small γ coincides with a small energy value. In addition, the contour is not moved across the boundary since the step size automatically decreases near the object boundary.

The minimization process terminates if less than 5% of the control points are moved in an iteration. In any case, the minimization process terminates after a maximum of 120 iterations. The optimization procedure completes the fully automatic detection of neurons.

6. Morphology and protein analysis

Quantitative assessment of protein distributions along neurites requires the extraction of intensity distributions of the stained proteins along the neurites. In the third phase, first structural parts of the neuron, in particular the neurites, are identified. Subsequently, in the fourth phase the intensity distributions along all neurites are extracted to obtain a distinctive profile of the protein analyzed.

6.1. Identification of structural neuron parts

In general, a neuron can be divided into two types of structural parts, the soma and the neurites. A complete neurite extends from its origin at the border of the somatic area along the neurite shaft to the tip of the growth cone (cf. Fig. 5A). To identify these structural parts, first a skeleton graph is extracted to represent the skeleton of the neuron region as a directed acyclic graph. Subsequently, the soma region is coarsely detected, neurite origins are roughly localized, and finally a wavelet based approach is used to exactly localize the neurites and detect the neurite parts, namely shaft and growth cone.

Based on the segmented neuron region (cf. Section 5), the skeleton of the neuron is extracted (Zhang and Suen, 1984) and a directed acyclic *skeleton graph* is generated, allowing a convenient handling of the skeleton. The nodes of the skeleton graph are defined by branch and end points of the skeleton, whereas the edges of the skeleton graph are defined as skeleton segments between such nodes, containing a sorted list of skeleton pixels. The edges are directed to point from a start point towards end points, which is achieved during construction of the graph. In addition, the soma of the segmented neuron region is coarsely localized by morphological opening. Fig. 4 shows the segmented soma region colored in gray within the black segmented neuron region. A large structuring
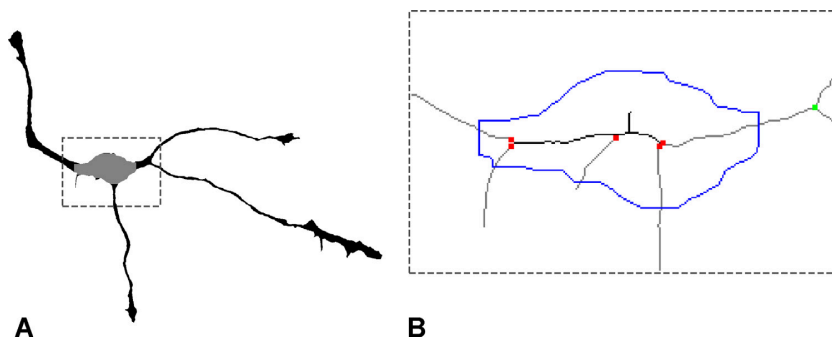
Fig. 4. Identification of soma and neurites as structural neuron parts. (A) The putative soma region of the detected neuron is coarsely localized by morphological opening with a large structuring element to eliminate non-soma regions (**black**) and leave a putatively coarse soma region (**gray**). (B) Zoom into the putative soma region, including the skeleton graph of the neuron region. The edges of the skeleton graph are marked as **gray** or **black lines**. The soma is marked with a **blue polygon**. Edges intersecting the soma boundary (**gray**) are used to split the graph into subgraphs, while edges inside the soma (**black**) are removed. The **red nodes** define the start nodes for each of the neurite skeleton subgraphs. These neurite skeleton subgraphs represent the individual neurites of the neuron and give the basis for localizing the start and end points of each neurite. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

element of 21 × 21 pixels is applied to eliminate non-soma regions and leave a putatively coarse soma region (cf. box in Fig. 4A).

Soma and skeleton graph are used for further analysis. To localize the neurite origins, only graph edges are considered which intersect the soma boundary (cf. Fig. 4B). To this end, the skeleton graph is divided into multiple subgraphs by splitting the graph at branch nodes inside the soma and consider only subgraphs with

an edge intersecting the soma boundary (cf. red nodes in Fig. 4B). In the special case that no branch node is present, the single edge is split at the intersections of this edge with the soma's boundary. This results in a set of skeleton graphs representing the skeletons of the individual neurites (cf. Fig. 4B).

For exact localization of a neurite's start and end point, the border between soma and neurite origin as well as between neurite shaft and growth cone needs to be detected. Considering the structural parts of the neuron in our images, it is obvious that these differ in their morphology. The somatic area is given by a large compact region extended by elongated neurites, consisting of a thin neurite shaft and a thickened terminal end, the growth cone. Hence, to identify the different structural parts these morphological characteristics, in particular the region width, are exploited. We interpret the width of the neuron region along each skeleton graph as a 1D function and denote it as *width profile*. Fig. 5A shows a binarized neurite and the corresponding width profile. The slope of the profile is steep at the transitions between the structural parts. Accordingly, the width of the neuron region can be used to detect the borders between the different structural parts, in particular between neurite shaft and soma, and shaft and growth cone, and to finally identify the neurites.

To solve this task of automatically identifying the structural parts from the width profile, we have adopted a wavelet based approach. It is derived from *centWave*, an algorithm to detect features in mass spectrometry data (Tautenhahn et al., 2008). The algorithm aims to detect chromatographic peaks and determines their boundaries based on 2D signals given by LC/MS raw data (liquid chromatography coupled to mass spectrometry) (Tautenhahn et al., 2008; Kuhl et al., 2011; Neumann et al., 2013). The main aim of the centWave algorithm is to detect peaks of different width in a chromatographic signal. Such peaks are characterized by significant deflections in the signal. centWave detects the peaks by localizing the center and their boundaries.

Considering the deflections in our width profile, two significant peaks exist that characterize the soma and the growth cone. To localize their boundaries and separate the neurite shaft from the soma and growth cone we search for positions in the profile where the slope of the profile considerably changes. These positions are given by the boundaries of the peaks related to the soma and growth cone inside the width profile.

centWave is able to detect peaks of varying width by using a *Continuous Wavelet Transform* (CWT) on multiple scales. In general, wavelet transforms are used to analyze signal structures by decomposing the signal with dilated and translated wavelets.
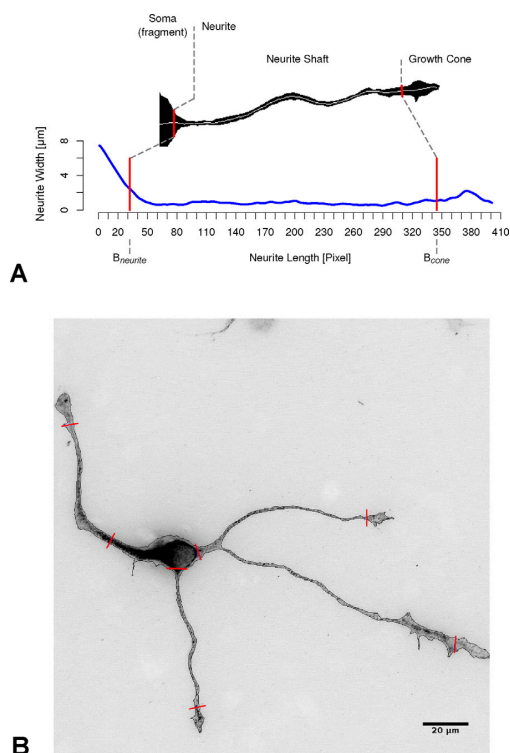

Fig. 5. (A) Binary image of a neurite (**black**) with the corresponding width profile (**blue**) along the neurite skeleton from the soma to growth cone. The results from the detection of the neuron parts are shown (**red**), separating the soma from the neurite and the neurite shaft from the growth cone. (B) Inverted image (adjusted brightness and contrast) showing a segmented neuron with located boundaries (**red**) between the different structural parts of the neuron. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The wavelet transform of a signal $f \in L^2(R)$ is then given by

$$W_f(\tau, s) = \int_{-\infty}^{\infty} f(t)\, \psi_{\tau,s}(t) \ dt, \quad s \in R^+, \quad \tau \in R \ , \tag{4}$$

where $\psi_{\tau,s}(t)$ is a family of normalized wavelets resulting from scaling and translating the mother wavelet $\psi(t)$ by s and τ (cf. Mallat, 1998). This transformation yields a matrix of wavelet coefficients W_f, for discrete values of τ and s. A *Mexican Hat* wavelet (normalized negated second derivative of a Gaussian) is used as mother wavelet, representing the model peak of the centWave algorithm. In the following the detection of the boundaries between the structural parts of the neuron is explained in detail.

Only width profiles with a minimum length of 20 μm (approx. 105 pixel) are considered for detection. Smaller profiles characterize filopodia-like protrusions (small protrusions of the cell surface), which are not considered in the protein profile extraction. Each profile is smoothed by a one dimensional Gaussian ($\sigma_{smooth} = 2$). After smoothing, the width profile is split into two parts at the center to improve the baseline estimation in the subsequently applied centWave algorithm. Based on the local maxima of the wavelet coefficients at each scale the centroids of the peaks and their left and right boundaries are determined (Tautenhahn et al., 2008; Du et al., 2006). For every single width profile, a set of peaks is returned by centWave. Each peak detected in the profile represents a bulge along the neurite. We are interested in the boundaries of the left- and right-most peaks. The right boundary ($B_{neurite}$) of the left-most peak is used to identify the origin of the neurite, whereas the left boundary (B_{cone}) of the right-most peak is used to identify the transition from neurite shaft to growth cone (cf. Fig. 5A). The positions of the peak boundaries are calculated as

$$B_{neurite} = C_{left} + 1.5 * \sigma \ , \tag{5}$$

$$B_{cone} = C_{right} - 1.5 * \sigma \ , \tag{6}$$

where C_{left} and C_{right} are the centers of the left- and right-most peaks, and σ is the parameter of the Gaussian fit to these peaks. A value of $\pm 1.5 * \sigma$ for boundary positions was identified as suitable for our application.

Based on these boundaries, the neuron is split into its structural parts. Fig. 5B shows a segmented neuron with the located boundaries, splitting the neuron into a soma region and three neurites with their neurite shafts and growth cones. The neurite to the right branches into two neurite arbors. The structural parts yield the base for extracting the protein profiles along the neurite shafts and growth cones in the fourth phase.

6.2. Protein profile extraction

The profile of an analyzed protein is quantitatively assessed by the extraction of the fluorescence intensity along the localized neurites (cf. Section 6.1). To extract the intensity distributions, each pixel of the neurite region is linked to the closest skeleton pixel using a Euclidean distance metric, which is equivalent to a Voronoi tesselation of the neurite region. The centers of the Voronoi cells are given by the skeleton points of the neurite skeleton. Based on the average fluorescence intensity of each Voronoi cell the intensity distribution of the protein along the neurite is extracted. Finally, a set of intensity distributions per protein is obtained by performing this analysis for all neurites. All distributions are normalized to unit length, to account for the different lengths of the neurites. Normalization is done by resampling the neurite with a fixed number of 100 sampling segments and summing up the intensities within each segment (integrated intensity). Subsequently, the median for each segment is calculated across all profiles for a specific protein, yielding a median intensity distribution for each protein to remove variations and extract one distinctive distribution for each protein.
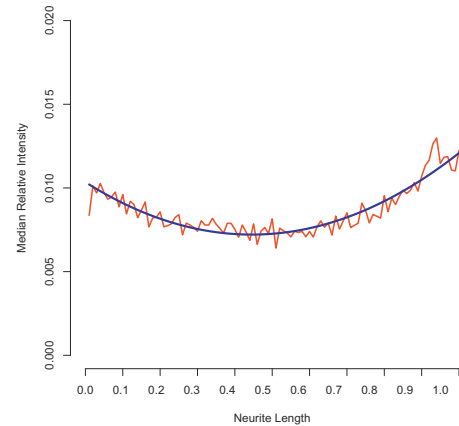

Fig. 6. Approximation of the protein profile. The median intensity profile (**red**) is calculated from the set of intensity distributions. Fitting a polynomial to the median intensity profile results in the final protein profile (**blue**) of the analyzed protein. The *x*-axis represents the length of the neurite from its origin to the tip of the growth cone normalized to range [0, 1]. The *y*-axis represents the median relative fluorescence intensity in a range of [0, 1], with regard to the integrated intensity (sum of neurite pixel intensities) at each length position. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In general the median intensity profiles are noisy. To get smooth *protein profiles* a polynomial is fitted to the median intensity profiles (cf. Fig. 6), using the *Nonlinear Least Squares* (NLS) (Bates and Watts, 1988) method, implemented in R[7]. Second order polynomials were identified as suitable for the median intensity profiles, while higher order polynomials tend to overfitting.

7. Results and discussion

7.1. Data

The neurons analyzed were extracted from the hippocampus of E17 mouse embryo brains. Subsequently, they were cultivated in culture medium for two days and then paraformaldehyde (PFA) fixed and immunohistochemically stained. The nuclei of the neuronal cells were labeled with a 4′,6-diamidino-2-phenylindole (DAPI) blue-fluorescent stain, that binds to DNA molecules (Kapuscinski, 1995). Furthermore, α-tubulin protein was stained to label the bundled microtubules that are localized in the neurite shaft and in addition ZBP1 (Zipcode Binding Protein 1) was stained. Both proteins were labeled via specific antibodies. To label the growth cones with their filopodia (cell surface protrusion with bundled actin filaments in its core), filamen-tous actin (F-actin) was stained with phalloidin.

Images were acquired with an epifluorescence microscope (Nikon TE2000-E) using a $60 \times$ oil objective (NA 1.4). The size of the images is 1000×1000 pixels, corresponding to a physical image size of 191 μm $\times$ 191 μm with a resolution of 0.191 μm $\times$ 0.191 μm per pixel. An image consists of four channels with an effective depth of 12 bit per pixel and channel. In total a set of 60 images was analyzed, including one or two cells and fragments of not completely visible cells.
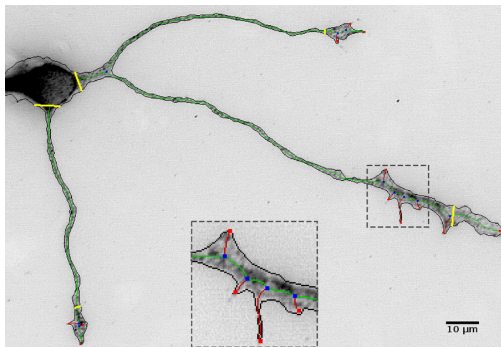
---

[7] http://www.r-project.org/

Fig. 7. Inverted gray-value MIP image. Bright pixels indicate low fluorescence intensities, dark pixels indicate high fluorescence intensities in the original image. The clip shows the contour of the neuron (**black**) as well as the boundaries (**yellow**) between the structural neuron parts (soma, neurite shafts, growth cones). The neurite traces (**green**), branch points (**blue dots**), end points (**red dots**) and filopodia-like protrusions (**red**) are also shown.

7.2. Neuron detection results

The final detection results for each image consist of the complete neuron contour from the active contour model (cf. Section 5.2), the neurite regions and their structural parts as well as the neurite trace lines (centerline along the skeleton of the neurite). Fig. 7 shows a clip of the MIP from a multichannel fluorescence image, inverted for better visualization. The segmented neuron is shown with its contour in black, the soma and two neurites, where the right neurite branches into two neurite arbors. The neurite branches have a length of 88.24 μm (462 pixel), 127.40 μm (667 pixel), and 71.05 μm (372 pixel). Neurite trace lines are shown in green. Boundaries between the structural parts of the neuron are marked by yellow lines. Fig. 7 also shows ten branch (blue dots) and twelve end points (red dots) as well as nine filopodia-like protrusions (red) with length ≤ 10 μm (approx. 52 pixel). Additionally, the area of the soma (342.16 μm^2), average width of neurite branches (0.79, 1.12, 0.88 μm), growth cone areas (33.02, 68.73, 32.10 μm^2) and number of filopodia-like protrusions inside the growth cone regions (2, 0, 2) are obtained.

For assessing the detection accuracy the neuron areas are considered. Twenty from the total of 60 neurons were randomly selected and manually analyzed by a biomedical expert, yielding ground truth data. Subsequently, segmented neuron regions are compared to the ground truth data, and *recall* and *precision* rates are calculated.

In Fig. 8 the recall and precision rates are presented for all evaluated neurons. An average median recall of 0.76 is observed with a minimum of 0.45 and a maximum of 0.91. For precision rate we observe a median of 0.85 with a minimum of 0.74 and a maximum of 0.94. The interquartile ranges of the boxplots are small, indicating a small variance of recall and precision rates. In the recall plot, two outliers are shown. These outliers result from thin and weak neurites that were not correctly segmented in some parts by the algorithm (for explanation see below). However, compared to the manually labeled neuron areas, the overall quality of the detection by our algorithm is high. As an example, the clip of a segmentation result in Fig. 9 underlines that the detected regions coincide well with given ground truth data. It can be observed that some small filopodia-like protrusions and the lower right neurite are not correctly segmented. This is due to low contrast boundaries and small width of these regions. In addition, the segmented soma region is marginally larger compared to the expert labeling. This is induced
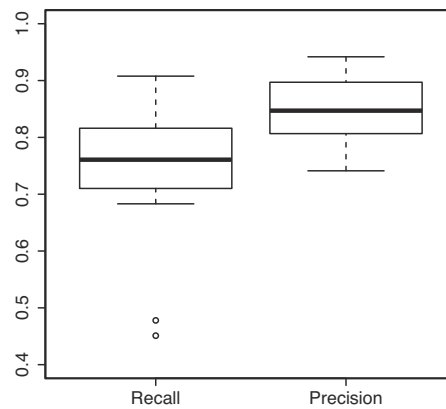

Fig. 8. Boxplots of recall and precision resulting from comparing twenty randomly selected neurons with ground truth data. A small variance of recall and precision rates is indicated by small interquartile ranges, illustrating the overall quality of the detection by our algorithm. In the recall plot, two outliers are shown which result from two neurons with thin and weak neurites.
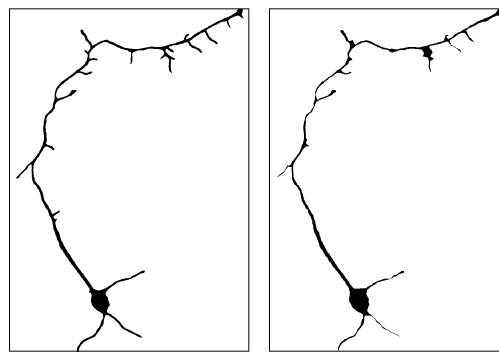

Fig. 9. Sample segmentation result for a detected neuron area. The figure compares expert labeling (left) and the automated segmentation result of our approach (right). Due to low contrast boundaries and small width of some regions, the lower right neurite and a few small filopodia-like protrusions are not correctly segmented. The segmented soma region is slightly larger compared to the expert labeling induced by perinuclear accumulation of proteins, overexposing the region of the soma. But in general, the detected regions coincide well with given ground truth data.

by perinuclear accumulation of proteins which overexposes the region of the soma.

To quantitatively evaluate the neuron segmentation of Neuron-Analyzer2D, we compared our results with those of NeuriteTracer (Pool et al., 2008) and Vaa3D-Neuron (Peng et al., 2010). Both are automatic tools for neurite tracing and produce complete areas of the detected neurons. The resulting neuron areas of all three tools are compared with ground truth data by using the *F-score*. This score is defined as the harmonic mean of precision and recall

$$F_{\text{score}} \;\; = \;\; 2\frac{\text{recall} \;\;\cdot\;\; \text{precision}}{\text{recall} \;\;+\;\; \text{precision}} \quad .$$

The best value for the *F*-score is reached at 1 and the worst value at 0.

NeuriteTracer is an ImageJ plugin for automated tracing of dissociated cultured neurons in two dimensional fluorescence microscopy images. It requires user defined thresholds for nuclei and neuronal stains as well as a range of admissible nuclear sizes. All thresholds and parameters should be reasonable to process multiple images. A nuclear threshold of 130 and a size range of [100, 4000] were identified as suitable. In contrast, it is not simple to
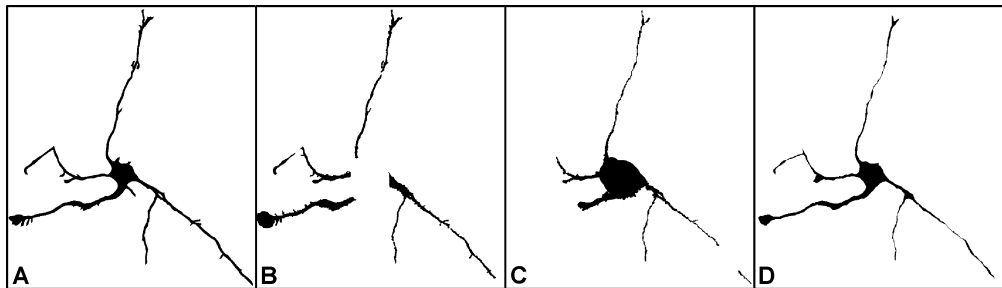
Fig. 10. Binary regions of neuron 21. (A) Manually labeled ground truth data. (B–D) Detection results of NeuriteTracer, Vaa3D-Neuron and NeuronAnalyzer2D.
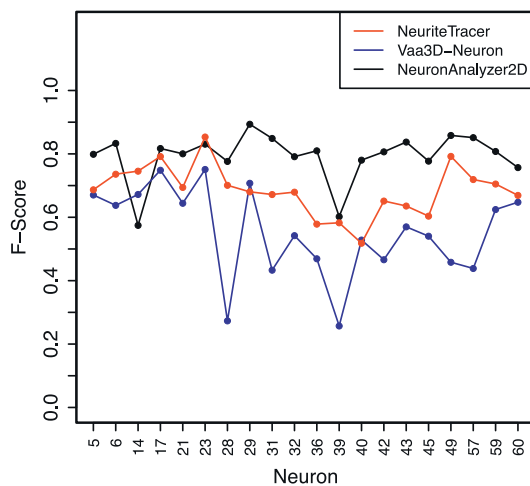

Fig. 11. *F*-scores of neuron areas obtained by comparing the manually labeled ground truth data of each neuron with the results of NeuriteTracer (**red**), Vaa3D-Neuron (**blue**) and the presented NeuronAnalyzer2D (**black**). The *F*-score is defined as the harmonic mean of precision and recall and demonstrates that the NeuronAnalyzer2D outperforms the other tools. The figure illustrates that most neurons are better detected by our approach, except neuron number 14, showing a thin neurite that exhibits a very weak fluorescence. Also the outliers from the recall boxplot in Fig. 8 are shown, represented by the small *F*-scores of neurons 14 and 39.

determine a suitable neuronal threshold for all images. To ensure a proper comparison we used different thresholds in range [40, 50, ..., 110] to detect the neurons in all test images. A value of 60 is chosen as neuronal threshold for NeuriteTracer, yielding the highest mean *F*-score (0.68) and smallest standard deviation (0.08) for all images.

The second tool is Vaa3D-Neuron, a module for neuron tracing and reconstruction, included in the Vaa3D[8] software suite for bioimage visualization and analysis. Upon request, we received a version[9] including the all-path pruning (APP) algorithm for automatic tracing and reconstruction (Peng et al., 2011). Parameters were set to default and auto-downsampling and auto-thresholding were enabled.

Figs. 10–12 show the results of the comparison of all three tools with ground truth data. In Fig. 11 the *F*-scores for each neuron area are shown for NeuriteTracer (red), Vaa3D-Neuron (blue) and the presented NeuronAnalyzer2D (black). These *F*-scores demonstrate
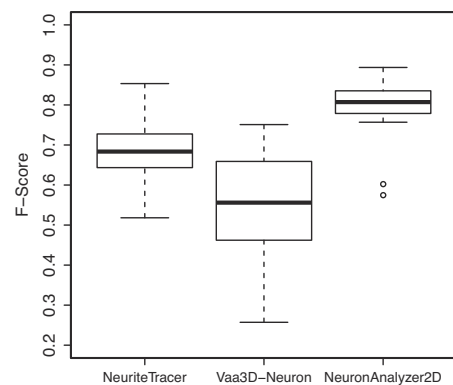
Fig. 12. Summary of the overall *F*-scores for each of the three tools from comparing twenty randomly selected neurons with ground truth data. A mean *F*-score of 0.81 is observed for our NeuriteAnalyzer2D, NeuriteTracer achieves a mean *F*-score of 0.68 and Vaa3D-Neuron-APP1 achieves a mean *F*-score of 0.56. The interquartile ranges of the boxplots of NeuriteTracer and NeuriteAnalyzer2D are small, which indicates a small variance of the scores, while a higher variance appears within the Vaa3D-Neuron scores. Outliers are plotted as small circles.

that on average our presented approach outperforms the other tools.

Compared to NeuronAnalyzer2D, NeuriteTracer detects neuron 14 better, whereas neurons 17, 23 and 39 are equally well detected. For the remaining neurons, NeuriteTracer's global image threshold for neuronal detection appears to underperform. In some cases, parts of the soma get lost, neurites are fragmented and often neuronal regions are thickened at the soma and neurites (cf. Fig. 10B).

Neuron number 14 is also better detected by Vaa3D-Neuron, compared to NeuronAnalyzer2D. However, in the detection of the remaining neurons it performs worse. Compared to ground truth data, the detected soma regions are often larger and neurites are fragmented or parts of them are lost (cf. Fig. 10C). A reason for this could be the characteristics of our image data, given by overexposed soma regions and two-dimensionality.

Fig. 12 shows boxplots of the *F*-scores for all three tools. For NeuriteTracer a mean *F*-score of 0.68 is observed, Vaa3D-Neuron-APP1 has a mean *F*-score of 0.56 and our NeuriteAnalyzer2D achieves a mean *F*-score of 0.81. The interquartile ranges of the NeuriteTracer and NeuriteAnalyzer2D boxplots are small, indicating a small variance of the scores, while a higher variance exists within the Vaa3D-Neuron scores.

As already mentioned above, on average the NeuronAnalyzer2D performs better than NeuriteTracer and Vaa3D-Neuron. An exception are the neurons 14 (cf. Fig. 13A) and 39 (cf. Fig. 13B), which are also apparent as outliers in the boxplot (cf. Fig. 12). A closer examination shows that these false detections occur on very long and thin
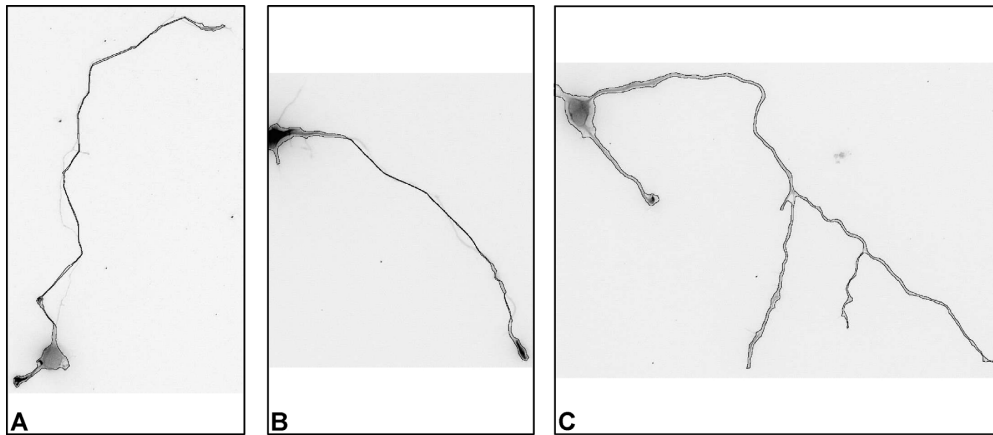
Fig. 13. Inverted gray-value image of detected neuron contours (**black**) by NeuronAnalyzer2D. (A, B) Outlier neurons 14 and 39 are shown. False detections along the neurites occur due to weak fluorescence stains. (C) Completely detected area of neuron 31.

neurites that exhibit a weak fluorescence. Such structures contain less stained proteins, which are also spread over a long distance. As a result, such structures show a weak fluorescence and frequently inhomogeneous stained regions (cf. Fig. 13A and B). Thus, certain areas of these neurites are better detected than others and some errors in detection occur. But in general, our NeuronAnalyzer2D is able to detect long and thin neurites as shown in Figs. 9 and 13C.

7.3. Profile extraction results

In Fig. 14 boundaries separating the different structural parts of a neuron are shown. These boundaries are obtained from the extended wavelet based approach (cf. Section 6.1). All images show a fragment of the soma on the left side and two boundaries (red lines), separating the soma from the neurite and the neurite shaft from the growth cone. It can be seen that the soma and neurite separation works well (cf. Fig. 14A–C), whereas the separation of
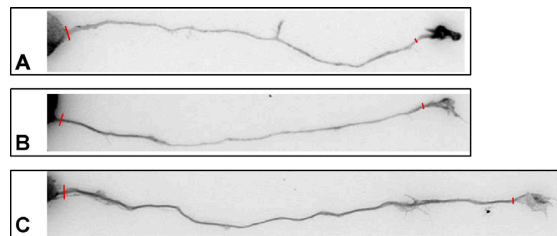
Fig. 14. Inverted gray-value images of neurites, cropped at the soma region. The boundaries (**red lines**) between the structural neuron parts are obtained applying the extended wavelet based approach. Elongated or misshaped growth cones of neurites lead to a premature cutting of neurite shafts (right boundary in A and B). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
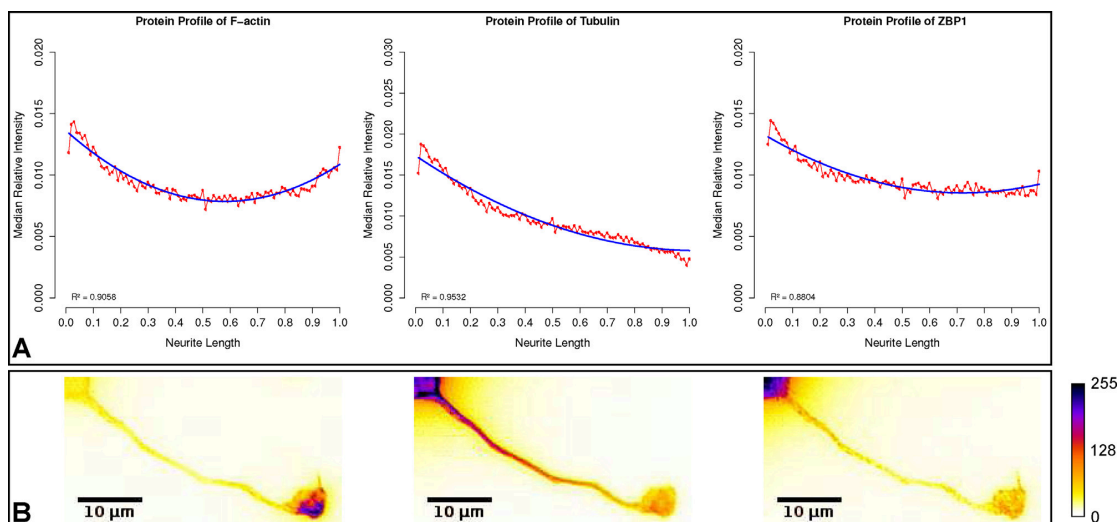
Fig. 15. (A) Extracted protein profiles for F-actin, α-tubulin, and ZBP1 (from left to right) using the present approach. The **red lines** show the median intensity profile of the corresponding protein. The **blue lines** show the polynomial interpolation of the median intensity profile. (B) Inverted images of one sample neurite using a fire-color LUT. The images show the protein distribution along this neurite (from left to right: F-actin, α-tubulin, and ZBP1), according to the extracted profiles in (A). Pixels scaled to 0 indicate low fluorescence intensities, whereas pixels scaled to 255 indicate high fluorescence intensities.

neurite shaft and growth cone can vary. This mainly depends on the shape of the growth cone, whereby elongated or misshaped cones lead to a premature cutting of the neurite shaft (cf. Fig. 14A and B). Based on the located structural parts, protein profiles are extracted.

The profile extraction results in median intensity distributions and corresponding polynomial interpolations, yielding a distinct distribution for each protein analyzed. In Fig. 15A the extracted median intensity profiles and corresponding polynomial interpolations are shown for F-actin, α-tubulin, and ZBP1. To validate our extraction method, the profiles of F-actin and α-tubulin are compared to published distributions. From the literature, for example, Dent and Gertler (2003), Lowery and Vactor (2009) and Spooner and Holladay (1981), it is known that growth cone motility and protrusion depend on actin, in particular F-actin, dynamics. Therefore, F-actin is mainly located in the growth cone periphery. The amount of F-actin is large in the tip of the growth cone and low along the neurites. Microtubules are transported from the cell body into neurites for structuring and elongation of neurites. Consequently, α-tubulin is enriched in neurite shafts, whereas α-tubulin is rarely observed in the growth cone periphery.

As can be seen from Fig. 15 the extracted profiles of F-actin (Fig. 15A, left) and α-tubulin (Fig. 15A, middle) correspond well with these experimental findings. The concentration of F-actin protein decreases along the neurites, compared to the soma, while F-actin is enriched in the growth cones. Opposed to this, α-tubulin concentration decreases from the soma along the neurites to the growth cones, and only a small amount of α-tubulin is located inside the growth cones. These results underline the suitability of our approach for quantitative assessment of protein profiles.

For ZBP1, Fig. 15A (right) shows that it is mainly localized in the soma and growth cone regions. Along the neurites the amount of ZBP1 decreases, while the terminal amount of ZBP1 is smaller than the amount at the origin near the soma. The three images in Fig. 15B show the distributions of the stained proteins along one sample neurite in comparison to the extracted profiles in Fig. 15A.

8. Conclusion

This paper presents a new integrated approach for the fully automated extraction of protein profiles along automatically extracted neurites in fluorescence microscopy images. Thus, our approach has the unique feature that allows one to assess the distribution of labeled proteins *inside* neurites in a spatial and quantitative manner that allows conclusions regarding relationships between the spatial arrangement of proteins together with morphological changes of the cells.

The approach uses active contour models for segmentation of neurons and performs a wavelet based identification of structural neuron parts. It has proven successful in neuron detection as well as in the extraction of protein profiles along neurites. A distinctive profile of ZBP1 was acquired that shows, for the first time, how the protein is distributed in primary hippocampal neurons.

In general, the analysis of neuronal morphology together with quantitative measurements of neurite associated proteins will allow new findings in biomedical research. Beyond that, our approach can be used to characterize morphological and biomolecular changes in neurons due to stress or influence of other molecular stimuli. The detection of structural neuron parts offers new options for investigating the impact of proteins on morphological changes in neurite or growth cone shapes.

In contrast to other tools, our approach offers important novel features in the field of morphology and structure analysis for biomedical investigations. Fully automated analysis of a large number of images enables high-content analysis, while the tool can easily be used by scientists without prior image processing experience. A small number of configuration parameters permits an unbiased and rapid quantification.

In the future, we aim to adapt this approach for growth cone motility analysis with regard to neurite turning and growth cone guidance (Gomez and Zheng, 2006; Mai et al., 2009). This would allow investigations regarding shape and orientation changes under defined conditions, for example, neuronal stimuli acting as attractive or repulsive signals. Furthermore, the approach will be extended for joined analysis of living cells to offer population and cell differentiation analysis over time, based on environmental constraints.

Appendix A. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jneumeth. 2013.12.009.

References

Ananthakrishnan R, Ehrlicher A. The forces behind cell movement. International Journal of Biological Sciences 2007];3:303–17.

Bas E, Erdogmus D. Principal curves as skeletons of tubular objects – locally characterizing the structures of axons. Neuroinformatics 2011];9:181–91.

Bates DM, Watts DG. Nonlinear regression analysis and its applications. Wiley series in probability and mathematical statistics. New York: Wiley; 1988].

Bauer C, Bischof H. A novel approach for detection of tubular objects and its application to medical image analysis. In: Rigoll G, editor. Proceedings of the pattern recognition, 30th DAGM symposium. Springer; 2008]. p. 163–72.

Borgefors G. Distance transformations in digital images. Computer Vision, Graphics, and Image Processing 1986];34:344–71.

Cabuk AD, Alpay E, Acar B. Detecting tubular structures via direct vector field singularity characterization. Conference Proceedings: IEEE Engineering in Medicine and Biology Society 2010];1:4801–4.

Capowski JJ. An automatic neuron reconstruction system. Journal of Neuroscience Methods 1983];8:353–64.

Chan TF, Vese LA. Active contours without edges. IEEE Transactions on Image Processing 2001];10:266–77.

Chang S, Metaxas DN, Axel L. Scan-conversion algorithm for ridge point detection on tubular objects. In: Ellis RE, Peters TM, editors. Medical image computing and computer-assisted intervention – MICCAI 2003, 6th international conference. Springer; 2003]. p. 158–65.

Chothani P, Mehta V, Stepanyants A. Automated tracing of neurites from light microscopy stacks of images. Neuroinformatics 2011];9:263–78.

Dehmelt L, Poplawski G, Hwang E, Halpain S. Neuritequant: an open source toolkit for high content screens of neuronal morphogenesis. BMC Neuroscience 2011];12:100.

Dent EW, Gertler FB. Cytoskeletal dynamics and transport in growth cone motility and axon guidance. Neuron 2003;40:209–27.

Du P, Kibbe WA, Lin SM. Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. Bioinformatics 2006;22:2059–65.

Erz G, Veste M, Anlauf H, Breckle SW, Posch S. A region and contour based technique for automatic detection of tomato roots in minirhizotron images. Journal of Applied Botany and Food Quality 2005;79:83–8.

Gavis ER, Singer R, Hüttelmaier S. Localized translation through messenger RNA localization. In: Mathews MB, Sonenberg N, Hershey JW, editors. Translational control in biology and medicine. Cold Spring Habor, New York: Cold Spring Harbor Laboratory Press; 2007. p. 687–717.

Glaser EM, Van Der Loos H. A semi-automatic computer-microscope for the analysis of neuronal morphology. IEEE Transactions on Biomedical Engineering 1965;12:22–31.

Gomez TM, Zheng JQ. The molecular basis for calcium-dependent axon pathfinding. Nat Rev Neurosci 2006;7:115–25.

Ho SY, Chao CY, Huang HL, Chiu TW, Charoenkwan P, Hwang E. NeurphologyJ: an automatic neuronal morphology quantification method and its application in pharmacological discovery. BMC Bioinformatics 2011;12:230.

Kapuscinski J. DAPI: a DNA-specific fluorescent probe. Biotechnic and Histochemistry 1995;70:220–33.

Kass M, Witkin AP, Terzopoulos D. Snakes: active contour models. International Journal of Computer Vision 1988;1:321–31.

Kirbas C, Quek FKH. A review of vessel extraction techniques and algorithms. ACM Computing Surveys 2004;36:81–121.

Kuhl C, Tautenhahn R, Böttcher C, Larson TR, Neumann S. CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. Analytical Chemistry 2011;84:283–9.

Läthén G, Jonasson J, Borga M. Blood vessel segmentation using multi-scale quadrature filtering. Pattern Recognition Letters 2010;31:762–7.

Liang J, McInerney T, Terzopoulos D. United snakes. Medical Image Analysis 2006;10:215–33.

Lowery LA, Vactor DV. The trip of the tip: understanding the growth cone machinery. Nature Reviews Molecular Cell Biology 2009;10:332–43.

Mai J, Fok L, Gao H, Zhang X, Poo MM. Axon initiation and growth cone turning on bound protein gradients. Journal of Neuroscience 2009;29:7450–8.

Mallat S. A wavelet tour of signal processing. San Diego: Academic Press; 1998.

Mattila PK, Lappalainen P. Filopodia: molecular architecture and cellular functions. Nature Reviews Molecular Cell Biology 2008;9:446–54.

Meijering E. Neuron tracing in perspective. Cytometry 2010;77:693–704.

Meijering E, Jacob M, Sarria JC, Steiner P, Hirling H, Unser M. Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images. Cytometry 2004;58:167–76.

Mikl M, Vendra G, Doyle M, Kiebler MA. RNA localization in neurite morphogenesis and synaptic regulation: current evidence and novel approaches. Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology 2010;196:321–34.

Mikl M, Vendra G, Kiebler MA. Independent localization of MAP2, CaMKIIα and β-actin RNAs in low copy numbers. EMBO Reports 2011;12:1077–84.

Misiak D, Posch S, Stöhr N, Hüttelmaier S, Möller B. Automatic analysis of fluorescence labeled neurites in microscope images. In: Workshop on applications of computer vision (WACV), 2009; 2009. p. 1–7.

Neumann S, Thum A, Böttcher C. Nearline acquisition and processing of liquid chromatography-tandem mass spectrometry data. Metabolomics 2013;9:84–91.

Niblack W. An introduction to digital image processing. Upper Saddle River, NJ: Prentice-Hall, Inc; 1986.

Osher SJ, Fedkiw RP. Level set methods and dynamic implicit surfaces. 1st ed. New York: Springer; 2002.

Otsu N. A threshold selection method from gray-level histograms. Trans on Systems, Man, and Cybernetics (SMC) 1979;9:62–6.

Peng H, Ruan Z, Long F, Simpson JH, Myers EW. V3d enables real-time 3d visualization and quantitative analysis of large-scale biological image data sets. Nature Biotechnology 2010;28:348–53.

Peng H, Long F, Myers G. Automatic 3d neuron tracing using all-path pruning. Bioinformatics [ISMB/ECCB] 2011;27:239–47.

Pool M, Thiemann J, Bar-Or A, Fournier AE. Neuritetracer: a novel ImageJ plugin for automated quantification of neurite outgrowth. Journal of Neuroscience Methods 2008;168:134–9.

Ramm P, Alexandrov Y, Cholewinski A, Cybuch Y, Nadon R, Soltys BJ. Automated screening of neurite outgrowth. Journal of Biomolecular Screening 2003;8:7–18.

Ross AF, Oleynikov Y, Kislauskis EH, Taneja KL, Singer RH. Characterization of a *beta*-actin mRNA zipcode-binding protein. Mol Cell Biol 1997;17:2158–65.

Schmitz SK, Hjorth JJ, Joemai RM, Wijntjes R, Eijgenraam S, de Bruijn P, et al. Automated analysis of neuronal morphology, synapse number and synaptic recruitment. Journal of Neuroscience Methods 2011;195:185–93.

Shen H, Nelson G, Kennedy S, Nelson D, Johnson J, Spiller D, et al. Automatic tracking of biological cells and compartments using particle filters and active contours. Chemometrics and Intelligent Laboratory Systems 2006;82:276–82.

Sholl DA. Dendritic organization in the neurons of the 19 visual and motor cortices of the cat. Journal of Anatomy 1953;87:387–406.

Spooner BS, Holladay CR. Distribution of tubulin and actin in neurites and growth cones of differentiating nerve cells. Cell Motility 1981;1:167–78.

Tautenhahn R, Böttcher C, Neumann S. Highly sensitive feature detection for high resolution LC/MS. BMC Bioinformatics 2008;9:504.

Tübing F, Vendra G, Mikl M, Macchi P, Thomas S, Kiebler MA. Dendritically localized transcripts are sorted into distinct ribonucleoprotein particles that display fast directional motility along dendrites of hippocampal neurons. Journal of Neuroscience 2010;30:4160–70.

Türetken E, González G, Blum C, Fua P. Automated reconstruction of dendritic and axonal trees by global optimization with geometric priors. Neuroinformatics 2011;9:279–302.

Unde A. A novel edge detection approach on active contour for tumor segmentation. In: 2012 students conference on engineering and systems (SCES); 2012. p. 1–6.

Uylings HBM, van Pelt J. Measures for quantifying dendritic arborizations. Network-Computation in Neural Systems 2002;13:397–414.

Wang D, Lagerstrom R, Sun C, Bishof L, Valotton P, Götte M. HCA-vision: automated neurite outgrowth analysis. Journal of Biomolecular Screening 2010;15:1165–70.

Wang Y, Narayanaswamy A, Tsai CL, Roysam B. A broadly applicable 3-d neuron tracing method based on open-curve snake. Neuroinformatics 2011;9:193–217.

Xu C, Prince JL. Snakes, shapes, and gradient vector flow. IEEE Transactions on Image Processing 1998;7:359–69.

Xu X, Wong S. Optical microscopic image processing of dendritic spines morphology. IEEE Signal Processing Magazine 2006;23:132–5.

Zhang TY, Suen CY. A fast parallel algorithm for thinning digital patterns. Communications of the ACM 1984;27:236–9.

Zhang HL, Eom T, Oleynikov Y, Shenoy SM, Liebelt DA, Dictenberg JB, et al. Neurotrophin-induced transport of a β-actin mRNP complex increases β-actin levels and stimulates growth cone motility. Neuron 2001;31:261–75.

Zhang Y, Zhou X, Degterev A, Lipinski M, Adjeroh D, Yuan J, et al. A novel tracing algorithm for high throughput imaging screening of neuron-based assays. Journal of Neuroscience Methods 2007;160:149–62.

Zhao T, Xie J, Amat F, Clack N, Ahammad P, Peng H, et al. Automated reconstruction of neuronal morphology based on local geometrical and global structural models. Neuroinformatics 2011;9:247–61.

Note on the following publication:

Analysis of Arabidopsis Root Images — Studies on CNNs and Skeleton-Based Root Topology

Birgit Möller, Berit Schreck, and Stefan Posch

Institute of Computer Science, Martin Luther University Halle-Wittenberg
Von-Seckendorff-Platz 1, 06099 Halle (Saale), Germany
`birgit.moeller@informatik.uni-halle.de`

Abstract

Roots and their temporal development play an important role in plant research. Over the decades image-based monitoring of root growth has become a key methodology in this research field. The growing amount of image data is often tackled with automatic image analysis approaches. In particular convolutional neural networks (CNNs) recently gained increasing interest for root segmentation. This segmentation of roots is usually only the first step of an analysis pipeline and needs to be supplemented by topological reconstruction of the complete root system architecture.

In this paper we present a comprehensive study of different CNN architectures, loss functions and parameter settings for root image segmentation. In addition, we show how main and lateral roots can be identified based on the skeletons of segmented root components as a first step towards topological reconstruction of root system architecture. We present quantitative and qualitative results on data released in the course of the CVPPA Arabidopsis Root Segmentation Challenge 2021.

1. Introduction

Roots are an important organ of plants. They play essential roles in ensuring secure anchorage and in uptaking vital nutrients from the soil. Hence, root physiology and status have a major impact on growth and development of plants and render them a key topic in plant research. Quantitative data about root mass and root system development over time provide a solid basis for understanding functional relationships between environmental conditions, plant development and the status of ecosystems as a whole [3]. As roots are the plant organs least accessible, in the beginning of root studies direct manual measurements from excavated roots dominated the collection of quantitative data.

Meanwhile image-based techniques are well established. For experiments in soil minirhizotrons are available which

Figure 1. Left, prototypical root image from one of the challenge data sets, right, corresponding ground truth annotation with background in black, lateral roots in white and main roots in gray.

allow for non-destructive acquisition of time series data of roots as basis for developmental studies [9]. Such image data does not provide enough information to reconstruct the architecture of the complete root system, but data is restricted to overall root length or mass in the sample area. As an alternative, growing plants in culture medium within transparent plates and regular acquisition of images from such plates (Fig. 1) has become a popular protocol for studying the development of complete root systems [1].

Automatic analysis of such data typically subsumes two analysis stages. In the first stage images are segmented to separate roots from the image background on the pixel level. Essentially this allows to estimate root mass. Root topology is extracted in the second stage where different levels are considered. Identification of the center pixels of the roots allow for length measurements. The reconstruction of complete root system architectures subsumes in addition explicit annotation of branching points, as well as distinction between the main root (MR) and lateral roots.

Over the decades the methodology for root image analysis has emerged from pure manual image annotation to fully automatic approaches. During the last years particularly deep learning with convolutional neural networks (CNNs) has proven suitable for solving the root image segmentation task. As all methods based on deep learning, such ap-

proaches strongly rely on a sufficiently large set of annotated training data which often forms a serious bottleneck.

Two data sets of annotated root images were released for the Arabidopsis Root Segmentation Challenge organized in conjunction with the 7th Workshop on Computer Vision in Plant Phenotyping and Agriculture held as part of the ICCV 2021 (for details see Sec. 4). The task of the challenge is to segment all roots from a given image and to identify the MR and all lateral roots of individual plants (Fig. 1).

In this paper we present a fully automatic approach for solving the challenge task. Root segmentation is performed by applying CNNs to individual images of a time series. We decided to treat the images of a time series independently, as for the challenge training data annotations are available only for few, non-consecutive images of each series. To subsequently identify the MRs in the segmentation results we rely on topological analysis of root skeletons extracted from the segmentation results. Our CNN model with which we successfully participated in the challenge comprises a U-Net architecture with VGG16 backbone and was trained by first applying a loss function combining dice loss and cross entropy, and then by fine-tuning with focal loss. The model was selected from pre-studies on the challenge data.

The main contributions of this paper are two-fold. On the one hand we not only discuss our challenge model and the results, but extend our pre-studies towards a comprehensive comparative overview of additional architectures and loss functions for the segmentation of Arabidopsis root images. On the other hand, we present our approach for post-processing the segmentation results towards proper reconstruction of the complete root system architecture (RSA), focusing on the challenge task of extracting the MRs. Root segmentation results and outcomes of MR extraction are comprehensively evaluated quantitatively as well as qualitatively on the challenge training and test data sets.

The remainder of this paper is organized as follows. In Sec. 2 we give an overview of related work before we present our studies and methods in Sec. 3. Details about the data sets are provided in Sec. 4, while experimental results are presented in Sec. 5. A conclusion is given in Sec. 6.

2. Related Work

Deep learning and CNNs became the prevailing technique in image analysis with the publication of AlexNet [11] for the task of object classification. In the following years they have been extended to the task of semantic segmentation with the proposal of FCN in [13] and encoder-decoder architectures with SegNet [2] and U-Net [18]. These architectures have been extended based on residual-blocks [7], inception modules [24], and the hour glass architecture [16], see also [5] for an early review.

Besides network architecture another important ingredient for good performance is training with an adequate loss

function. [8] discuss and evaluate several common ones like cross entropy and dice loss, and also less common ones like focal loss. These functions have the drawback to penalize false positives or false negatives without considering distances to the nearest annotated and the next predicted foreground pixel, respectively. The Weighted Hausdorff Distance (WHD) is proposed in [27] to overcome this problem.

The task of segmenting roots from images is often tackled with conventional segmentation techniques like intensity thresholding in BRAT [22] or ridge filtering in MyRoot [6], sometimes still relying on manual user intervention, e.g., for selecting appropriate threshold values like in EZ-Root-VIS [19]. Recently CNNs gained larger importance in this field. SegRoot [25] adopts the SegNet architecture and applies a dice loss function for extracting roots from minirhizotron images. [23] build upon the U-Net architecture with a loss combining cross entropy and the dice loss for the same task. RootNav 2.0 [28] is specifically designed for assay images using an encoder-decoder configuration integrating an hourglass network at the interface between encoder and decoder. PhenomNet [29] not only tackles the root segmentation task, but also integrates Recurrent Neural Networks based on Long Short-Term Memory to couple phenotypic predictions with genotypic analysis. In [4] a CNN based on U-Net employing residual blocks is proposed. It applies deep supervision of intermediate results and adds convolutional layers at the end of the U-Net core.

The separation of roots and background is only the first step in root image analysis. In many cases subsequent post-processing steps are applied, e.g., to close gaps and link segmented components which belong to the same root. In MyRoot 2.0 [6] a tracking algorithm is implemented which aims to link all fragments of a root between root tips and the hypocotyl based on distance heuristics. RootNav [17] adopts the A^* search algorithm to extract paths from root tips to seeds along lateral and MRs. While in the original paper [17] seeds and tips had to be selected manually by the user, in RootNav 2.0 [28] these are now automatically predicted by the CNN in parallel to potential root pixels.

3. Methods

3.1. Semantic Segmentation

For our studies we choose the basic encoder-decoder variants SegNet and U-Net due to their popularity especially in the life sciences and their use for the root segmentation task [23, 25]. They differ mainly in how they incorporate information from the encoder stage into the decoder. In Seg-Net, the positions of maximal values selected in the max-pooling operations are used in the corresponding upsampling step in the decoder to initialize the upsampled feature maps. The other values are filled with zeros and then interpolated with convolution. In contrast, U-Net upsamples

the last feature map in a resolution level with a trainable transpose convolution and concatenates the last feature map from the encoder with the same spatial resolution.

In addition we investigate hierarchical feature integration (Hi-Fi) proposed in [31] for the task of skeleton detection in the wild. This can be viewed as an extension of FCN [13] and Holistically-Nested Edge Detection (HED) [26]. The latter uses the last feature map of all resolution levels to compute multiple predictions, the so-called side outputs, which are subject to intermediate supervision. In addition they are fused to the final prediction. Hi-Fi proposes a richer way to incorporate the features of the encoder. First, not only the last, but all feature maps of each resolution level are fed into the side outputs. Second, the feature maps are not directly fused into prediction with a 1×1 kernel, but first convolved with 3×3 kernels into features maps as a basis for side output predictions. Third, features from neighboring resolution levels are combined, which results in Hi-Fi level 1. This combination of neighboring resolutions may be recursively repeated yielding further Hi-Fi levels and, as in HED, all side outputs are supervised. [31] advise to use one or two levels of the hierarchy. The tasks of edge or skeleton detection and semantic segmentation share common challenges. E.g., HED, proposed for edge detection, was applied to skeleton detection [10, 20, 31]. In previous studies we found HED suitable for root detection in minirhizotron images. As in addition roots exhibit strong symmetries we explore the potential of Hi-Fi for root segmentation.

A second focus in our study are loss functions as they optimize different characteristics of the segmentation. These are the cross entropy (CE) commonly used for semantic segmentation and the dice loss (DI) [14] as the inverse of the dice score which optimizes one of the evaluation metrics. In addition we combine DI with CE weighted by 0.3 (CombCED) as suggested in [23] to overcome a drawback of DI yielding a zero loss if no pixel is annotated as root.

We also use the Weighted Hausdorff Distance (WHD) defined in [27] as:

$$
\begin{aligned}
L_{\mathrm{WHD}} = \quad & \frac{1}{\left|\tilde{Y}_+\right| + \varepsilon} \sum_{b \in \Omega} p_f(b) \min_{a \in Y_+} d(b, a) \\
+ \quad & \frac{1}{|Y_+|} \sum_{a \in Y_+} \min_{b \in \Omega} \frac{d(a, b) + \varepsilon}{(p_f(b))^{\alpha} + \frac{\varepsilon}{d_{\max}}}, \quad (1)
\end{aligned}
$$

where Y_+ is the set of annotated foreground pixels, $p_f(b)$ the probability of pixel b to be predicted as foreground, Ω the image domain, $|\tilde{Y}_+| = \sum_{b \in \Omega} p_f(b)$, $d(a, b)$ the Euclidean distance between two pixels, α a weighting factor, and $d_{\max}$ the maximal distance between two pixels. The first term is a proxy to the average distance of predicted foreground pixels to the nearest annotated foreground pixel. $\varepsilon = 10^{-6}$ is added to avoid division by zero. The second term is an approximation of the averaged minimal distance

of foreground pixels to the nearest prediction. Setting $\alpha > 1$ emphasizes the second term with respect to the first one.

The WHD as defined is vulnerable in case no or few pixels are annotated as foreground and the majority of pixels is predicted as background with a large probability. Although this prediction is near the correct answer the first term yields a large value. We cure this problem setting predictions $p_f(b) < 0.1$ to zero. If no pixel is annotated as foreground $d(a, b)$ in the first term is undefined and we define $d(a, b)$ as the minimal distance of b to the border of Ω plus one. The second term is defined as zero in this case.

In [27] it is reported that WHD leads to unstable training and is therefore combined with the patch-based point loss (PPL) defined as

$$
L_{\mathrm{PPL}} = \sum_{\Omega_{i,j}} \left| \sum_{b \in \Omega_{i,j}} \tilde{p}_f(b) - |\Omega_{i,j} \cap Y_+| \right| \quad (2)
$$

where the windows $\Omega_{i,j}$ are a partitioning of the image domain, and $\tilde{p}_f(b)$ is the predicted probability $p_f(b)$ if $p_f(b) > \lambda_T$, zero otherwise. Thus, PPL compares the sum of these clamped probabilities and the number of foreground pixels in a window summed over all non overlapping windows. Similar to [27] we use a linear combination

$$
\mu \cdot WHD + (1 - \mu) \cdot PPL, \qquad \mu \in [0, 1], \quad (3)
$$

as loss function after an initial training, where in our experiments we use CombCED.

In analogy we use the focal loss (FL) [12] as a loss function subsequent to an initial training of weights in an attempt to improve a pretrained network. FL generalizes CE by adding a modulation factor:

$$
L_{\mathrm{FL}} = - \sum_{b \in \Omega} (1 - p_t(b))^{\gamma} \log(p_t(b)), \quad (4)
$$

where $p_t(b)$ is the predicted probability for the true class. In case of $\gamma = 0$, FL reduces to CE. FL emphasizes the hard to predict examples during training.

3.2. Post-processing and main root extraction

Applying the CNNs to the input images provides us with binary predictions for root pixels and background. However, as until now no topological knowledge has been considered in the segmentation process, roots sometimes decompose into multiple connected components. While this is not a serious problem with regard to overall segmentation accuracy as, e.g., measured by recall and precision rates, gaps significantly hamper the extraction of overall root system architecture and, with regard to the challenge task, in particular the extraction of the MR for each plant.

Therefore we use a post-processing pipeline on the segmentation results where we also consider temporal information from time series. The post-processing consists of two

main stages. First we aim at closing gaps in roots and reconnect branches to the main component of a plant that might have been detached. Subsequently, assuming that each plant is now represented by a single connected component, the MR of the corresponding plant is extracted based on topological skeleton analysis and graph-based path search. Both stages rely on the skeletons of the connected components which are extracted with the algorithm of Zhang et al. [30].

Gap closing and branch reconnection Gaps splitting the segmentation of a root system can be distinguished into two main categories: (I) gaps within a stretch of a root, (II) gaps disconnecting the root system at branching points. We consider both cases in turn.

Gaps of type (I) are characterized by pairs of end points in the skeleton where both points are only a short distance apart from each other and the skeleton segments located next to the end points do not significantly differ in their orientations. Thus, we initially locate skeleton end points as points with not more than one neighboring pixel in the skeleton. Subsequently all pairs of end points are detected which satisfy the following three criteria: (i) small distance, (ii) similar orientation of the skeletons at both end points, (iii) which are in turn similar to the orientation of the line connecting both end points. Thresholds for these criteria are set empirically. To connect such pairs of end points we apply a Dijkstra shortest path search. To this end we convert the local image patch around both end points into a graph representation with the pixels as graph nodes and their 8-neighbors connected by edges. The weight of the edges is defined as the response of an anisotropic vesselness filter measuring the correlation between the local intensity structure and the theoretical landscape of locally linear root structures (for more details see [15] where the same idea is used to close cell contours). If the final path is not longer than 1.5 times the distance between the end points they are connected. The width of the connecting segment is derived from the width of the root segments to be connected.

In case of type (II) gaps typically only a skeleton end point exists in the detached branch, but not in the root to be connected to. To check if an end point of a branch should be reconnected to a nearby root component we estimate the orientation of the final part of the skeleton branch and search in its direction for nearby root pixels. If at least one pixel is found within a maximum distance we insert a line segment and derive its width from the width of the branch.

Main root extraction The root system of an Arabidopsis plant consists of a single MR and any number of lateral roots. According to biological experts there is no clear definition of the MR except that it starts at the hypocotyl and is usually the longest root. In many images of the challenge data sets the hypocotyl cannot easily be localized, as it is often hidden by leaves and their stems. Thus, we define it as the end or branch point of the skeleton of the plant component located topmost in the image. As tip of the MR we use the pixel of the component closest to the bottom image border. The MR can then be found as the shortest path between root tip and hypocotyl applying a Dijkstra path search.

Obviously this approach assumes that each plant is represented by a single component. In practice this assumption is often not fulfilled as even after gap closing multiple components may survive for one root system, and further components may result from clutter or leaves. Hence, to select a single component per plant we define position priors for the hypocotyls and restrict locations of tip points with positions in the previous frame. In all experiments of the challenge test data set four seeds are initially planted in the upper third of the images at approximately constant positions. Thus, we define four regions of interest (ROIs) in the upper third of the image around the four seed positions and process each of these ROIs independently. We identify the largest connected component within the ROI and detect the MR within this component as described above.

The localization of the tip of the MR as described fails in some cases. One such situation occurs if the roots of two or more plants overlap and the corresponding components merge. We detect such cases by comparing the size of the components of each plant between subsequent frames. In case of overlapping roots the size usually almost doubles and the connected component covers more than one plant. As a consequence it contains the tip points of several MRs and the lowest pixel selected as tip may be the wrong one. We avoid this by enforcing additional constraints on the tip point of a plant, i.e., enforce a certain maximal distance from the position in the previous frame. This maximum distance, however, may lead to cases where no tip is found at all. This happens if a wrong connected component was initially selected, e.g., due to clutter in the ROI. In such cases we process the second largest component which in the majority of situations is the correct one.

Finally we remove all components which are too small and/or too far away from any MR component (see Sec. 5.2).

4. Data Sets

We conduct our experiments on the data sets of the Arabidopsis Root Segmentation Challenge 2021. For model training two data sets with partial annotations were provided. Due to time constraints we use only the one with binary labels consisting of 34 video sequences yielding a total of 1542 images. For 207 images annotations of roots and background is given, yielding data set D_{bin}. The challenge test data set D_{test} provided without annotations comprises 22 video sequences with a total number of 933 images.

We randomly partition D_{bin} into training D_{bin}^{train}, validation D_{bin}^{val} and test D_{bin}^{test} data with the ratio of 70:20:10. We crop images to size 512×512 pixels with an offset of 384 or to size 256×256 pixels with an offset of 192.

5. Results

5.1. Semantic Segmentation

Experimental Setup For SegRoot training we use the SegNet implementation[1] and optimizer described in [25].

The U-Net architecture in the original structure and in the VGG16 structure as well as the Hi-Fi architecture are our own implementations in PyTorch. Training and test is based on the code[2] used in [23]. We implement FL as well as WHD and PPL in addition to the existing dice and CE losses. In PPL we set the patch size to 32, as suggested in [27], and $\lambda_T = 0.5$ as this is our threshold for prediction. During training we use the stochastic gradient descent as optimizer with a momentum 0.99 and weight decay 10^{-5}. As a result we always choose the epoch with the best validation results and also test our net on this epoch.

SegRoot The SegRoot network [25] was initially designed for root segmentation in minirhizotron images. Here we retrain the network from scratch in different configurations. The SegRoot code supports to configure the network structure, i.e., the number of feature maps in the first convolutional layer (width), which is then doubled subsequent to each max pooling, and also the number of resolution levels (depth). In the original work a width of 8 and a depth of 5, i.e., an 8-5-net, was proposed as best compromise between network complexity and segmentation performance. Configurations with larger widths showed slightly better performance, but generally required more training time.

We tested configurations with widths of 8, 16 and 32, and depths of 4 and 5. A learning rate $lr = 10^{-2}$ was applied for the 8-5-net in [25], and for wider configurations $lr = 10^{-4}$ was chosen. For widths of 16 and 32 $lr = 10^{-4}$ worked well in our studies, but with $lr = 10^{-2}$ no learning process was observed on the 8-4- and 8-5-nets, i.e., scores dropped instantly close to zero and never recovered. Thus, we tested also learning rates of $10^{-3}, 10^{-4}$ and 10^{-5}, but with $lr = 10^{-3}$ learning neither happened. Batch sizes varied with available memory from 64 for the 8-5- and 16-4-nets, 32 for the 16-5- and 32-4-nets to 16 for the 32-5-net. SegRoot requires to use crops of size 256×256. The dice scores for the various experiments on D_{bin}^{test} are shown in Tab. 1.

The best dice score on D_{bin}^{test} was achieved with the 32-4-net performing slightly better than 16-4 with $lr = 10^{-4}$ and 32-5 with $lr = 10^{-5}$. Configurations like 8-4 and 16-5 with $lr = 10^{-4}$ performed also well. On the contrary, trainings with $lr = 10^{-5}$ performed significantly worse, except for the 32-5-net, but with $lr = 10^{-4}$ no training at all could be initiated for this net. According to studies on configurations with a width of 64 in [25] it might be hypothesized that such network models could boost the SegRoot performance further, and we plan to extend our study towards these models.

---

[1]https://github.com/wtwtwt0330/SegRoot

[2]https://github.com/Abe404/segmentation_of_roots_in_soil_with_unet

width	depth = 4		depth = 5	
	$lr = 10^{-4}$	$lr = 10^{-5}$	$lr = 10^{-4}$	$lr = 10^{-5}$
8	0.833	0.614	0.825	0.699
16	0.870	0.776	0.847	0.781
32	0.874	0.817	fail	0.870

Table 1. Dice scores for D_{bin}^{test} for different SegRoot widths (8, 16, 32), depths (4, 5) and learning rates ($lr = 10^{-4}, 10^{-5}$). For the 32-5-net with $lr = 10^{-4}$ no learning process could be observed.

U-Net and Hi-Fi As an alternative to SegNet we investigate U-Net and Hi-Fi. As the structure of the encoder resp. backbone we employ VGG16 [21] and the one proposed in the original U-Net paper [18]. As loss functions CE, DI, and CombCED are used, group normalization as opposed to no normalization, and learning rates of 10^{-3} and 10^{-4}. Thus, a total of 48 combinations of hyper parameters result.

First, we conducted 8 replicates of the hyper parameter combination used for the challenge submission (see below) to estimate the variance due to random initialization. The mean dice score on D_{bin}^{test} is 0.910 with a maximal difference of 0.0020. In the following we consider differences of the dice score in the second decimal place as considerable and not to be attributed to random effects.
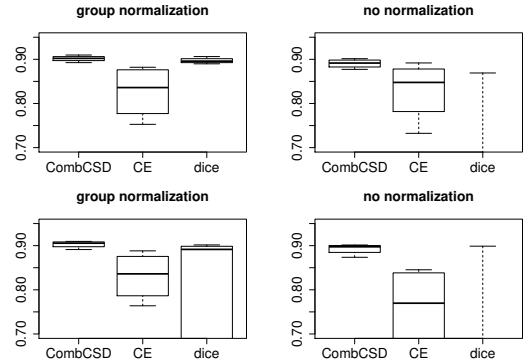

Figure 2. Boxplot of dice scores on D_{bin}^{test} for different backbone structures, loss functions, and normalizations. Top row: VGG16 structure, bottom row: original U-Net structure.

The boxplots in Fig. 2 show that CombCED yields a better dice score in almost all cases compared to CE. This may be partly attributed to the fact, that we assess performance with the dice score which is part of the CombCED loss function. In addition, however, CE is less robust with respect to hyper parameters which are thus more difficult to tune. DI can be expected to perform well, as it coincides with the performance measure. However, training is quite unstable in this case resulting in a dice score of less than 0.02 for 8 of the experiments. As evident from Tab. 2 only one of

lr	CombCED	CE	DI
U-Net, VGG16 Structure			
10^{-3}	0.902	0.882	0.895
10^{-4}	0.903	0.802	0.900
U-Net, Original U-Net Structure			
10^{-3}	0.910	0.890	0.890
10^{-4}	0.904	0.809	0.900
Hi-Fi, VGG16 Structure			
10^{-3}	0.910	0.870	0.906
10^{-4}	0.893	0.753	0.890
Hi-Fi, Original U-Net Structure			
10^{-3}	0.907	0.863	0.902
10^{-4}	0.891	0.764	0.006

Table 2. Dice scores on D_{bin}^{test} using group normalization.

these instable trainings occurs when using group normalization. In contrast we find only one such experiment for CE, and none for CombCED. In case of successful training in most cases competitive dice scores result compared to CombCED, which is true for all experiments with group normalization and VGG16. We speculate that CE alleviates the instability of DI in the combination still giving the advantage of dice as the performance measure.

Next we analyze the effect of group normalization. Fig. 2 indicates that adding normalization tends to produce a more stable performance. Comparing all experiments with respect to normalization variants, CombCED gives a considerable better dice score in two thirds of experiments and comparable results otherwise. For CE no clear tendency can be observed if the single experiment with a lack in proper learning is omitted. Due to these findings we only consider training with group normalization in the following.

Tab. 2 shows that using CombCED the four combinations of network type and backbone structure perform comparable with the exception of Hi-Fi and $lr = 10^{-4}$. Training with DI yields comparable results except for one experiment where no learning happened at all. This dice score of 0.90 ± 0.01 for these experiments is the best performance on our test set we observed. The CE delivers considerable worse dice scores for most of the cases, especially for $lr = 10^{-4}$. With respect to learning rate $lr = 10^{-3}$ outperforms 10^{-4} for several combinations. However, we feel that more experiments should be performed and expect that the appropriate one depends on the other hyper parameters.

In all experiments where learning was successful precision and recall vary slightly and quite symmetric around the stable dice score. In Fig. 3 the evolution of performance during training on D_{bin}^{val} is displayed.

In summary, we find a slight advantage of CombCED loss and group normalization, while both network types – U-Net and Hi-Fi – and backbone structures perform compa-
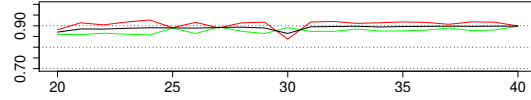
Figure 3. Performance on D_{bin}^{val} in the course of training. Black: dice score, red: precision, green: recall.

μ	$lr = 10^{-8}$	$lr = 10^{-9}$	$lr = 10^{-10}$
0.4	0.846/0.883/0.864	0.859/0.886/0.872	0.862/0.896/0.879
0.5	0.844/0.845/0.845	0.864/0.888/0.876	0.862/0.895/0.878
0.6	0.818/0.601/0.693	0.863/0.891/0.877	0.862/0.896/0.879

Table 3. Test results on training with a linear combination of WHD and PPL with varying weighting factor μ as loss function and different learning rates. For $lr = 10^{-8}$, 10^{-9} and 10^{-10} recall/precision/dice are given.

lr	Recall	Precision	Dice
10^{-7}	0.856	0.946	0.899
10^{-8}	0.889	0.929	0.908
10^{-9}	0.907	0.914	0.910
10^{-10}	0.908	0.912	0.910

Table 4. Test scores for focal loss with $\gamma = 1.0$ trained on pretrained weights with different learning rates.

rable. Obviously, still more combinations of hyper parameters could be examined. However, we speculate that no significant improvements may be achieved especially taking ambiguities of groundtruth annotation into account, see also "Qualitative results" below.

WHD and PPL loss We employ a linear combination of WHD and PPL to train a U-Net with the VGG16 structure which was pretrained using CombCED, $lr = 10^{-4}$ and group normalization. Due to memory limitations caused by WHD crops sized 256×256 pixels are used and α is set to 4. We train 15 epochs and then choose the epoch with the best validation dice score for testing and present results in Tab. 3. Training with $lr = 10^{-9}$ and $lr = 10^{-10}$ results in a slight decrease of the dice score compared to the pretrained network and a slight imbalance between precision and recall developments. Using $lr = 10^{-8}$ intensifies this effect especially with increasing μ from 0.4 to 0.5 and 0.6. To summarize, at least based on these three performance metrics WHD does not give an improvement, but rather a decline in performance. Potential improvements with respect to the aim of geometry-awareness are hard to quantify and need to be scrutinized more carefully.

Focal loss As a second loss to further train the same U-Net with VGG16 structure we use the focal loss. In Tab. 4, we show recalls, precisions and dice scores for different learning rates. With respect to these performance measures $lr = 10^{-10}$ and 10^{-9} are obviously too small

to make a difference. With the increase of the learning rate, precision and recall diverge with considerable higher precision while reducing the recall. This leads to an insignificant change of the dice score. With $lr = 10^{-7}$ the imbalance of recall and precision further increases and the decrease of the dice score gets considerable. Whether this increase in precision for the price of smaller recall is an advantage and if so to which degree is to be answered by the application.

Qualitative results For our challenge results we choose the U-Net with VGG16 structure, CombCED, $lr = 10^{-4}$, and group normalization due to best performance in our preliminary tests on D_{bin}^{test} (data not shown). After 40 epochs we continued training with FL setting $\gamma = 1.0$ using $lr = 10^{-8}$ as visual inspection indicated superior performance on D_{test}. We observed that including training with FL decreases false positives (FPs) in the leaf regions. While it also induces more false negatives (FNs) in root gaps this is at least partially compensated by gap closing in the postprocessing stage. Note, that we trained using crops of size 256×256 during the preliminary tests, thus the performances given in Tab. 2 and 5 slightly differ.

Qualitative aspects of the results achieved with this network are discussed next. While the roots are usually quite well segmented in the middle and lower parts of the images, errors seem to appear more frequently in the upper parts containing the leaves. To validate this quantitatively, we evaluate the region at the top including most leaves and the rest of the images separately. The leaf region is sized 2000×645 pixels and located with its top left corner at position $(620, 0)$. In Tab. 5 recalls, precisions, and dice scores are given.

	Recall	Precision	Dice
Image complete	0.884	0.940	0.911
Leaf region	0.844	0.912	0.879
Non-leaf region	0.902	0.951	0.926

Table 5. Evaluation results on leaf and non-leaf regions in D_{bin}^{test}.

Performance is best in non-leaf regions and outperforms all scores on the complete images. For the leaf regions performance significantly drops compared to the non-leaf regions, but also with regard to complete images, e.g., the dice score drops from 0.926 and 0.911, respectively, to 0.879. Thus, improving segmentation in particular in leaf regions seems to be promising for boosting segmentation quality.

Given the above observations we further investigated root segmentation in leaf regions by visual groundtruth comparisons. It turns out that root annotations seem inconsistent sometimes which occasionally causes errors with regard to groundtruth although the segmentation appears reasonable according to the image data. First, roots covered by leaves are sometimes annotated as foreground, sometimes
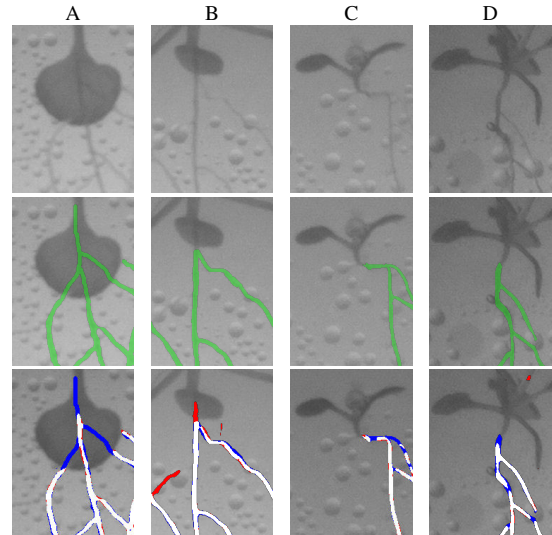
Figure 4. Examples for inconsistent annotations of roots covered by leaves (columns A and B), and examples for rough hypocotyl localization (columns C and D). Top row: input images, middle row: groundtruth, bottom row: overlay of our segmentations (white: true positive, red: FP, blue: FN).

not. For the samples in Fig. 4 our network predicts parts of the roots behind the leaves, which for the example in the left column (A) results in FNs as well as TPs, while for the example in the second column (B) FPs result. Second, the hypocotyl position is sometimes very roughly localized so that parts of the stem are marked as root. In Fig. 4, third column (C), the hypocotyl is properly localized and our segmentation is consistent with the annotation, while in the forth column (D) parts of the stem are also annotated as foreground which we miss in our segmentation.

Additionally, in some images there are spurious pixels annotated as foreground distant to the nearest root system. While these incorrect annotations have only weak impact on the dice score the effect on the Hausdorff Distance, one of the challenge metrics, may be considerable.

5.2. Root Segmentation Challenge

According to the challenge organizers our CNN described in the previous paragraph achieved a dice score of 0.761 and completeness and correctness scores of 0.894 and 0.955, respectively, on a subset of 132 images of D_{test} which were used for producing the challenge results. Compared to a human annotator who provided annotations for comparison and achieved a dice score of 0.802 with completeness and correctness scores of 0.957 and 0.948, respectively, this is only slightly worse. However, if we compare to the dice scores on D_{bin}^{test} of roughly 0.91 that we observed
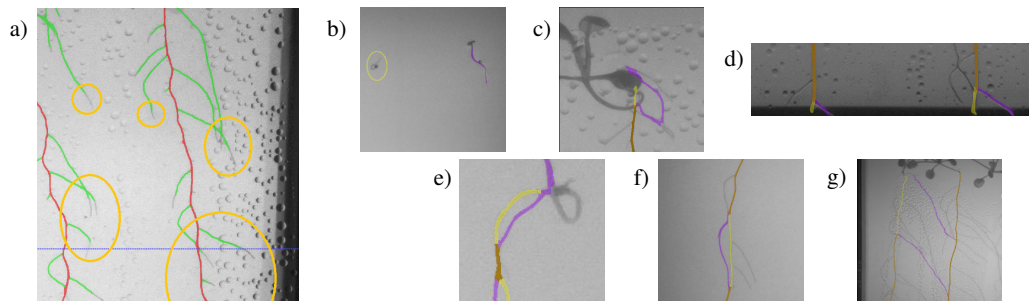
Figure 5. a), Sample segmentation result on D_{test}. Tip segments of the lateral roots are frequently missed (red: TP of MRs, green: TP of lateral root). Image courtesy to the Root Segmentation Challenge Team. b)-g), examples for errors in MR root segmentation according to our own judgement of groundtruth (orange: TP, yellow: FN, purple: FP). For further details refer to the text.

during training of our model, the challenge results are significantly worse. The primary reason for these large deviations seem to originate from the fact that we often miss larger parts of lateral root tips as shown in Fig. 5 a. This could point to significant differences in image characteristics between our training data D_{bin} and the test data D_{test}.

Our post-processing stage for extracting the MRs relies on several empirically chosen parameters which were set rather liberal. As the challenge evaluation metrics considers the total number of detected components, our intention was to reduce the number of components as much as possible while accepting some small erroneously closed gaps. Hence, we extract paths between end points with a maximum distance of 40 pixels and a maximum deviation in orientations of 50°. For reconnecting branches the maximal distance is set to 20 pixels. After extracting the MR we remove all components smaller than 75 pixels, components smaller than 250 pixels if more distant than 100 pixels to all plant components, and components smaller than 500 pixels if more distant than 350 pixels. This allows to eliminate clutter, but keep branches and other root parts that could not be linked to the main component of a plant.

We performed a thorough analysis of the quality of our MR extraction by visually inspecting all 933 images of D_{test}. In 17 out of the 22 video sequences almost 90% of the MRs seem to be identified correctly. This is also confirmed by our challenge results where we achieve a completeness of 0.918 and a correctness of 0.952 on the MR pixels. If MR extraction fails to a large extent this is mostly attributable to one of three typical issues. In the early images of a video sequence where plants start to grow it is often hard to distinguish between small components resulting from clutter and correct root components (Fig. 5 b). From our subjective assessment and without knowing the groundtruth we may sometimes miss small roots and MRs in the first images of a sequence. Likewise, sometimes wrong components are selected as root components lead-

ing to MRs located in noise components. A third more serious source of errors are path errors where the correct component is traced, however, the MR path includes wrong root segments. This mainly happens at the top or bottom of root components if the hypocotyl or tip point is wrongly detected, e.g., due to the roots growing out of the image or plate, or the seed points and roots being covered by parts of a leaf or stem (Fig. 5 c, d). In some rare cases the MR path is wrongly extracted due to ambiguities in the image data (Fig. 5 e,f) where it is hardly possible to correctly trace the MR without considering additional temporal information. Including such data in MR extraction would be one of the most promising directions for improving the quality of MRs. Anyway, most of the MR errors due to these issues affect only small portions of the MR. In three video sequences, however, MR extraction fails seriously for a couple of images. In these sequences components for different plants merge and the path extraction in parts follows roots of the wrong plant (Fig. 5 g). Here also temporal information and path alignment between successive time points might help to extract MR paths more robustly.

6. Conclusions

CNNs are a common approach for semantic segmentation and have also gained interest to segment root images. Here we present our approach for the segmentation task of this year's CVPPA Arabidopsis Root Segmentation Challenge. In general, we achieve fair segmentation scores and particularly succeed in identifying the main roots, while complete segmentation of the tips of lateral roots remains challenging. Thus, our study on alternative CNN architectures, loss functions and parameters could guide the development of more powerful models. Together with considering additional temporal information from time series in the post-processing stage this may lead to further performance boosts and foster progress in root image segmentation.

References

[1] Jonathan A. Atkinson, Michael P. Pound, Malcolm J. Bennett, et al. Uncovering the hidden half of plants using new advances in root phenotyping. *Current Opinion in Biotechnology*, 55:1–8, 2019.

[2] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495, 2017.

[3] Hans de Kroon, Liesje Mommer, and Aya Nishiwaki. Root competition: towards a mechanistic understanding. In *Root Ecology*, pages 215–234. Springer, 2003.

[4] Nicolás Gaggion, Federico Ariel, Vladimir Daric, et al. ChronoRoot: High-throughput phenotyping by deep segmentation networks reveals novel temporal parameters of plant root system architecture. *bioRxiv*, 2020. DOI: 10.1101/2020.10.27.350553.

[5] Alberto Garcia-Garcia, Sergio Orts-Escolano, Sergiu Oprea, et al. A survey on deep learning techniques for image and video semantic segmentation. *Applied Soft Computing*, 70:41–65, 2018.

[6] Alejandro González, Xavier Sevillano, Isabel Betegón-Putze, et al. MyROOT 2.0: An automatic tool for high throughput and accurate primary root length measurement. *Comp. and Electronics in Agriculture*, 168:105125, 2020.

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, et al. Deep residual learning for image recognition. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[8] Shruti Jadon. A survey of loss functions for semantic segmentation. In *Proc. of the IEEE Conf. on Computational Intelligence in Bioinformatics and Computational Biology*, pages 1–7, 2020.

[9] Mark G. Johnson, David T. Tingey, Donald L. Phillips, et al. Advancing fine root research with minirhizotrons. *Environmental and Experimental Botany*, 45(3):263–289, 2001.

[10] Wei Ke, Jie Chen, Jianbin Jiao, et al. SRN: Side-output residual network for object symmetry detection in the wild. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pages 1068–1076, 2017.

[11] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

[12] Tsung-Yi Lin, Priya Goyal, Ross Girshick, et al. Focal loss for dense object detection. In *Proc. of the IEEE Int. Conf. on Computer Vision*, pages 2980–2988, 2017.

[13] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.

[14] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *Proc. of Fourth IEEE Int. Conf. on 3D Vision*, pages 565–571, 2016.

[15] Birgit Möller and Katharina Bürstenbinder. Semi-automatic cell segmentation from noisy image data for quantification of microtubule organization on single cell level. In *Proc. of IEEE 16th Int. Symposium on Biomedical Imaging (ISBI)*, pages 199–203, Venice, Italy, April 2019.

[16] Alejandro Newell, Kaiyu Yang, and Jia Deng. Stacked hourglass networks for human pose estimation. In *European Conf. on Computer Vision*, pages 483–499. Springer, 2016.

[17] Michael P. Pound, Andrew P. French, Jonathan A. Atkinson, et al. RootNav: Navigating Images of Complex Root Architectures. *Plant Physiology*, 162(4):1802–1814, 2013.

[18] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Proc. of Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241. Springer, 2015.

[19] Zaigham Shahzad, Fabian Kellermeier, Emily M. Armstrong, et al. EZ-Root-VIS: A Software Pipeline for the Rapid Analysis and Visual Reconstruction of Root System Architecture. *Plant Physiology*, 177(4):1368–1381, 2018.

[20] Wei Shen, Kai Zhao, Yuan Jiang, et al. Object skeleton extraction in natural images by fusing scale-associated deep side outputs. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pages 222–230, 2016.

[21] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *Proc. of the Int. Conf. on Learning Representations (ICLR)*, 2015.

[22] Radka Slovak, Christian Göschl, Xiaoxue Su, et al. A Scalable Open-Source Pipeline for Large-Scale Root Phenotyping of Arabidopsis. *The Plant Cell*, 26(6):2390–2403, 2014.

[23] Abraham G. Smith, Jens Petersen, Raghavendra Selvan, et al. Segmentation of roots in soil with U-Net. *Plant Methods*, 16(1):1–15, 2020.

[24] Christian Szegedy, Wei Liu, Yangqing Jia, et al. Going deeper with convolutions. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pages 1–9, 2015.

[25] Tao Wang, Mina Rostamza, Zhihang Song, et al. SegRoot: A high throughput segmentation method for root image analysis. *Comp. and Electronics in Agriculture*, 162:845–854, 2019.

[26] Saining Xie and Zhuowen Tu. Holistically-nested edge detection. In *Proc. of the IEEE Int. Conf. on Computer Vision*, pages 1395–1403, 2015.

[27] Weijian Xu, Gaurav Parmar, and Zhuowen Tu. Geometry-Aware End-to-End Skeleton Detection. In *Proc. of British Machine Vision Conf.*, volume 2, page 7, 2019.

[28] Robail Yasrab, Jonathan A. Atkinson, Darren M. Wells, et al. RootNav 2.0: Deep learning for automatic navigation of complex plant root architectures. *GigaScience*, 8(11), 2019.

[29] Robail Yasrab, Michael P. Pound, Andrew P. French, et al. PhenomNet: Bridging Phenotype-Genotype Gap: A CNN-LSTM Based Automatic Plant Root Anatomization System. *bioRxiv*, 2020. DOI: 10.1101/2020.05.03.075184.

[30] Tongjie Y. Zhang and Ching Y. Suen. A fast parallel algorithm for thinning digital patterns. *Communications of the ACM*, 27(3):236–239, 1984.

[31] Kai Zhao, Wei Shen, Shanghua Gao, et al. Hi-Fi: Hierarchical Feature Integration for Skeleton Detection. In *Proc. of Int. Joint Conf. on Artificial Intelligence*, pages 1191–1197, 7 2018.

Chapter 4

Object Shape and Texture

RESEARCH PAPER

Microtubule-associated protein IQ67 DOMAIN5 regulates morphogenesis of leaf pavement cells in *Arabidopsis thaliana*

Dipannita Mitra[1,*], Sandra Klemm[1,*], Pratibha Kumari[1,*], Jakob Quegwer[1], Birgit Möller[2],
Yvonne Poeschl[2,3], Paul Pflug[1], Gina Stamm[1], Steffen Abel[1,4,5] and Katharina Bürstenbinder[1,†]

[1] Department of Molecular Signal Processing, Leibniz Institute of Plant Biochemistry (IPB), D-06120 Halle (Saale), Germany

[2] Institute of Computer Science, Martin Luther University Halle-Wittenberg, D-06120 Halle (Saale), Germany

[3] iDiv, German Integrative Research Center for Biodiversity, D-04103 Leipzig, Germany

[4] Institute of Biochemistry and Biotechnology, Martin Luther University Halle-Wittenberg, D-06120 Halle (Saale), Germany

[5] Department of Plant Sciences, University of California, Davis, CA 95616, USA

* These authors contributed equally to this work.
† Correspondence: Katharina.Buerstenbinder@ipb-halle.de

Abstract

Plant microtubules form a highly dynamic intracellular network with important roles for regulating cell division, cell proliferation, and cell morphology. Their organization and dynamics are co-ordinated by various microtubule-associated proteins (MAPs) that integrate environmental and developmental stimuli to fine-tune and adjust cytoskeletal arrays. IQ67 DOMAIN (IQD) proteins recently emerged as a class of plant-specific MAPs with largely unknown functions. Here, using a reverse genetics approach, we characterize Arabidopsis IQD5 in terms of its expression domains, subcellular localization, and biological functions. We show that *IQD5* is expressed mostly in vegetative tissues, where it localizes to cortical microtubule arrays. Our phenotypic analysis of *iqd5* loss-of-function lines reveals functions of IQD5 in pavement cell (PC) shape morphogenesis. Histochemical analysis of cell wall composition further suggests reduced rates of cellulose deposition in anticlinal cell walls, which correlate with reduced anisotropic expansion. Lastly, we demonstrate IQD5-dependent recruitment of calmodulin calcium sensors to cortical microtubule arrays and provide first evidence for important roles for calcium in regulation of PC morphogenesis. Our work identifies IQD5 as a novel player in PC shape regulation and, for the first time, links calcium signaling to developmental processes that regulate anisotropic growth in PCs.

Keywords: Arabidopsis, calcium, calmodulin, cell wall, IQ67 DOMAIN, microtubules, pavement cell shape, signaling

Introduction

The plant cytoskeleton, comprised of actin filaments and microtubules (MTs), forms a three-dimensional intracellular network that determines cell division and cell morphology, and serves as tracks for cellular transport of various cargoes, including organelles, proteins, and other macromolecular complexes (Wasteneys and Yang, 2004b; Hussey *et al.*, 2006; Akhmanova

and Hammer, 2010). Networks of MTs form highly dynamic arrays and adopt specific functions during the plant life cycle, including essential roles in cell division and expansion, intra- and intercellular transport, cellular organization, and the deposition of cell wall material. In interphase cells, MTs reorganize into cortical networks tethered to the plasma membrane (PM),

185

which serve as tracks for PM-localized cellulose synthase complexes (CSCs) and thereby define the direction of cellulose deposition (Paredez *et al.*, 2006; Liu *et al.*, 2015). In addition to their important roles in development, MT arrays function during growth adaptation in response to changing environmental conditions, thereby contributing to plant fitness (Wasteneys and Yang, 2004a).

To engage in these diverse cellular functions, MT organization and dynamics are tightly controlled (Wasteneys, 2002). Developmental and environmental stimuli can induce rapid reorganization of the MT cytoskeleton (e.g. in response to mechanical stimulation) which can occur within a few minutes and involves changes in MT trajectories, as well as altered rates of (de-) polymerization (Hardham *et al.*, 2008). Phytohormones exert control over MT orientation (Shibaoka, 1994; Locascio *et al.*, 2013; Takatani *et al.*, 2015), and signaling via the second messenger calcium (Ca^{2+}) has been implicated in cytoskeletal control, as suggested by sensitivity of MT stability to elevated Ca^{2+} concentrations (Hepler, 2005, 2016). MT-associated proteins (MAPs), which bind to tubulin subunits, play essential roles for regulating cytoskeletal behavior (Lloyd and Hussey, 2001; Sedbrook, 2004) and are likely candidates to integrate incoming signals into appropriate responses. Numerous MAPs have been identified in plants, which mediate bundling, cross-linking, nucleation, or severing of MTs, or, in the case of plus end-tracking MAPs, control dynamic instability at polymerizing plus ends (Akhmanova and Steinmetz, 2008; Horio and Murata, 2014). Other MAPs facilitate physical connections between MTs and protein complexes, such as CSCs (Bringmann *et al.*, 2012), or cross-linking to the actin cytoskeleton (Schneider and Persson, 2015). MAPs also mediate tethering of MTs to the PM (Bayer *et al.*, 2017; Oda, 2018), which is required for stabilization against the pushing forces of CSCs (Liu *et al.*, 2016) and potentially contributes to subcompartmentalization of PMs into functional subdomains (Sugiyama *et al.*, 2017). Still, the modes by which external signals are integrated into MT (re-) orientation and how MAPs contribute to it are poorly understood.

We previously identified IQ67 DOMAIN (IQD) family proteins as the largest known class of MAPs in plants (Bürstenbinder *et al.*, 2017b), which are encoded by multigene families of 23–66 members in several angiosperms, including *Arabidopsis thaliana*, *Oryza sativa* (rice), *Solanum lycopersicum* (tomato), and *Glycine max* (Abel *et al.*, 2005; Huang *et al.*, 2013; Feng *et al.*, 2014). The family-defining IQ67 domain harbors motifs with predicted roles in binding to calmodulin (CaM) Ca^{2+} sensor proteins that are an integral part of the cellular Ca^{2+} decoding machinery (Abel *et al.*, 2005). Thus, IQDs are likely candidates for integration of CaM-dependent Ca^{2+} signaling into MT (re-)organization and growth regulation (Bürstenbinder *et al.*, 2017a). First experimental data point to important roles for IQD proteins in plant development, as indicated by altered fruit shape and grain size in plants with elevated expression levels of tomato *SUN/IQD12* and rice *GRAIN SIZE ON CHROMOSOME 5(GSE5)/IQD21*, respectively (Xiao *et al.*, 2008; Duan *et al.*, 2017). Functions of IQDs in MT organization are supported by differential MT patterns, which are induced upon overexpression of individual

family members in transient expression assays in *Nicotiana benthamiana* (Bürstenbinder *et al.*, 2017b). Divergent MT patterns in *N. benthamiana* correlate with divergent effects on MT organization and growth in transgenic Arabidopsis *IQD* overexpression plants, as indicated by altered plant morphology and cell shape upon ectopic overexpression of, for example, *IQD14* and *IQD16* visible during development and in diverse tissues (Bürstenbinder *et al.*, 2017b). Mechanistic studies on IQD functions, however, are still limited because (i) phenotypes are almost exclusively reported in *IQD* gain-of-function or overexpression lines; (ii) multigene families are not easily amenable to reverse genetics approaches due to functional redundancies; and (iii) insights into the spatial and temporal control of Ca^{2+} signal generation during development are sparse due to limited sensitivities of intracellular Ca^{2+} imaging methods (Kudla *et al.*, 2018).

To identify functions of family members, we selected Arabidopsis IQD5, because MT pattern analysis upon overexpression of *YFP-IQD5* in transient expression assays in *N. benthamiana* suggested unique and specific roles for this family member in MT organization (Bürstenbinder *et al.*, 2017b). Moreover, IQD5, which belongs to phylogenetic subgroup IIIa of the IQD family, clusters separately from the other members of this subgroup, namely IQD6–IQD10 (Abel *et al.*, 2005), and thus may have specialized functions *in planta*. In this study, we present a systematic analysis of Arabidopsis *IQD5* using reverse genetics approaches. We identified expression domains of *IQD5* by analysis of transgenic *pIQD5::GFP-GUS* reporter lines, and determined its subcellular localization in transgenic *pIQD5::IQD5-GFP/iqd5-1* lines. We show that IQD5–green fluorescent protein (GFP) decorates cortical MTs at neck regions of leaf epidermis pavement cells (PCs). Loss of *IQD5* results in strongly reduced growth restriction at neck regions, which correlates with a reduced deposition of cellulose in anticlinal walls of PCs. Recombinant IQD5 interacts with apo-CaM and Ca^{2+}–CaM *in vitro*, and IQD5 recruits CaM to MTs *in planta*. Moreover, we show that PC shape is sensitive to elevated external Ca^{2+} concentrations. Together, our research provides evidence for functions of IQD5 in shape establishment of leaf epidermis PCs and for the first time links Ca^{2+} signaling to the control of interdigitated growth of PCs.

Materials and methods

Plant material, growth conditions, and macroscopic phenotyping

Wild-type (WT) seeds (Col-0 accession) were originally obtained from the Arabidopsis Biological Resource Center. T-DNA insertion lines SALK_015580 and GK-288E09, referred to as *iqd5-1* and *iqd5-2*, respectively, were obtained from the Nottingham Arabidopsis Stock Centre. Genomic DNA was extracted as described in Bürstenbinder *et al.* (2007). Homozygous mutant lines were identified by PCR-based genotyping with the following primer combinations: *iqd5-1*, WT allele IQD105 (5'-GATTATCTCTGCCAAACAGCG-3') and IQD106 (5'-GGAGAGTGACTTGGGCTGAC-3'), insert IQD105+A004 (5'-ATTTTGCCGATTTCGGAAC-3'); *iqd5-2*, WT allele IQD075 (5'-ATGGGAGCTTCAGGGAGATG-3')+IQD076 (5'-GCGTTACAGCAGCTTGTTTTC-3'), and insert IQD076+A009 (5'-ATAATAACGCTGCGGACATCTACATTT-3'). A 2201 bp (*pIQD5_long*) and a 1207 bp (*pIQD5_short*) fragment of the *IQD5* promoter sequence were amplified from genomic DNA with IQD180

(5'-CACCTCTATATATGGTTCACAATCGAGACAC-3') and IQD345 (5'-CACCATAAATCACATCACTGTTTTTGGGT-3') forward primers, respectively, in combination with the IQD181 reverse primer (5'-TCTATCTCAATTCCAACGATCAG-3'), and mobilized into the pENTR/dTOPO vector. A genomic *pIQD5short::IQD5(w/-stop codon)* fragment was amplified with forward primer IQD1521 (5'-attB1-TCTCTATATATGGTTCACAATCGAGACAC-3') and reverse primer IQD1522 (5'-attB2-CTGCAAGCCTCTGTTTTATTGGGTCGG-3'), and mobilized into pDONR221. Fidelity of inserts was verified by sequencing. For generation of transgenic *pIQD5::GFP-GUS* and *pIQD5::IQD5-GFP* lines, the inserts were mobilized into pBGWFS7 and pB7FWG,0, respectively (Karimi *et al.*, 2002). Arabidopsis plants were transformed by *Agrobacterium tumefaciens*-mediated transfection using the floral dip method (Clough and Bent, 1998). Per construct, 10–24 independent lines were identified in the T_1 generation by Basta selection. T_2 plants were screened for the presence of single-copy T-DNA insertion by segregation analysis (Basta). For analysis of GFP fluorescence and β-glucuronidase (GUS) expression, 2–4 homozygous T_3 lines were included, which showed representative GFP fluorescence or *GUS* expression patterns.

Seeds were surface sterilized with chlorine gas, stratified for 2 d at 4 °C on *Arabidopsis thaliana* Salts (ATS) medium [1× ATS, 0.5% (w/v) agar gel, 1% (w/v) sucrose] (Lincoln *et al.*, 1990), and grown at 21°C under long-day conditions (16 h light, 8 h dark). For oryzalin treatments, seedlings were incubated for 1–2 h in liquid medium supplemented with 10 μM oryzalin in a final concentration of 0.25% (v/v) DMSO or the DMSO control as described in Bürstenbinder *et al.* (2013). Macroscopic growth parameters were analyzed in 5-day-old seedling and in 3-week-old plants. Root length was quantified with RootDetection (http://www.labutils.de/rd.html, last accessed 19 November 2018). Cotyledon and leaf area were measured with the Easy Leaf Area software (http://www.plant-image-analysis.org/software/easy-leaf-area, last accessed 19 November 2018) according to the manual. For treatment with Ca²⁺, seedlings were grown for 5 d on half-strength Murashige and Skoog (1/2 MS) medium supplemented with the indicated concentrations of $CaCl_2$ according to Chen *et al.* (2014).

RNA extraction and expression analysis

Total RNA was extracted from 2-week-old plants using TRI reagent. Synthesis of cDNA via reverse transcription and RT–PCRs were performed according to Bürstenbinder *et al.* (2007) with the following primers: *IQD5*, primer IQD117 (5'-CTATGCAAGCCTCTGTTTTATTGG-3'); *ACTIN2*, primer A005 (5'-CAAAGACCAGCTCTTCCATC-3'); and A006 (5'-CTGTGAACGATTCCTGGACCT-3'). For quantitative real-time PCR (qRT-PCR), conditions were selected as described in Bürstenbinder *et al.* (2017b) and the following primers were used: *IQD5*, IQD1777 (CAACTAAAGCCAACCGAGCA-3') and IQD1778 (GGTTTTGGGCAGATTTTTCC-3'), *PP2A* A015 (AGCCAACTAGGACGGATCTGGT-3') and A016 (CTATCCGAACTTCTGCCTCATTA-3'). In brief, 3–5 shoots of 2-week-old plants were pooled for RNA extraction. A 2 μg aliquot of DNase I-treated RNA was reverse transcribed with oligo(dT) primers using the Revert Aid First Strand cDNA synthesis kit (Thermo Fisher) to generate first-strand cDNA. Primer efficiencies were calculated from standard curves. A 1 μl aliquot of 1:10-diluted cDNA was used in a 10 μl reaction mix including Fast SYBR Green master mix (Applied Biosystems), and qPCRs were run on a 7500 Fast Real-Time PCR system with the following program: 10 min, 95 °C; 40 cycles of 3 s, 95 °C, and 30 s, 63 °C. Expression levels of *IQD5* were calculated relative to *PP2A*.

Microscopy, staining procedures, and image analysis

Whole-mount GUS staining of seedlings and plants was performed as described in Bürstenbinder *et al.* (2017b). Plant materials were cleared in chloral hydrate, and roots and seeds were imaged with a Zeiss axioplan 2 microscope using a differential interference contrast (DIC) objective. Imaging of whole seedlings, leaves, flowers, and siliques was performed with a Nikon SMZ 1270 stereo microscope.

Confocal imaging was performed with a Zeiss LSM 780 inverted microscope using a ×40 water immersion objective, unless stated otherwise. Generation of yellow fluorescent protein (YFP)–IQD5, Y_N–TRM1, red fluorescent protein (RFP)–TUA5, and mCherry–CaM2 is described in Bürstenbinder *et al.* (2017b) and Gantner *et al.* (2018). GFP was excited using a 488 nm laser and emission was detected between 493 nm and 555 nm. YFP was excited by a 514 nm laser, and emission was detected between 525 nm and 550 nm. For mCherry excitation, a 555 nm laser was used, and emission was detected between 560 nm and 620 nm. Fluorescence intensities of IQD5–GFP adjacent to the periclinal wall at convex and concave sides of lobes were quantified according to Armour *et al.* (2015). Average fluorescence intensities were measured with FiJi (Schindelin *et al.*, 2012) in a total of six cells from three independent seedlings, and five lobes per cell were analyzed. The vector series of Gehl *et al.* (2009) was used for generation of the bimolecular fluorescence complementation (BiFC) construct. For all samples included in the BiFC experiment, imaging was performed with an identical laser setting. In co-expression assays, mCherry and GFP fluorescence were recorded in the sequential mode.

For visualization of cell contours, cell outlines were visualized by propidium iodide (PI) staining as described in Bürstenbinder *et al.* (2017b), and imaged with a ×20 objective [5–10 days after germination (DAG)] or with a ×40 objective (2 and 3 DAG). PI was excited with a 555 nm laser, and emission was detected between 560 nm and 620 nm. Segmentation, feature quantification, and graphical visualization of PC shapes were conducted with the ImageJ plugin PaCeQuant and the associated R script (Möller *et al.*, 2017). For cells in cotyledons 5–10 DAG and in the third true leaf, the threshold for size filtering implemented in PaCeQuant was set to the default value of 240 μm². For cells in cotyledons 2 and 3 DAG, the threshold for size filtering was reduced to 75 μm². For time series analysis of cells during cotyledon development, cells were grouped by their sizes into the following categories: tiny <240 μm²; small, 240–1400 μm²; medium, 1400–4042 μm²; and large >4042 μm². Thresholds for small, medium, and large cells were chosen according to Möller *et al.* (2017).

For histochemical cellulose staining in cell walls, 5-day-old seedlings were incubated for 90 min in 0.04% (v/v) calcofluor white M2R dissolved in Tris–HCl buffer (pH 9.2). To stain callose and the cuticle, seedlings were incubated for 3 h and 5 min in 0.1% (v/v) aniline blue in 100 mM Na_2PO_4 buffer (pH 7.2) and 0.1% (w/v) auramine O in 50 mM Tris–HCl buffer (pH 7.2), respectively. Subsequently, seedlings were co-stained with PI to visualize cell contours. Dissected cotyledons were imaged with a Zeiss LSM 700 inverted microscope, using a ×40 water immersion objective. Calcofluor white, aniline blue, and auramine O were excited with a 405 nm laser, and emission was detected with a 490 nm short pass filter. Co-staining was recorded in the sequential mode.

To quantify fluorescence intensities along the boundaries of the cells, we established a workflow combining automatic segmentation based on the method implemented in PaCeQuant and quantification of fluorescence intensities along the contour segments. For each boundary pixel in an image, the set of adjacent cell regions in a 15 × 15 neighborhood around the pixel is determined and the fluorescence intensity value of the pixel is added to the total intensity sum of each of these regions. Finally, an average intensity value for the boundary of each cell region is calculated by dividing the intensity sum of the region through the total number of pixels that contributed to the specific region, which we implemented in MiToBo (Möller *et al.*, 2016).

Structure prediction, protein expression, and calmodulin binding assays

Structural prediction of the IQ67 domain of IQD5 spanning amino acids E87–L153 was performed using PHYRE2 (Kelley *et al.*, 2015), which revealed the highest similarities with the crystal structures of the CaM-binding domains of mouse myosin V (PDB:2IX7, 99.5% similarity) (Houdusse *et al.*, 2006) and mouse myosin-1c (PDB:4R8G, 99.8%) (Lu *et al.*, 2015). The predicted structure of the IQ67 domain of IQD5 was aligned with PDB:2IX7, which contains the crystal structure of apo-CaM bound to the first two IQ motifs of myosin V, using PyMol (DeLano, 2009). CaM was fitted to adjust for the different spacing of IQ

motifs by 11 and 12 amino acids in the CaM-binding domains of IQD5 and myosin V, respectively.

Expression of glutathione *S*-transferase (GST)–IQD5 and *in vitro* CaM binding assays were performed according to Levy *et al.* (2005). Generation of IQD5 pENTR vectors is described in Bürstenbinder *et al.* (2017b). The coding sequence (CDS) of IQD5 was mobilized into the pDEST15 vector (Invitrogen) to generate an N-terminal GST fusion, and GST–IQD5 and the GST control were expressed in the *Escherichia coli* strain KRX (Novagen) upon induction with 0.1% (w/v) rhamnose and 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). Cells were resuspended in CaM pull-down buffer [5.8 mM Tris–HCl, pH 7.3; 2.7 mM KCl; 127 mM NaCl; 0.1% (v/v) Tween 20; 0.002% (w/v) NaN$_3$]. Bovine CaM immobilized on Sepharose beads (GE Healthcare) was incubated with cleared protein extracts in CaM buffer containing either 5 mM EGTA or 1 mM CaCl$_2$. After four steps of washing, the last washing fraction and the bead fraction were collected, and, together with the unbound fraction, separated by SDS–PAGE. GST-tagged proteins were visualized by immunoblot analysis using a horseradish peroxidase (HRP)-coupled α-GST antibody (Santa Cruz).

Statistical analysis

Statistical analysis of root length, cotyledon and leaf area, and *IQD5* expression was performed using ANOVA implemented in the R software, followed by a Tukey's post-hoc test, and Benjamini–Hochberg adjustment of *P*-values. For statistical analysis of fluorescence intensities at convex and concave sides of pavement cells in *pIQD5::IQD5-GFP* seedlings, a *t*-test was performed. Statistical analysis of PC shape features was performed using the Kruskal–Wallis test, followed by a Dunn's post-hoc test and Benjamini–Hochberg adjustment of *P*-values, which is part of the R script provided in the PaCeQuant package.

Results

IQD5 *is expressed in vegetative tissues*

To identify *in planta* sites of IQD5 function, we determined spatio–temporal expression domains of *IQD5* in transgenic *pIQD5$_{short}$::GFP-GUS* and *pIQD5$_{long}$::GFP-GUS* reporter lines, in which a 1207 bp and 2201 bp DNA fragment upstream of the translational start site of the *IQD5* gene were fused to the reporter, respectively. Histochemical GUS analysis of *pIQD5$_{short}$::GFP-GUS* lines throughout development revealed strong promoter activity in cotyledons and leaves, in the vasculature of leaves and the hypocotyl, as well as in the shoot apical meristem (Fig. 1A). In roots, GUS staining was detectable mostly in older parts of the root. In root tips, *IQD5* promoter activity was restricted to the lateral root cap of primary and lateral root meristems. GUS activity was largely absent from reproductive organs, such as flower buds, flowers, siliques, and seeds, and during embryo development. The GUS patterns are consistent with developmental *IQD5* expression data obtained from publicly available microarray data sets (Fig. 1B) (Winter *et al.*, 2007), which confirm higher *IQD5* expression levels in vegetative tissues when compared with reproductive tissues. Similar expression patterns were observed in *pIQD5$_{long}$::GFP-GUS* lines (Supplementary Fig. S1 available at *JXB* online), suggesting that the 1207 bp fragment was sufficient to report authentic *IQD5* expression patterns. Our analysis thus reveals preferential expression of *IQD5* in vegetative tissues of shoots and roots.

IQD5–GFP localizes to cortical microtubules

To examine the subcellular localization of IQD5, we generated a fluorescent protein fusion construct, in which GFP was fused to the C-terminus of IQD5 within a genomic fragment containing the native *IQD5$_{short}$* promoter (Fig. 2). The *pIQD5::IQD5-GFP* construct was introduced into an *iqd5* knockout background to avoid dosage effects of *IQD5* copy number (Fig. 2A, B). We obtained two independent Arabidopsis T-DNA insertion lines for *IQD5*, which we termed *iqd5-1* and *iqd5-2* (Fig. 2A). RT–PCR analysis revealed the complete absence of full-length *IQD5* transcripts in *iqd5-1* and *iqd5-2* lines when compared with the WT, demonstrating that both T-DNA insertion lines are null mutant alleles (Fig. 2B). Based on macroscopic examination, both *iqd5* mutants were phenotypically indistinguishable from WT plants, as shown for root length and shoot growth (Supplementary Fig. S2). The *iqd5-1* mutant was transformed with the *pIQD5::IQD5-GFP* construct by *Agrobacterium*-mediated floral dip. qRT-PCR analysis of steady-state *IQD5* mRNA levels revealed comparable expression in two independent *pIQD5::IQD5-GFP/iqd5-1* complementation lines, which was moderately higher than in the reference WT (Fig. 2C).

We investigated the subcellular localization of IQD5–GFP by confocal imaging and observed that IQD5 localized in punctate patterns along filamentous structures at the cell cortex of hypocotyl cells, reminiscent of cortical MTs (Fig. 2D). Treatment with oryzalin, a drug that binds to tubulin subunits and prevents MT polymerization (Morejohn *et al.*, 1987), abolished IQD5–GFP localization to filaments, while MTs remained intact upon mock treatment (Fig. 2D). Transgenic *pCaMV 35S::GFP-MAP4* (Marc *et al.*, 1998) and *pCaMV 35S::GFP-ABD2* (Sheahan *et al.*, 2004; Wang *et al.*, 2004) lines were included as controls for the MT and actin cytoskeleton, respectively (Fig. 2D). While oryzalin treatment disrupted MTs decorated with GFP–MAP4, labeled (GFP–ABD2) actin filaments remained intact, demonstrating the efficiency and specificity of the treatment. Co-expression of *pCaMV 35S::YFP-IQD5* with *pCaMV 35S::*RFP-TUA5 in transient expression assays in *N. benthamiana* further corroborated co-localization of YFP–IQD5 with MTs (Fig. 2E). GFP fluorescence was very weak in hypocotyls of *pIQD5::IQD5-GFP/iqd5-1* seedlings. Moderately stronger IQD5–GFP fluorescence was detectable at cortical MT arrays in epidermal PCs of cotyledons (Fig. 2F). PCs adopt highly complex jigsaw puzzle-like shapes with interlocking lobes and necks. Within individual PCs, IQD5–GFP accumulated at the convex side of necks at the interface of anticlinal and outer periclinal walls, as indicated by increased average fluorescence intensities when compared with the concave side (Fig. 2F, G).

Loss of IQD5 *causes aberrant PC shape*

To assess whether IQD5 contributes to growth regulation of PCs, we analyzed PC shapes in cotyledons of *iqd5* mutants and of the two *pIQD5::IQD5-GFP/iqd5-1* lines. Cell outlines were visualized by PI staining, and groups of PCs on the adaxial
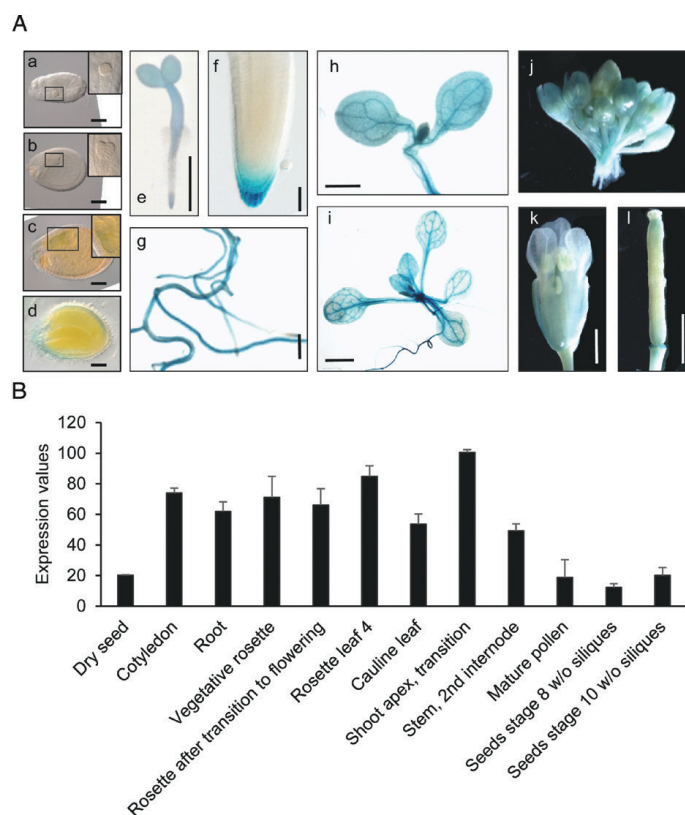
Fig. 1. *IQD5* expression analysis. Whole-mount histochemical GUS staining of *ProIQD5$_{short}$:GFP-GUS* reporter lines (A) in seeds and embryos of globular (a), heart (b), torpedo (c), and mature (d) stage, in 2-day-old seedlings (e); insets show close-ups of embryos, indicated by black boxes (a–c), in the primary root meristem (f), in lateral roots (g), and in cotyledons (h) of 5-day-old seedlings, in the shoot of 10-day-old seedlings (i), and in flower buds (j), flowers (k), and siliques (l) of 5-week-old plants. Scale bars represent 100 μm (a–d), 1 mm (e, g–l), and 10 μm (f). *In silico* expression data of *IQD5* in different tissues and organs were obtained from the publicly available eFP browser database (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi, last accessed 19 November 2018) (B). Data show mean values ±SD from three independent biological experiments.

side of cotyledons from 5-day-old seedlings were imaged by confocal microscopy (Fig. 3A). PC shape features were quantified with PaCeQuant (Möller *et al.*, 2017), an ImageJ-based open source tool for fully automatic quantification and graphical visualization of PC shape features. As evidenced by similar areas of individual cells and similar area distributions, expansion was largely unaffected in *iqd5* mutants (Fig. 3B). Cell shapes on the other hand differed strongly in both *iqd5* mutant alleles when compared with the WT or the two independent *pIQD5::IQD5-GFP/iqd5-1* lines. Mutants displayed an increased cellular circularity (Fig. 3C). Circularity values range between 0 and 1, where a circularity value of 1 represents a perfect circle. Increased circularity thus indicates reduced cellular complexity in *iqd5* mutants. Reduced cellular complexity correlates with a moderately reduced average number of lobes from 15 lobes per cell in WT to 13 lobes per cell in *iqd5* mutants (Fig. 3D). In addition, *iqd5* mutants displayed a strongly reduced growth of lobes, indicated by an ~30% reduction of average lobe length (Supplementary Fig. S3A). The width of the cellular core region, measured as the maximum (Fig. 3E) core width, was increased by 22%. Maximum core width provides a clearly defined value as an estimate for the

growth restriction of the cellular core region (Möller *et al.*, 2017), which is similar to neck width values manually quantified by, for example, Fu *et al.* (2005). An increased maximum core width indicates reduced growth restriction at neck regions. The phenotypic differences were highly similar between *iqd5-1* and *iqd5-2* mutant alleles. Expression of *pIQD5::IQD5-GFP* in the *iqd5-1* mutant background restored PC shape to WT-like patterns (Fig. 3; Supplementary Fig. S3), which demonstrates functionality of the IQD5–GFP fusion protein and sufficiency of the amplified promoter region for restoring *IQD5* expression levels. Collectively, our data suggest that IQD5 is required for lobe initiation, lobe growth, and growth restriction at neck regions of cotyledon PCs, which is consistent with its predominant localization to cortical MT arrays at necks.

Cell shape defects in iqd5 *mutants occur early during cotyledon development*

Morphogenesis of PCs in cotyledons is established during distinct phases (Fu *et al.*, 2002; Zhang *et al.*, 2011). In the early phase, 1–3 DAG, lobe formation is initiated and cells start to expand anisotropically, which is followed (3–7 DAG) by diffuse

189
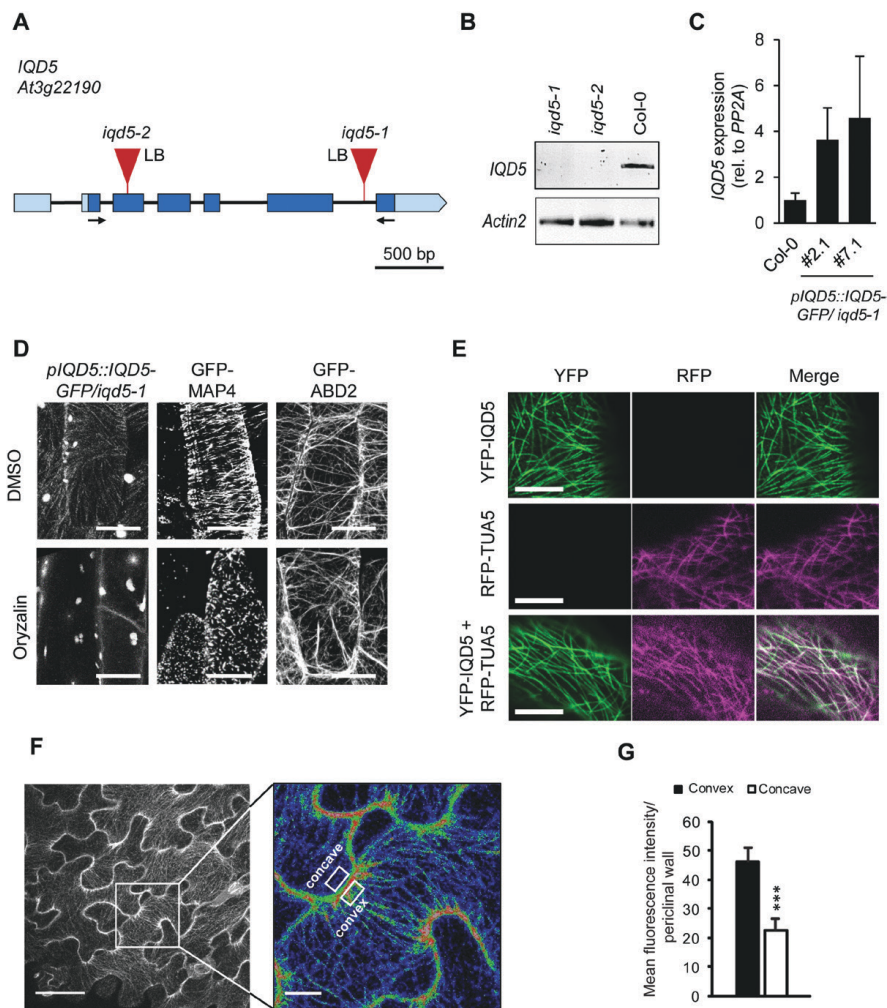
Fig. 2. Subcellular localization of IQD5–GFP in transgenic Arabidopsis *pIQD5::IQD5-GFP/iqd5-1* lines. Gene model and position of T-DNA insertions in two independent mutant lines, *iqd5-1* and *iqd5-2* (A). Boxes indicate the 5'UTR and 3'UTR (light blue), and exons (dark blue). Introns are represented by the black line. Loss of *IQD5* full-length transcript in *iqd5-1* and *iqd5-2* plants compared with the WT (Col-0) was validated by RT–PCR (B). Arrows in (A) indicate the position of primers used for amplification of *IQD5* transcripts. *Actin2* was included as a control for cDNA integrity. Relative *IQD5* expression levels were analyzed by qRT-PCR in two independent transgenic *pIQD5::IQD5-GFP/iqd5-1* lines (#2.1 and #7.1) compared with the WT (C). Data show mean values ±SD of three independent biological experiments. Subcellular localization of IQD5–GFP in hypocotyls of transgenic *pIQD5::IQD5-GFP/iqd5-1* seedlings after mock (DMSO) or oryzalin (10 µM) treatment; scale bars=20 µm (D). Subcellular localization of YFP–IQD5 (top), RFP–TUA5 (middle), and of YFP–IQD5 and RFP–TUA5 (bottom) in transient (co-)expression assays in leaves of *N. benthamiana* (E). Transgenic *pCaMV 35S::GFP-MAP4* and *pCaMV 35S::GFP-ABD2* seedlings were included as controls for the microtubule and actin cytoskeleton, respectively. *Z*-stack images of GFP fluorescence in epidermis pavement cells of cotyledons from 5-day-old *pIQD5::IQD5-GFP/iqd5-1* seedlings (F). Overview images (left column) and close-up of lobe regions (right column; fluorescence intensities are shown by false coloring, red, high fluorescence intensity; blue, low fluorescence intensity). Scale bars=50 µm and 5 µm, respectively. Mean fluorescence intensities measured at the convex and concave side of lobe regions in the upper periclinal wall (G). Data show mean values ±SD from a total of 50 lobes, quantified in 10 cells from five seedlings.

growth and expansion of shape patterns. Growth ceases at later stages (10–18 DAG), and PCs as well as cotyledons reach their final size (Belteton *et al.*, 2018). To determine at which stage IQD5 functions, we studied PC shape in *iqd5* mutants during cotyledon development and imaged WT and *iqd5* mutant seedlings at 2, 3, 5, 7, and 10 DAG (Fig. 4A). During development, the average cell size increased (Supplementary Figs S4–S8), which is consistent with earlier reports (Zhang *et al.*, 2011;

Möller *et al.*, 2017). To examine shape and geometries in cell populations of similar sizes, referred to as small, medium, and large, we applied the size thresholds of t_s=1400 µm^2 and t_m of 4040 µm^2 at 3, 5, 7, and 10 DAG. These thresholds were experimentally determined in our previous work (Möller *et al.*, 2017), and resemble cells at early (small), intermediate (medium), and late stages of cellular expansion. To distinguish between very small (tiny) and small cell populations in
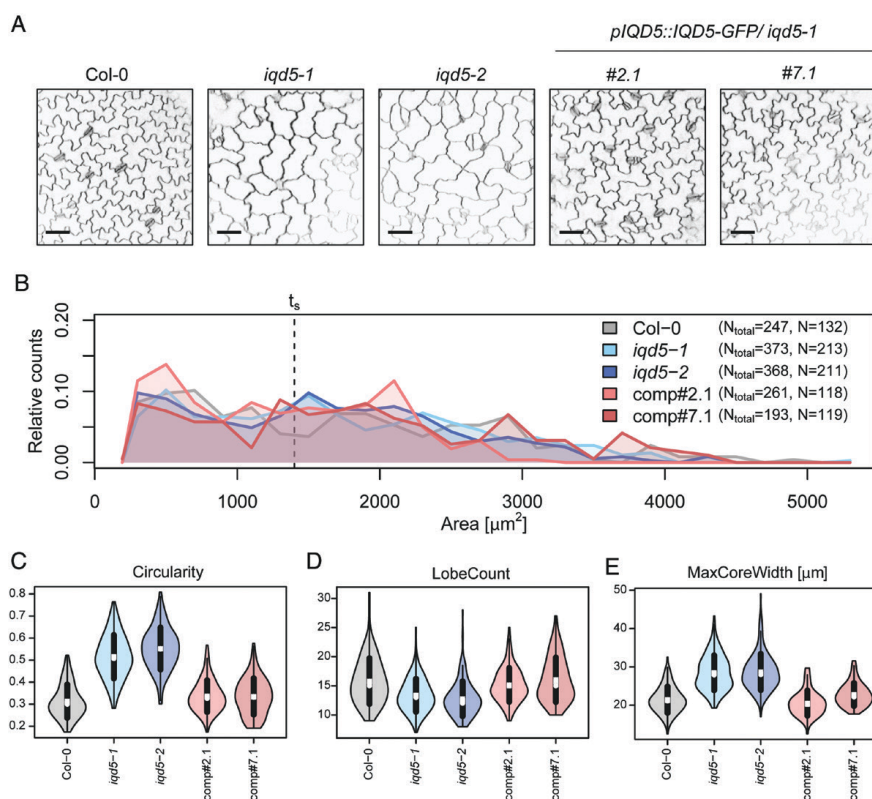
Fig. 3. Pavement cell (PC) morphology on the adaxial side of cotyledons from 5-day-old seedlings of the wild type (Col-0), two independent *iqd5* knockout lines (*iqd5-1* and *iqd5-2*), and two independent transgenic complementation lines (*pIQD5:IQD5-GFP/iqd5-1*; lines #*2.1* and #*7.1*). Representative images of PC morphology (A). Cell outlines were visualized with PI. Images are single optical sections. Scale bars=50 µm. Quantification of cell shape features by PaCeQuant (B–H). Relative distribution of cell areas in the analyzed genotypes (B). Numbers in the key refer to the total number of cells from 10 images of the different genotypes. Cells larger than size threshold t_s=1400µm^2 were used for further analysis. Violin plots of feature distributions for circularity (C), lobe count (D), and maximum core width (MaxCoreWidth, E). Circles and crosses refer to medians and means; the vertical black lines represent the SD (thick lines) and the 95% confidence intervals (thin lines). The width of each violin box represents the local distribution of feature values along the *y*-axis. For an overview of all shape features and statistical analysis, see Supplementary Fig. S3.

cotyledons at 2 and 3 DAG, a time span during which lobe formation and anisotropic expansion are initiated, we included an additional size threshold of t_{tiny}=240 µm^2. Quantification of PC shape features revealed first differences in cell shapes of *iqd5* mutants already at 2 and 3 DAG (Fig. 4D, E). When compared with the WT, cellular circularity was moderately but significantly increased in both *iqd5* mutant alleles in tiny and in small-sized cell populations (Fig. 4D), and margin roughness, a measure for the (ir-)regularity of local curvature values along the cell contour, as well as average basal lobe length were reduced (Fig. 4E; Supplementary Fig. S4). Similar results were observed in seedlings at 3 DAG (Supplementary Fig. S5). In medium to large-sized cell populations, analyzed in cotyledons between 5 and 10 DAG, phenotypic differences became more pronounced with increasing cell size (Fig. 4F; Supplementary Figs S6–S8). The time series analysis thus suggests important roles for IQD5 already during early phases of PC morphogenesis in cotyledons. Analysis of *pIQD5$_{short}$::GFP-GUS* and *pIQD5::IQD5-GFP/iqd5-1* lines revealed promoter activity and accumulation of IQD5–GFP at neck regions, respectively,

in cotyledons and in the shoot apical meristem between 2 and 10 DAG (Fig. 4B, C). Thus, our data demonstrate that *IQD5* is expressed early during cotyledon development and that loss of *IQD5* causes reduced lobe initiation and anisotropic expansion during early growth phases.

IQD5 regulates PC shape during embryogenesis and post-embryonic growth

Cotyledons resemble true leaves in many aspects, and thus provide a convenient system to study leaf development (Tsukaya *et al.*, 1994). However, while cotyledons emerge in embryogenesis, true leaves post-embryonically differentiate from the shoot apical meristem and, unlike cotyledons, differ in their final leaf shape (Tsukaya, 2002). Moreover, some mutations affect exclusively the development of cotyledons or true leaves (Tsukaya, 1995). To test if *IQD5* also functions in true leaf development, we analyzed PC shape in rosette leaves of 3-week-old plants (Fig. 5A). Morphologically, the first two true leaves in Arabidopsis are similar to cotyledons (Poethig, 1997; Kerstetter and Poethig, 1998), and phenotypes in some rosette leaf-specific
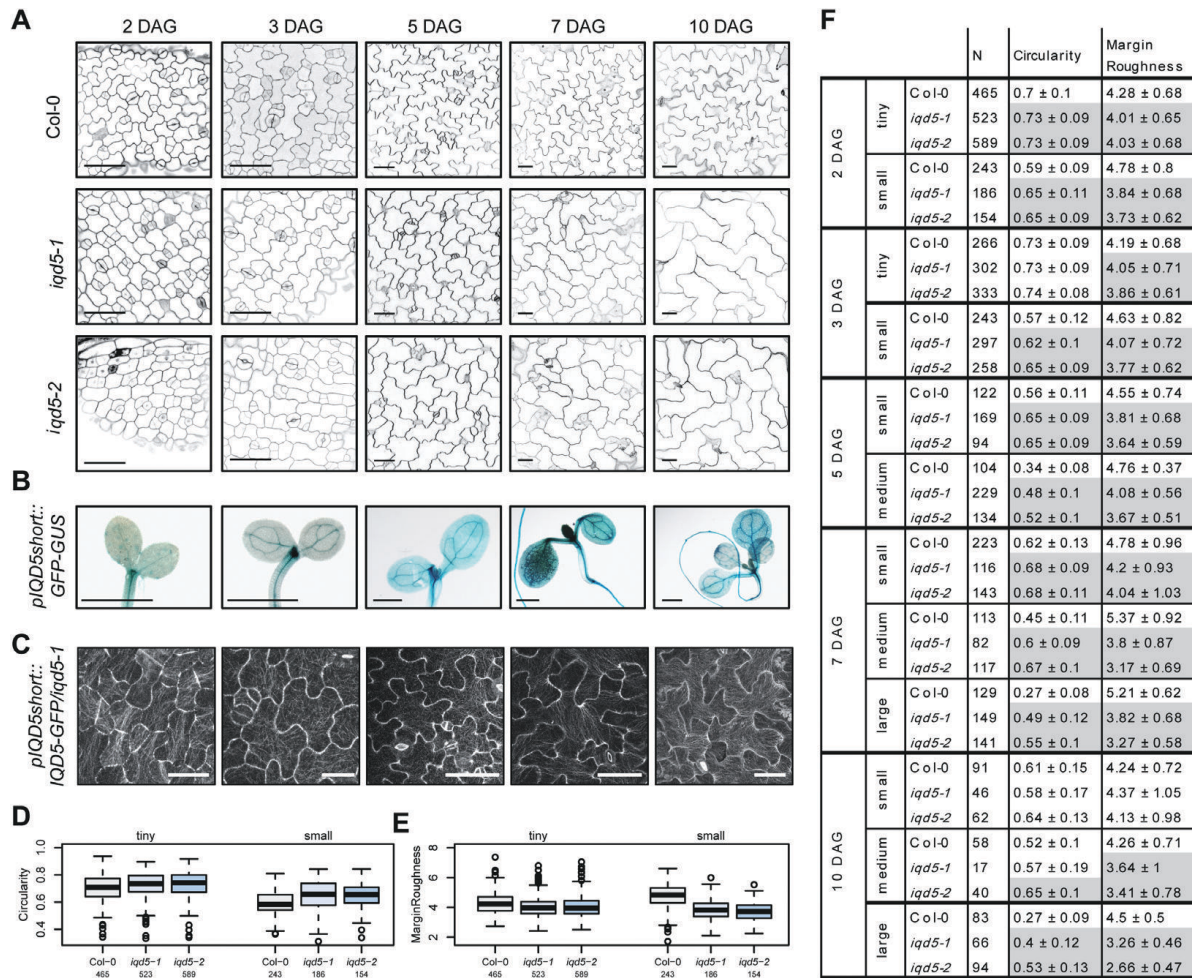
			N	Circularity	Margin Roughness
2 DAG	tiny	Col-0	465	0.7 ± 0.1	4.28 ± 0.68
		iqd5-1	523	0.73 ± 0.09	4.01 ± 0.65
		iqd5-2	589	0.73 ± 0.09	4.03 ± 0.68
	small	Col-0	243	0.59 ± 0.09	4.78 ± 0.8
		iqd5-1	186	0.65 ± 0.11	3.84 ± 0.68
		iqd5-2	154	0.65 ± 0.09	3.73 ± 0.62
3 DAG	tiny	Col-0	266	0.73 ± 0.09	4.19 ± 0.68
		iqd5-1	302	0.73 ± 0.09	4.05 ± 0.71
		iqd5-2	333	0.74 ± 0.08	3.86 ± 0.61
	small	Col-0	243	0.57 ± 0.12	4.63 ± 0.82
		iqd5-1	297	0.62 ± 0.1	4.07 ± 0.72
		iqd5-2	258	0.65 ± 0.09	3.77 ± 0.62
5 DAG	small	Col-0	122	0.56 ± 0.11	4.55 ± 0.74
		iqd5-1	169	0.65 ± 0.09	3.81 ± 0.68
		iqd5-2	94	0.65 ± 0.09	3.64 ± 0.59
	medium	Col-0	104	0.34 ± 0.08	4.76 ± 0.37
		iqd5-1	229	0.48 ± 0.1	4.08 ± 0.56
		iqd5-2	134	0.52 ± 0.1	3.67 ± 0.51
7 DAG	small	Col-0	223	0.62 ± 0.13	4.78 ± 0.96
		iqd5-1	116	0.68 ± 0.09	4.2 ± 0.93
		iqd5-2	143	0.68 ± 0.11	4.04 ± 1.03
	medium	Col-0	113	0.45 ± 0.11	5.37 ± 0.92
		iqd5-1	82	0.6 ± 0.09	3.8 ± 0.87
		iqd5-2	117	0.67 ± 0.1	3.17 ± 0.69
	large	Col-0	129	0.27 ± 0.08	5.21 ± 0.62
		iqd5-1	149	0.49 ± 0.12	3.82 ± 0.68
		iqd5-2	141	0.55 ± 0.1	3.27 ± 0.58
10 DAG	small	Col-0	91	0.61 ± 0.15	4.24 ± 0.72
		iqd5-1	46	0.58 ± 0.17	4.37 ± 1.05
		iqd5-2	62	0.64 ± 0.13	4.13 ± 0.98
	medium	Col-0	58	0.52 ± 0.1	4.26 ± 0.71
		iqd5-1	17	0.57 ± 0.19	3.64 ± 1
		iqd5-2	40	0.65 ± 0.1	3.41 ± 0.78
	large	Col-0	83	0.27 ± 0.09	4.5 ± 0.5
		iqd5-1	66	0.4 ± 0.12	3.26 ± 0.46
		iqd5-2	94	0.53 ± 0.13	2.66 ± 0.47

Fig. 4. PC morphology during cotyledon development. Representative images of epidermis cells on the adaxial side of cotyledons of WT, *iqd5-1*, and *iqd5-2* seedlings at 2, 3, 5, 7, and 10 days after germination (DAG) (A). Cell outlines were visualized by PI; scale bars=50 μm. Histochemical GUS staining of *pIQD5short::GFP-GUS* seedlings at the indicated time points; scale bars=1 mm (B). Analysis of IQD5–GFP subcellular localization in *pIQD5short::IQD5-GFP/iqd5-1* lines at the indicated time points. Scale bars=20 μm (2 and 3 DAG) and 50 μm (5–10 DAG) (C). Quantification of PC shape parameters by PaCeQuant. Cell populations were grouped according to their size as tiny, 75–240 μm², small, 240–1400 μm², medium 1400–4042 μm², and large ≥4042 μm². Boxplots show feature distributions for circularity (D) and margin roughness (E) in seedlings at 2 DAG. Results are medians; boxes range from the first to third quartile. Feature values for circularity and margin roughness in tiny, small, medium, and large-sized cell populations during cotyledon development (F). Results show mean values ±SD. Statistically significant differences (*P*≤0.05) of *iqd5* mutants relative to the WT are highlighted in gray. For an overview of all shape features and statistical analysis, see Supplementary Fig. S4–S8.

mutants are only visible beyond the second true leaf (Guo *et al.*, 2015). To reflect characteristics of true leaves, we thus focused on the third rosette leaf and analyzed PC shape on the adaxial side (Fig. 5C). Quantification of PC shape features revealed similar shape defects in rosette leaves to those observed for cotyledons. Loss of *IQD5* caused a reduced initiation of lobes, as indicated by reduced lobe counts in *iqd5-1* and *iqd5-2* mutants when compared with the WT (Fig. 5F), and the average lobe length of *iqd5* mutants was reduced (Fig. 5G). Reduced formation and growth of lobes were additionally reflected by increased circularity values (Fig. 5D), as well as reduced margin roughness (Fig. 5E) in PCs of *iqd5* mutant plants. Values of minimum (Supplementary Fig. S9) and maximum core width (Fig. 5H) increased, indicative of reduced growth restriction at

neck regions. IQD5 thus controls lobe growth and anisotropic expansion in cotyledons and true leaves. In agreement with functions of IQD5 in true leaves, histochemical GUS activity was detectable in *pIQD5short::GFP-GUS* lines within the entire leaf, indicating that *IQD5* is expressed throughout PC growth (Fig. 5B). Taken together, our data identify IQD5 as a novel regulator of leaf epidermis PC shape, which controls growth restriction at necks in embryonic and post-embryonic tissues.

Reduced growth restriction correlates with altered cellulose deposition

MTs guide CSCs and determine the deposition and direction of newly forming cellulose fibrils in the cell wall (Paredez
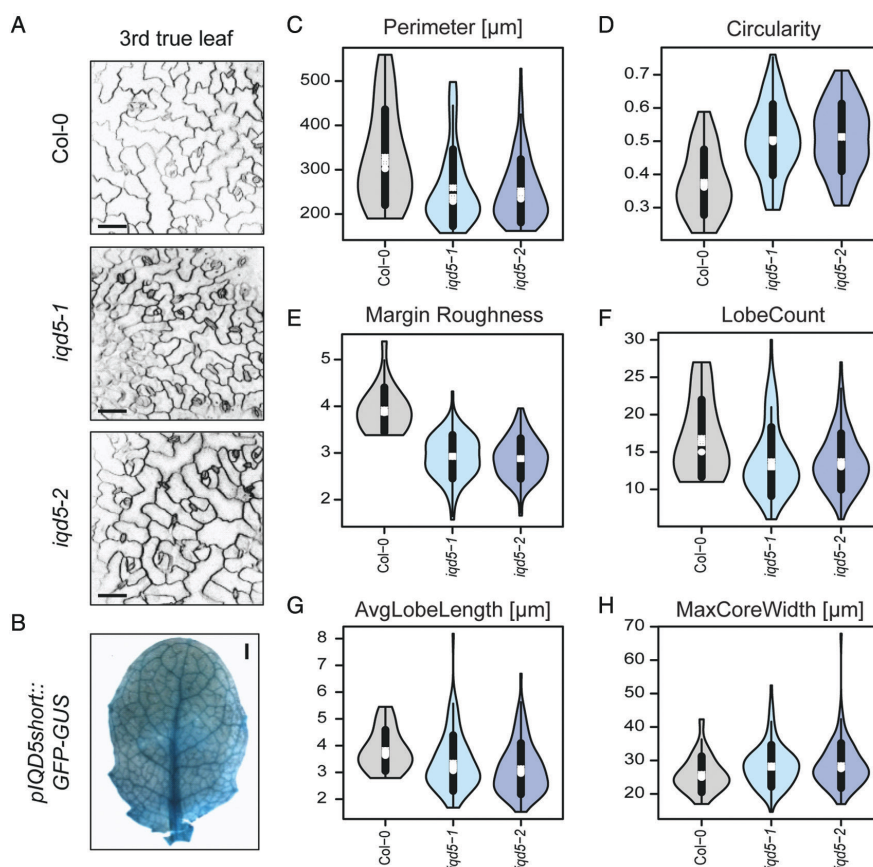
Fig. 5. PC shapes in the epidermis of rosette leaves of the WT and *iqd5* mutants. Images are single optical sections of PI-labeled epidermis cells on the adaxial side of the third rosette leaf in 3-week-old plants; scale bars=50 µm (A). Whole-mount GUS staining of the third rosette leaf in 3-week-old *pIQD5short::GFP-GUS* plants; scale bar=1 mm (B). Quantification of PC shape features. Violin plots show feature distributions from *n*=23–110 cells from 9–13 images for perimeter (C), circularity (D), margin roughness (E), lobe count (F), average lobe length (G), and maximum core width (H). Circles and crosses refer to medians and means; the vertical black lines represent the SD (thick lines) and the 95% confidence intervals (thin lines). The width of each violin box represents the local distribution of feature values along the *y*-axis. For an overview of all shape features and statistical analysis, see Supplementary Fig. S9.

et al., 2006; Gutierrez *et al.*, 2009; Endler and Persson, 2011). During cell expansion, cellulose fibrils are aligned perpendicular to the growth axis and promote anisotropic expansion. Because IQD5–GFP labels MTs, and mutants defective in *iqd5* display shape defects reminiscent of decreased growth restriction at necks, we aimed to investigate whether IQD5 affects cellulose deposition. Staining with calcofluor white, a dye used for visualization of cellulose fibrils (Seagull, 1986; Anderson *et al.*, 2010), revealed reduced staining intensities in *iqd5-1* and *iqd5-2* when compared with the WT, which were reverted to WT levels in the complementation line (Fig. 6A). To assess differences in fluorescence intensities at anticlinal cell walls quantitatively, we segmented the contour of individual cells after visualization of cell walls by co-staining with PI using the segmentation mode implemented in PaCeQuant. We measured an ~45% reduction of calcofluor white fluorescence intensities along the cell contour of *iqd5* mutant cells compared with the WT and the complementation line

(Fig. 6D). Reduced intensities suggest reduced deposition of cellulose in anticlinal cell walls of PCs. Calcofluor white, however, does not discriminate between β-1,3- and β-1,4-glucan chains (Anderson *et al.*, 2010), which are the building blocks of callose and cellulose, respectively. To test whether loss of *IQD5* specifically affects cellulose deposition, we included aniline blue staining to visualize callose (Wood, 1984) and quantified fluorescence intensities. No differences in callose deposition were observed in the mutants when compared with the WT or the complementation lines (Fig. 6B, E). Similarly, only minor differences (5–10%) in fluorescence intensities were observed upon auramine O staining, which labels the cuticle (Considine and Knox, 1979) (Fig. 6C, F). Taken together, our data suggest that the reduced calcofluor white signals are not an artifact of reduced penetration or uptake of the dyes due to general defects in cell wall composition, and probably reflect reduced cellulose deposition caused by the loss of *IQD5*.

193
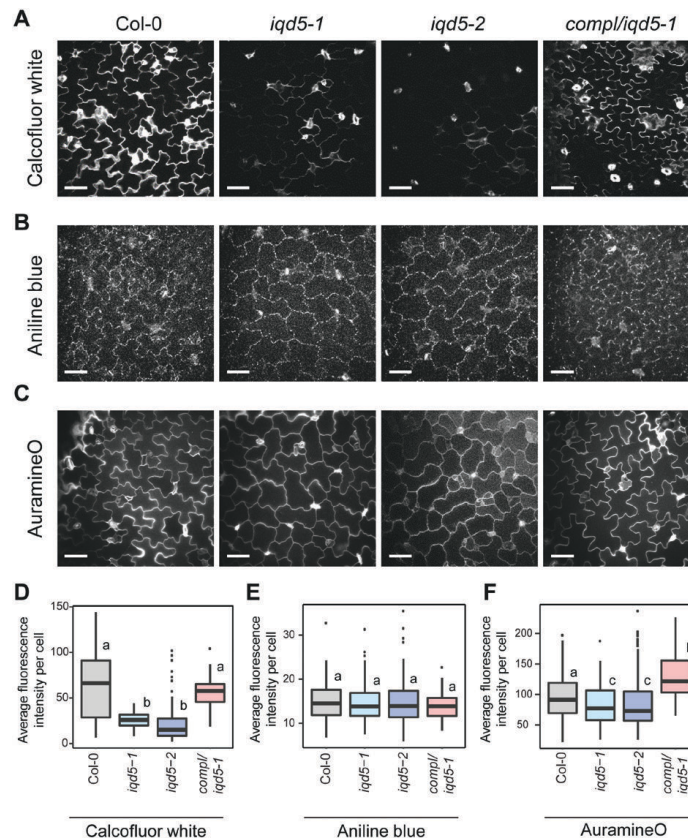
Fig. 6. Histochemical analysis of cell wall composition in the WT, the two *iqd5* mutant alleles *iqd5-1* and *iqd5-2*, and in one transgenic *pIQD5::IQD5-GFP/iqd5-1* line. Confocal images show single optical sections of epidermis pavement cells in cotyledons of 5-day-old seedlings. Cellulose staining by calcofluor white (A), aniline blue staining of callose (B), and auramine O staining of the cuticle (C). Scale bars=50 μm. Quantification of relative fluorescence intensities along the anticlinal cell wall (D–F). Shown are medians; boxes range from the first to third quartile. Different letters denote a statistically significant difference by one-way ANOVA, *P*<0.01.

IQD5-dependent recruitment of CaM to cortical microtubules

A hallmark of IQD proteins is the presence of their eponymous IQ67 domain, which contains a repetitive arrangement of predicted CaM-binding motifs (Abel *et al.*, 2005). In IQD5, the IQ67 domain contains three copies each of three different classes of CaM-interacting motifs, including the IQ motif and motifs of the 1-5-10 and 1-8-14 classes, with presumed roles for binding to apo-CaM (IQ) and holo-CaM (1-5-10 and 1-8-14), respectively (Fig. 7A). Homology modeling of IQD5 indicates that the IQ67 domain adopts an α-helical fold (Fig. 7B), similar to the apo-CaM-binding domain of myosin (Houdusse *et al.*, 2006), and potentially interacts simultaneously with more than one CaM polypeptide. To assess whether IQD5 is a functional CaM target, we performed *in vitro* CaM binding assays (Fig. 7C). We expressed GST-tagged IQD5 and the GST core as a control in *E. coli* to investigate interaction with immobilized bovine CaM in the presence (Ca^{2+}) and absence (EGTA) of calcium. GST–IQD5, but not GST, co-sedimented with apo-CaM, and CaM binding of GST–IQD5 was enhanced in the presence of Ca^{2+}. CaM binding thus is independent of the GST tag, and the predicted CaM-binding motifs (Fig. 7A) are functional in mediating interaction with both states of CaM, the Ca^{2+}-free apo-CaM and Ca^{2+}-bound holo-CaM (Fig. 7C). To gain insight into subcellular sites of IQD5 interaction with CaM, we performed BiFC analyses. N-terminal fusions of IQD5 to the N-terminal half of YFP (Y_N–IQD5) were transiently co-expressed with N-terminal fusions of CaM2 to the C-terminal half of YFP (Y_C–CaM2) in *N. benthamiana* leaves by infiltration with *Agrobacterium* harboring the respective plasmids. As controls, we included Y_N and Y_C fusions of TON1 RECRUITMENT MOTIF1 (TRM1), a member of a plant-specific class of MAPs that interacts with TONNEAU1 (TON1) *in planta* (Drevensek *et al.*, 2012). Recovery of YFP fluorescence was visible along the MT lattice between Y_N–IQD5 and Y_C–CaM2, and between Y_N–TRM1 and Y_C–TON1, which served as positive control (Fig. 7D). No fluorescence complementation was detectable in the negative controls, in which Y_N–IQD5 and Y_C–CaM2 were combined with Y_C–TRM1 and Y_N–TRM1, respectively, demonstrating specificity of the BiFC assay. Additionally, CaM binding at MTs was validated in co-expression assays (Fig. 7E). Expression of *pCaMV 35S::mCherry-CaM2* resulted

194
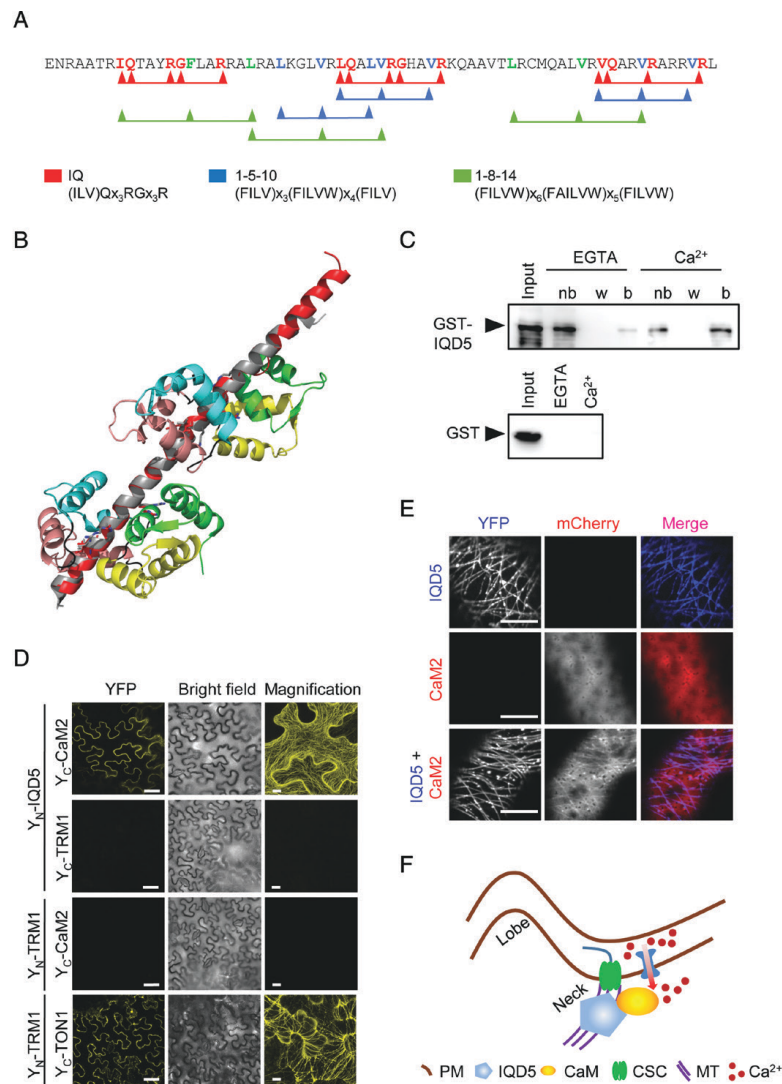
Fig. 7. Calmodulin (CaM) binding properties of IQD5. Amino acid sequence of the IQ67 domain of IQD5 (A). IQ motifs implicated in apoCaM (IQ) and Ca^{2+}–CaM (1-5-10 and 1-8-14) binding are highlighted in red, blue, and green, respectively. Structural alignment of the IQ67 domain of IQD5 (red) and the IQ motif-containing domain of myosin (gray), together with two apo-CaM proteins (first, second, third, and fourth EF hand in green, yellow, salmon, and cyan, respectively (aligned and fitted with PyMol) (B). *In vitro* pull-down of recombinant GST–IQD5 and as control of GST alone expressed in *E. coli* with bovine CaM immobilized on agarose beads in the presence (Ca^{2+}) and absence (EGTA) of calcium; nb, not bound; w, last wash; b, bead-immobilized fraction (C). *In planta* interaction of IQD5 with CaM (D, E). BiFC assays between Y_N–IQD5 and Y_C–CaM2 in leaves of *N. benthamiana* (D). Combinations of Y_N–IQD5 with Y_C–TRM1 and Y_C–CaM2 with Y_N–TRM1 served as negative controls. Y_N–TRM1 and Y_C–TON1 were included as positive control. Single optical sections of YFP fluorescence (left column) and corresponding bright field images (center column). Scale bars=50 μm. Right column, close-up *Z*-stack images of YFP fluorescence; scale bars=10 μm. Subcellular localization of mCherry–CaM2 (top), YFP–IQD5 (middle), and of YFP–IQD5 and mCherry–CaM2 (bottom) in transient (co-)expression assays in leaves of *N. benthamiana* (E). Scale bars=5 μm. Proposed model of IQD5 function in pavement cell morphogenesis (F). IQD5 localizes to microtubules and is required for growth restriction at neck regions, possibly by affecting cellulose deposition along anticlinal cell walls. Interaction of IQD5 with CaM at microtubules points to important roles for Ca^{2+} signaling during shape establishment.

in cytosolic accumulation of mCherry–CaM2, consistent with previous reports (Bürstenbinder *et al.*, 2013). Upon co-expression with YFP–IQD5, mCherry–CaM2 re-localized to cortical MTs (Fig. 7E). Important roles for Ca^{2+} in regulation of PC shape are supported by altered PC morphology in seedlings grown under elevated concentrations of Ca^{2+} (Supplementary Fig. S10). High external Ca^{2+} resulted in increased circularity and an increased minimum and maximum core width, indicative of a reduced growth restriction at neck regions. Thus, our data point to roles of IQD5 in CaM recruitment to cortical

195

MTs, and provide the first indications for CaM-dependent Ca^{2+} signaling in shape development of leaf epidermis PCs (Fig. 7F).

Discussion

Plant-specific IQD families emerged as one of the largest class of CaM targets with proposed roles in linking Ca^{2+} signaling to the regulation of plant growth and development (Bürstenbinder *et al.*, 2017*b*). Despite growing evidence for important roles as cellular scaffolds at the MT cytoskeleton, the precise molecular mechanisms of IQD functions are still enigmatic (Bürstenbinder *et al.*, 2017*a*). Here, we provide experimental evidence that (i) positions IQD5 at cortical MT arrays in vegetative tissues (Figs 1, 2) and (ii) identifies IQD5 as a novel regulator of shape establishment in epidermal PCs of cotyledons and leaves (Figs 3–5). We further show that (iii) phenotypes in *iqd5* mutants correlate with alteration of cell wall properties (Fig. 6) and (iv) provide experimental evidence for interaction of IQD5 with CaM2 *in vitro* and *in planta* (Fig. 7). Thus, our data support roles for IQD5 in regulation of MTs and cellulose deposition during PC morphogenesis. We further provide the first evidence for roles of Ca^{2+} signaling in the spatial co-ordination of cell expansion during interdigitated growth of PCs.

PCs are the most abundant cell type in the leaf epidermis, which are characterized by their jigsaw puzzle-like shape in Arabidopsis and in several other plant species (Ivakov and Persson, 2013; Jacques *et al.*, 2014). Their morphogenesis relies on lobe initiation and anisotropic expansion to generate the complex multi-lobed shapes of PCs (Zhang *et al.*, 2011). Our study establishes IQD5 as a novel regulator of PC shape formation as revealed by reduced lobe growth and growth restriction of neck regions in PCs of *iqd5* mutants. Lobe initiation, however, is only slightly reduced, and cell size as well as overall growth are unaffected, which indicates that IQD5 specifically functions in control of anisotropic expansion of PCs. Consistent with roles during anisotropic expansion, shape defects in *iqd5* mutants are established during early growth phases (i.e. in cotyledons at 2 DAG) at which lobe formation and lobe growth are initiated, and persist during later growth phases, in which cells expand within the lateral cell borders defined during the early growth phase (Zhang *et al.*, 2011). A role for IQD5 in regulating anisotropic expansion is further supported by its expression pattern identified in *pIQD5::GFP-GUS* reporter lines, which revealed uniform promoter activity within cotyledons and leaves. Similarly, growth and shape changes of PCs occur throughout the entire leaf, and growth rates display large heterogeneity between neighboring cells within expanding leaves (Elsner *et al.*, 2012). In contrast, cell cycle activity ceases in a longitudinal gradient during leaf maturation (Asl *et al.*, 2011). At later growth stages, cell division is restricted to the basal part of cotyledons and leaves, as shown by analysis of the cell division marker CYCLINB1;1 (CYCB1;1) in transgenic *pCYC1;1::GUS* reporter lines (Ferreira *et al.*, 1994; Dhondt *et al.*, 2010; Carter *et al.*, 2017). The combined analysis of mutant phenotypes and of spatio-temporal expression domains

thus establishes IQD5 as a novel factor controlling anisotropic growth of leaf epidermis PCs.

The plant MT cytoskeleton plays important roles in regulation of anisotropic expansion during PC morphogenesis (Fu *et al.*, 2002, 2005; Jacques *et al.*, 2014). Disturbances in MT organization, stability, or dynamics by pharmacological agents or by mutations in MAPs, such as TRM2/LONGIFOLIA1 or KATANIN (KTN1), reduce cellular complexity of PC morphogenesis (Lee *et al.*, 2006; Lin *et al.*, 2013; Akita *et al.*, 2015). Potential roles of IQD proteins as MAPs that control cell expansion are supported by our previous work, which revealed altered MT organization and PC shape in transgenic plants ectopically overexpressing *IQD11*, *IQD14*, or *IQD16* (Bürstenbinder *et al.*, 2017*b*). Moreover, altered expression of *IQD* genes is linked to regulation of grain size in rice (Duan *et al.*, 2017; Liu *et al.*, 2017; Yang *et al.*, 2018, Preprint) and fruit shape in tomato (Xiao *et al.*, 2008), water melon (*Citrullus lanatus* L.) (Dou *et al.*, 2018), and cucumber (*Cucumis sativus* L.) (Pan *et al.*, 2018), suggesting that IQDs are key determinants of cell and organ shape (Bürstenbinder *et al.*, 2017*a*). Here, by analysis of GFP fluorescence in transgenic *pIQD5::IQD5-GFP/iqd5-1* lines, we demonstrate subcellular localization of IQD5–GFP to cortical MT arrays in leaf epidermal PCs. Functionality of the GFP-tagged IQD5 protein is indicated by efficient complementation of PC shape defects in *pIQD5::IQD5-GFP/iqd5-1* lines. MT localization was validated by oryzalin treatment in transgenic Arabidopsis plants and by co-localization of YFP–IQD5 with the MT marker RFP–TUA5 in transient expression assays in *N. benthamiana*. Our work thus for the first time identifies a direct link between MTs and cell expansion in an *iqd* knockout mutant. Notably, in contrast to most reported mutants with defects in PC shape that have pleiotropic effects, including reduced plant growth, organ twisting, or swelling of cells (Qian *et al.*, 2009), *iqd5* mutants are macroscopically indistinguishable from the WT. Specific defects in PC morphogenesis thus indicate limited functional redundancy and compensation between the 33 IQD family members in Arabidopsis and point to unique roles for IQD5 in PC morphogenesis.

Quantification of fluorescence intensities along the outer periclinal cell wall suggests accumulation of IQD5–GFP at convex sides of indenting neck regions. Our findings are consistent with *iqd5* mutant phenotypes and patterns of subcellular IQD5 localization reported in an independent study, which was published while our work was under revision (Liang *et al.*, 2018). Similar to IQD5–GFP, preferential accumulation of MT bundles at neck regions has been reported in several studies, and MT accumulation correlates with reduced growth in the neck regions (Fu *et al.*, 2005; Sampathkumar *et al.*, 2014; Armour *et al.*, 2015). Cortical MTs serve as tracks for PM-localized CSCs and thereby determine the direction of cellulose deposition and consequently cell expansion (Endler and Persson, 2011). Here, using histochemical staining, we show that *iqd5* mutants accumulate reduced amounts of cellulose in their anticlinal walls while callose deposition and cuticle formation are unaffected. Plants with impaired cellulose deposition, for example upon cellulase treatment or in mutants of the

196

cellulose synthase AtCesA1, display strongly reduced lobing of PCs (Higaki *et al.*, 2016; Majda *et al.*, 2017), similar to phenotypes in *iqd5* lines. Our data suggest that IQD5 is required for efficient cellulose deposition, for example by controlling MT dynamics or organization, thereby affecting cellulose synthesis. In agreement with this hypothesis, Liang *et al.* (2018) reported increased disorder of cortical MTs in *iqd5* mutants when compared with the WT and provide first evidence for functions of IQD5 in stabilization of MTs. Alternatively, IQD5 may mediate coupling of cellulose synthase movement to MT tracks, possibly by direct interaction with KINESIN LIGHT CHAIN-RELATED (KLCR)/CELLULOSE MICROTUBULE UNCOUPLING (CMU) family members. KLCRs mediate PM tethering of MTs to stabilize cortical MTs against the pushing forces of CSCs (Liu *et al.*, 2016). Arabidopsis IQD1, IQD2, and IQD23 interact with KLCR family members in yeast, and IQD1 recruits KLCR1 to MTs in transient expression assays in *N. benthamiana* (Mukhtar *et al.*, 2011; Bürstenbinder *et al.*, 2013). Thus, the prospect arises that IQD:KLCR modules collectively co-ordinate MT organization and lateral stability of cortical MTs at the PM–MT nexus.

A hallmark of IQD families is their ability to bind CaM Ca^{2+} sensors, which suggests important roles for IQDs in linking CaM-mediated Ca^{2+} signaling to the regulation of the MT cytoskeleton via as yet unknown mechanisms (Abel *et al.*, 2005, Hepler, 2016; Bürstenbinder *et al.*, 2017b). The phenotypes in *iqd5* mutants, together with the IQD5-dependent recruitment of CaM to cortical MTs, provide the first experimental evidence for functions of Ca^{2+} signaling in PC morphogenesis, probably via CaM/CMLs. Roles for Ca^{2+} during PC morphogenesis are further supported by altered PC shapes in response to elevated exogenous Ca^{2+} supply. Interestingly, IQD5 interacts *in vitro* with both states of CaM, the Ca^{2+}-free and Ca^{2+}-bound apo- and holoCaM, respectively. Similarly, IQD1 and IQD20 interact with apoCaM and holoCaM *in vitro* (Abel *et al.*, 2005; Bürstenbinder *et al.*, 2013), which suggests functionality of the distinct CaM-binding motifs within the IQ67 domain. The repetitive alignment of multiple CaM-binding motifs may facilitate interaction with several CaM/CMLs simultaneously, in which individual CaM-binding motifs differentially contribute to CaM/CML binding. Additionally, CaM/CMLs may exert specific effects on IQD5 depending on their Ca^{2+} occupancy, which adds another potential layer of Ca^{2+}-dependent regulation. Identification of *in planta* IQD5-interacting CaM/CMLs, however, will be challenging because large multigene families of 7 and 50 members code for CaMs and CMLs in Arabidopsis, respectively, and many *CaM/CML* genes are expressed in cotyledons and leaves (Supplementary Fig. S11). Ca^{2+} signals are rapidly generated by exogenous application of several phytohormones, including auxin and cytokinin (Saunders and Hepler, 1981; Vanneste and Friml, 2013), which are key regulators of PC morphogenesis that antagonistically activate Rho-like GTPases from plants (ROPs) in lobes and necks, respectively (Fu *et al.*, 2002, 2005, 2009). IQD5 and related proteins of the IQD family may constitute promising candidates for integrating upstream signals (e.g. from phytohormones) into the reorganization of MT arrays, possibly via phytohormone-induced Ca^{2+} signals (Bürstenbinder

et al., 2017a; this study). Lastly, a recent study by Sugiyama *et al.* (2017) provides first indications for functions of IQD13 in spatial control of ROP signaling domains required for cell wall patterning during vessel development. A similar mechanism might apply to IQD5 during PC shape formation, thereby providing a potential link between phytohormone actions, Ca^{2+} signaling, and ROP GTPase activation. Collectively, our work identifies IQD5 as a novel regulator of PC shape and a potential hub for co-ordination of cellular signaling, cytoskeletal reorganization, and cell wall remodeling. Our work thus provides a framework for future mechanistic studies of cellular signaling networks at the cell wall–PM–MT continuum, which will aid a more holistic understanding of cellular processes guiding shape complexity.

Supplementary data

Supplementary data are available at *JXB* online.

Fig. S1. *IQD5* expression analysis in *pIQD5$_{long}$::GFP-GUS* reporter lines.

Fig. S2. Macroscopic analysis of growth parameters in the WT and *iqd5* mutants.

Fig. S3. Quantification and statistical analysis of PC shape features in 5-day-old seedlings of the WT and *iqd5* mutants.

Fig. S4. Quantification and statistical analysis of PC shape in cotyledons at 2 DAG.

Fig. S5. Quantification and statistical analysis of PC shape in cotyledons at 3 DAG.

Fig. S6. Quantification and statistical analysis of PC shape in cotyledons at 5 DAG.

Fig. S7. Quantification and statistical analysis of PC shape in cotyledons at 7 DAG.

Fig. S8. Quantification and statistical analysis of PC shape in cotyledons at 10 DAG.

Fig. S9. Quantification and statistical analysis of PC shape in true leaves.

Fig. S10. Calcium-dependent changes in PC shape.

Fig. S11. *In silico* expression analysis of Arabidopsis CaM/CMLs.

Acknowledgements

References

Abel S, Savchenko T, Levy M. 2005. Genome-wide comparative analysis of the *IQD* gene families in *Arabidopsis thaliana* and *Oryza sativa*. BMC Evolutionary Biology **5,** 72.

Akhmanova A, Hammer JA 3rd. 2010. Linking molecular motors to membrane cargo. Current Opinion in Cell Biology **22,** 479–487.

Akhmanova A, Steinmetz MO. 2008. Tracking the ends: a dynamic protein network controls the fate of microtubule tips. Nature Reviews. Molecular Cell Biology **9,** 309–322.

Akita K, Higaki T, Kutsuna N, Hasezawa S. 2015. Quantitative analysis of microtubule orientation in interdigitated leaf pavement cells. Plant Signaling & Behavior **10,** e1024396.

Anderson CT, Carroll A, Akhmetova L, Somerville C. 2010. Real-time imaging of cellulose reorientation during cell wall expansion in Arabidopsis roots. Plant Physiology **152,** 787–796.

Armour WJ, Barton DA, Law AM, Overall RL. 2015. Differential growth in periclinal and anticlinal walls during lobe formation in Arabidopsis cotyledon pavement cells. The Plant Cell **27,** 2484–2500.

Asl LK, Dhondt S, Boudolf V, Beemster GT, Beeckman T, Inzé D, Govaerts W, De Veylder L. 2011. Model-based analysis of Arabidopsis leaf epidermal cells reveals distinct division and expansion patterns for pavement and guard cells. Plant Physiology **156,** 2172–2183.

Bayer EM, Sparkes I, Vanneste S, Rosado A. 2017. From shaping organelles to signalling platforms: the emerging functions of plant ER–PM contact sites. Current Opinion in Plant Biology **40,** 89–96.

Belteton SA, Sawchuk MG, Donohoe BS, Scarpella E, Szymanski DB. 2018. Reassessing the roles of PIN proteins and anticlinal microtubules during pavement cell morphogenesis. Plant Physiology **176,** 432–449.

Bringmann M, Li E, Sampathkumar A, Kocabek T, Hauser MT, Persson S. 2012. POM-POM2/cellulose synthase interacting1 is essential for the functional association of cellulose synthase and microtubules in Arabidopsis. The Plant Cell **24,** 163–177.

Bürstenbinder K, Mitra D, Quegwer J. 2017*a*. Functions of IQD proteins as hubs in cellular calcium and auxin signaling: a toolbox for shape formation and tissue-specification in plants? Plant Signaling & Behavior **12,** e1331198.

Bürstenbinder K, Möller B, Plötner R, Stamm G, Hause G, Mitra D, Abel S. 2017*b*. The IQD family of calmodulin-binding proteins links calcium signaling to microtubules, membrane subdomains, and the nucleus. Plant Physiology **173,** 1692–1708.

Bürstenbinder K, Rzewuski G, Wirtz M, Hell R, Sauter M. 2007. The role of methionine recycling for ethylene synthesis in Arabidopsis. The Plant Journal **49,** 238–249.

Bürstenbinder K, Savchenko T, Müller J, Adamson AW, Stamm G, Kwong R, Zipp BJ, Dinesh DC, Abel S. 2013. Arabidopsis calmodulin-binding protein IQ67-domain 1 localizes to microtubules and interacts with kinesin light chain-related protein-1. Journal of Biological Chemistry **288,** 1871–1882.

Carter R, Sánchez-Corrales YE, Hartley M, Grieneisen VA, Marée AFM. 2017. Pavement cells and the topology puzzle. Development **144,** 4386–4397.

Chen X, Grandont L, Li H, Hauschild R, Paque S, Abuzeineh A, Rakusová H, Benkova E, Perrot-Rechenmann C, Friml J. 2014. Inhibition of cell expansion by rapid ABP1-mediated auxin effect on microtubules. Nature **516,** 90–93.

Clough SJ, Bent AF. 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. The Plant Journal **16,** 735–743.

Considine JA, Knox RB. 1979. Development and histochemistry of the cells, cell walls, and cuticle of the dermal system of fruit of the grape, *Vitis vinifera* L. Protoplasma **99,** 347–365.

DeLano WL. 2009. PyMOL molecular viewer: updates and refinements. Abstracts of Papers of the American Chemical Society **238**.

Dhondt S, Coppens F, De Winter F, Swarup K, Merks RM, Inzé D, Bennett MJ, Beemster GT. 2010. SHORT-ROOT and SCARECROW regulate leaf growth in Arabidopsis by stimulating S-phase progression of the cell cycle. Plant Physiology **154,** 1183–1195.

Dou J, Zhao S, Lu X, He N, Zhang L, Ali A, Kuang H, Liu W. 2018. Genetic mapping reveals a candidate gene (ClFS1) for fruit shape in watermelon (*Citrullus lanatus* L.). Theoretical and Applied Genetics **131,** 947–958.

Drevensek S, Goussot M, Duroc Y, *et al*. 2012. The Arabidopsis TRM1–TON1 interaction reveals a recruitment network common to plant cortical microtubule arrays and eukaryotic centrosomes. The Plant Cell **24,** 178–191.

Duan P, Xu J, Zeng D, *et al*. 2017. Natural variation in the promoter of *GSE5* contributes to grain size diversity in rice. Molecular Plant **10,** 685–694.

Elsner J, Michalski M, Kwiatkowska D. 2012. Spatiotemporal variation of leaf epidermal cell growth: a quantitative analysis of *Arabidopsis thaliana* wild-type and triple cyclinD3 mutant plants. Annals of Botany **109,** 897–910.

Endler A, Persson S. 2011. Cellulose synthases and synthesis in Arabidopsis. Molecular Plant **4,** 199–211.

Feng L, Chen Z, Ma H, Chen X, Li Y, Wang Y, Xiang Y. 2014. The *IQD* gene family in soybean: structure, phylogeny, evolution and expression. PLoS One **9,** e110896.

Ferreira PC, Hemerly AS, Engler JD, van Montagu M, Engler G, Inzé D. 1994. Developmental expression of the arabidopsis cyclin gene cyc1At. The Plant Cell **6,** 1763–1774.

Fu Y, Gu Y, Zheng Z, Wasteneys G, Yang Z. 2005. Arabidopsis interdigitating cell growth requires two antagonistic pathways with opposing action on cell morphogenesis. Cell **120,** 687–700.

Fu Y, Li H, Yang Z. 2002. The ROP2 GTPase controls the formation of cortical fine F-actin and the early phase of directional cell expansion during Arabidopsis organogenesis. The Plant Cell **14,** 777–794.

Fu Y, Xu T, Zhu L, Wen M, Yang Z. 2009. A ROP GTPase signaling pathway controls cortical microtubule ordering and cell expansion in Arabidopsis. Current Biology **19,** 1827–1832.

Gantner J, Ordon J, Ilse T, Kretschmer C, Gruetzner R, Löfke C, Dagdas Y, Bürstenbinder K, Marillonnet S, Stuttmann J. 2018. Peripheral infrastructure vectors and an extended set of plant parts for the Modular Cloning system. PLoS One **13,** e0197185.

Gehl C, Waadt R, Kudla J, Mendel RR, Hänsch R. 2009. New GATEWAY vectors for high throughput analyses of protein–protein interactions by bimolecular fluorescence complementation. Molecular Plant **2,** 1051–1058.

Guo X, Qin Q, Yan J, Niu Y, Huang B, Guan L, Li Y, Ren D, Li J, Hou S. 2015. TYPE-ONE PROTEIN PHOSPHATASE4 regulates pavement cell interdigitation by modulating PIN-FORMED1 polarity and trafficking in Arabidopsis. Plant Physiology **167,** 1058–1075.

Gutierrez R, Lindeboom JJ, Paredez AR, Emons AM, Ehrhardt DW. 2009. Arabidopsis cortical microtubules position cellulose synthase delivery to the plasma membrane and interact with cellulose synthase trafficking compartments. Nature Cell Biology **11,** 797–806.

Hardham AR, Takemoto D, White RG. 2008. Rapid and dynamic subcellular reorganization following mechanical stimulation of Arabidopsis epidermal cells mimics responses to fungal and oomycete attack. BMC Plant Biology **8,** 63.

Hepler PK. 2005. Calcium: a central regulator of plant growth and development. The Plant Cell **17,** 2142–2155.

Hepler PK. 2016. The cytoskeleton and its regulation by calcium and protons. Plant Physiology **170,** 3–22.

Higaki T, Kutsuna N, Akita K, Takigawa-Imamura H, Yoshimura K, Miura T. 2016. A theoretical model of jigsaw-puzzle pattern formation by plant leaf epidermal cells. PLoS Computational Biology **12,** e1004833.

Horio T, Murata T. 2014. The role of dynamic instability in microtubule organization. Frontiers in Plant Science **5,** 511.

Houdusse A, Gaucher JF, Krementsova E, Mui S, Trybus KM, Cohen C. 2006. Crystal structure of apo-calmodulin bound to the first two IQ motifs of myosin V reveals essential recognition features. Proceedings of the National Academy of Sciences, USA **103,** 19326–19331.

Huang Z, Van Houten J, Gonzalez G, Xiao H, van der Knaap E. 2013. Genome-wide identification, phylogeny and expression analysis of *SUN*, *OFP* and *YABBY* gene family in tomato. Molecular Genetics and Genomics **288,** 111–129.

Hussey PJ, Ketelaar T, Deeks MJ. 2006. Control of the actin cytoskeleton in plant cell growth. Annual Review of Plant Biology **57,** 109–125.

Ivakov A, Persson S. 2013. Plant cell shape: modulators and measurements. Frontiers in Plant Science **4,** 439.

Jacques E, Verbelen JP, Vissenberg K. 2014. Review on shape formation in epidermal pavement cells of the Arabidopsis leaf. Functional Plant Biology **41,** 914–921.

Karimi M, Inzé D, Depicker A. 2002. GATEWAY vectors for *Agrobacterium*-mediated plant transformation. Trends in Plant Science **7,** 193–195.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols **10,** 845–858.

Kerstetter RA, Poethig RS. 1998. The specification of leaf identity during shoot development. Annual Review of Cell and Developmental Biology **14,** 373–398.

Kudla J, Becker D, Grill E, Hedrich R, Hippler M, Kummer U, Parniske M, Romeis T, Schumacher K. 2018. Advances and current challenges in calcium signaling. New Phytologist **218,** 414–431.

Lee YK, Kim GT, Kim IJ, Park J, Kwak SS, Choi G, Chung WI. 2006. *LONGIFOLIA1* and *LONGIFOLIA2*, two homologous genes, regulate longitudinal cell elongation in *Arabidopsis*. Development **133,** 4305–4314.

Levy M, Wang Q, Kaspi R, Parrella MP, Abel S. 2005. Arabidopsis IQD1, a novel calmodulin-binding nuclear protein, stimulates glucosinolate accumulation and plant defense. The Plant Journal **43,** 79–96.

Liang H, Zhang Y, Martinez P, Rasmussen CG, Xu T, Yang Z. 2018. The microtubule-associated protein IQ67 DOMAIN5 modulates microtubule dynamics and pavement cell shape. Plant Physiology **177,** 1555–1568.

Lin D, Cao L, Zhou Z, Zhu L, Ehrhardt D, Yang Z, Fu Y. 2013. Rho GTPase signaling activates microtubule severing to promote microtubule ordering in Arabidopsis. Current Biology **23,** 290–297.

Lincoln C, Britton JH, Estelle M. 1990. Growth and development of the axr1 mutants of Arabidopsis. The Plant Cell **2,** 1071–1080.

Liu J, Chen J, Zheng X, et al. 2017. GW5 acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. Nature Plants **3,** 17043.

Liu Z, Persson S, Zhang Y. 2015. The connection of cytoskeletal network with plasma membrane and the cell wall. Journal of Integrative Plant Biology **57,** 330–340.

Liu Z, Schneider R, Kesten C, Zhang Y, Somssich M, Zhang Y, Fernie AR, Persson S. 2016. Cellulose–microtubule uncoupling proteins prevent lateral displacement of microtubules during cellulose synthesis in Arabidopsis. Developmental Cell **38,** 305–315.

Lloyd C, Hussey P. 2001. Microtubule-associated proteins in plants—why we need a MAP. Nature Reviews. Molecular Cell Biology **2,** 40–47.

Locascio A, Blázquez MA, Alabadí D. 2013. Dynamic regulation of cortical microtubule organization through prefoldin–DELLA interaction. Current Biology **23,** 804–809.

Lu Q, Li J, Ye F, Zhang M. 2015. Structure of myosin-1c tail bound to calmodulin provides insights into calcium-mediated conformational coupling. Nature Structural & Molecular Biology **22,** 81–88.

Majda M, Grones P, Sintorn IM, et al. 2017. Mechanochemical polarization of contiguous cell walls shapes plant pavement cells. Developmental Cell **43,** 290–304.e4.

Marc J, Granger CL, Brincat J, Fisher DD, Kao Th, McCubbin AG, Cyr RJ. 1998. A GFP–MAP4 reporter gene for visualizing cortical microtubule rearrangements in living epidermal cells. The Plant Cell **10,** 1927–1940.

Möller B, Glaß M, Misiak D, Posch S. 2016. MiToBo—a toolbox for image processing and analysis. Journal of Open Research Software **4,** e17.

Möller B, Poeschl Y, Plötner R, Bürstenbinder K. 2017. PaCeQuant: a tool for high-throughput quantification of pavement cell shape characteristics. Plant Physiology **175,** 998–1017.

Morejohn LC, Bureau TE, Molè-Bajer J, Bajer AS, Fosket DE. 1987. Oryzalin, a dinitroaniline herbicide, binds to plant tubulin and inhibits microtubule polymerization in vitro. Planta **172,** 252–264.

Mukhtar MS, Carvunis AR, Dreze M, et al. 2011. Independently evolved virulence effectors converge onto hubs in a plant immune system network. Science **333,** 596–601.

Oda Y. 2018. Emerging roles of cortical microtubule–membrane interactions. Journal of Plant Research **131,** 5–14.

Pan YP, Liang XJ, Gao ML, Liu HQ, Meng HW, Weng YQ, Cheng ZH. 2018. Round fruit shape in WI7239 cucumber is controlled by two interacting quantitative trait loci with one putatively encoding a tomato SUN homolog. Theoretical and Applied Genetics **130,** 573–586.

Paredez AR, Somerville CR, Ehrhardt DW. 2006. Visualization of cellulose synthase demonstrates functional association with microtubules. Science **312,** 1491–1495.

Poethig RS. 1997. Leaf morphogenesis in flowering plants. The Plant Cell **9,** 1077–1087.

Qian P, Hou S, Guo G. 2009. Molecular mechanisms controlling pavement cell shape in Arabidopsis leaves. Plant Cell Reports **28,** 1147–1157.

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

Saunders MJ, Hepler PK. 1981. Localization of membrane-associated calcium following cytokinin treatment in Funaria using chlorotetracycline. Planta **152,** 272–281.

Schindelin J, Arganda-Carreras I, Frise E, et al. 2012. Fiji: an open-source platform for biological-image analysis. Nature Methods **9,** 676–682.

Schneider R, Persson S. 2015. Connecting two arrays: the emerging role of actin–microtubule cross-linking motor proteins. Frontiers in Plant Science **6,** 415.

Seagull RW. 1986. Changes in microtubule organization and wall microfibril orientation during in vitro cotton fiber development—an immunofluorescent study. Canadian Journal of Botany **64,** 1373–1381.

Sedbrook JC. 2004. MAPs in plant cells: delineating microtubule growth dynamics and organization. Current Opinion in Plant Biology **7,** 632–640.

Sheahan MB, Staiger CJ, Rose RJ, McCurdy DW. 2004. A green fluorescent protein fusion to actin-binding domain 2 of Arabidopsis fimbrin highlights new features of a dynamic actin cytoskeleton in live plant cells. Plant Physiology **136,** 3968–3978.

Shibaoka H. 1994. Plant hormone-induced changes in the orientation of cortical microtubules—alterations in the cross-linking between microtubules and the plasma-membrane. Annual Review of Plant Physiology and Plant Molecular Biology **45,** 527–544.

Sugiyama Y, Wakazaki M, Toyooka K, Fukuda H, Oda Y. 2017. A novel plasma membrane-anchored protein regulates xylem cell-wall deposition through microtubule-dependent lateral inhibition of Rho GTPase domains. Current Biology **27,** 2522–2528.

Takatani S, Hirayama T, Hashimoto T, Takahashi T, Motose H. 2015. Abscisic acid induces ectopic outgrowth in epidermal cells through cortical microtubule reorganization in *Arabidopsis thaliana*. Scientific Reports **5,** 11364.

Tsukaya H. 1995. Developmental genetics of leaf morphogenesis in dicotyledonous plants. Journal of Plant Research **108,** 407–416.

Tsukaya H. 2002. Leaf development. The Arabidopsis Book **1,** e0072.

Tsukaya H, Tsuge T, Uchimiya H. 1994. The cotyledon—a superior system for studies of leaf development. Planta **195,** 309–312.

Vanneste S, Friml J. 2013. Calcium: the missing link in auxin action. Plants **2,** 650–675.

Wang YS, Motes CM, Mohamalawari DR, Blancaflor EB. 2004. Green fluorescent protein fusions to Arabidopsis fimbrin 1 for spatio-temporal imaging of F-actin dynamics in roots. Cell Motility and the Cytoskeleton **59,** 79–93.

Wasteneys GO. 2002. Microtubule organization in the green kingdom: chaos or self-order? Journal of Cell Science **115,** 1345–1354.

Wasteneys GO, Yang Z. 2004*a*. The cytoskeleton becomes multidisciplinary. Plant Physiology **136,** 3853–3854.

Wasteneys GO, Yang Z. 2004*b*. New views on the plant cytoskeleton. Plant Physiology **136,** 3884–3891.

Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ. 2007. An 'Electronic Fluorescent Pictograph' browser for exploring and analyzing large-scale biological data sets. PLoS One **2,** e718.

Wood P. 1984. Specific interaction of aniline blue with (1,3)-beta-D-glucan. Carbohydrate Polymers **4,** 49–72.

Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E. 2008. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. Science **319,** 1527–1530.

Yang BJ, Wendrich JR, De B, Weijers D, Xue HW. 2018. OsIQD14 regulates grain shape through modulating the microtubule cytoskeleton. bioRxiv doi: 10.1101/275552. [Preprint.]

Zhang C, Halsey LE, Szymanski DB. 2011. The development and geometry of shape change in *Arabidopsis thaliana* cotyledon pavement cells. BMC Plant Biology **11,** 27.

199

Quantification of Actin Structures using Unsupervised Pattern Analysis Techniques

Birgit Möller, Elisabeth Piltz
Institute of Computer Science,
Martin Luther University Halle-Wittenberg, Germany
Email: birgit.moeller@informatik.uni-halle.de

Nadine Bley
Institute of Molecular Medicine,
Martin Luther University Halle-Wittenberg, Germany
Email: nadine.stoehr@medizin.uni-halle.de

Abstract—The analysis of F-actin organization in cells is a key topic in many fields of biomedical research. While standard protocols for imaging immunostained actin are well established, assessment of the resulting microscopy images is most of the time still performed manually and with a high degree of subjectivity. In this paper, we present a new approach for the analysis of actin structures in microscopy images and the quantification of differences and similarities in actin organization between cells. Compared to existing methods, our approach does not require any previous knowledge about the cells or structures to be analyzed. It works in an unsupervised fashion, combining statistical texture measures and clustering techniques. By this, our method yields large flexibility and allows for application in a wide range of experimental scenarios, and also to heterogeneous cell populations. Experimental evaluation on sample data proves that our method yields meaningful results for biomedical investigations.

Keywords—*actin structures; quantification; texture; clustering*

I. INTRODUCTION

In the field of cell biology the actin cytoskeleton is a central issue underlying intensive investigations. On the one hand it is an essential determinant of cell division, and on the other hand it is crucial for cell migration during development and disease. Actin occurs in a monomeric (G-actin) and filamentous (F-actin; also called microfilaments) form in the cell. Thereby, the continuous assembly and disassembly of F-actin filaments (actin dynamics) is the driving force for cell migration. Since a large and growing number of factors can modulate the physical structure of the F-actin system, and since its organization is directly linked to its cellular functions, the analysis of actin filaments and structures within cells under varying conditions and treatments is a key topic in many fields of biomedical research, e.g., tumor cell research ([1], [2]).

The analysis of F-actin organization is usually performed by immunostaining which allows to visualize microfilament structures by microscopy. In Fig. 1 two sample clips of such images are shown which demonstrate how actin filament organization may vary in cells from the same origin, but with different treatments. Today standard protocols are established to acquire high-quality images of actin and related sub-cellular structures. Nevertheless, the analysis of the microscopy images is most of the time still done manually. This naturally limits the number of images that can be evaluated in experiments, and of course adds a high degree of subjectivity to the results.

One of the main reasons for the lack of automatic methods is the great variety in F-actin structures. Besides environmental
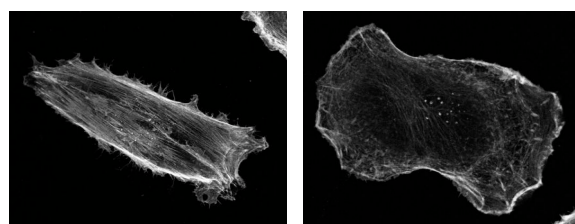
Fig. 1: Image clips illustrating possible variation in actin structure appearance: left, a cell of the control population with high regularity among its long fibers is shown, on the right, one of the IGF2BP1 knock-down population (cf. Sec. IV) with significantly less parallel structures and shorter and unordered actin filaments, but with increased actin along the cell borders.

and cellular factors having an impact on actin characteristics, also cell types by themselves already exhibit a notable variance within microfilament structures. This renders it quite difficult to establish generally applicable techniques for characterizing and quantifying these structures. This is also fortified by the fact, that although by now several software tools for automatic analysis have been released, most of them rely on specific assumptions about the cells to be analyzed like, e.g., contrast or length of actin filaments [3] or their distribution within cells [4], limiting the general applicability of those tools.

In this paper, we present an unsupervised approach for characterizing and quantifying actin structures in microscopy images. Compared to other techniques we do not require any previous knowledge about the actin structures to be analyzed, i.e. neither about specific properties of the cell line nor about fiber characteristics. Moreover, we do not even assume that distinct fibers are present. This allows our method to be applied in a wide range of different scenarios, and also to handle heterogeneous cell populations, resulting in a large flexibility most of the existing approaches cannot provide.

Our method relies on the analysis of textural features and their distributions within cells. This analysis is performed completely unsupervised applying co-occurrence matrices combined with statistical pattern analysis techniques, i.e. clustering approaches and principal component analysis. As result the approach yields an objective ranking of cells according to similarities and differences within their actin structures. By this it fosters biomedical investigations with regard to the cellular functions of actin under varying experimental conditions.

We have implemented our method in Java based on the open source toolkit MiToBo[1], which is compatible with ImageJ[2]. The operator is included in MiToBo 1.4, which is licensed under GPL and publicly available from its website.

The remainder of the paper is organized as follows. After reviewing related work in Section II, we will outline all details of our approach in Section III. Section IV presents results while Section V summarizes key aspects of the paper and gives an outlook on ongoing and future work.

II. RELATED WORK

For the analysis of actin fiber structures in microscopy images different approaches have been proposed. Many of them rely on the explicit segmentation of actin filaments. In [5] a method for quantification of F-actin-containing ruffles is presented, combining thresholding and line detection to segment relevant structures. Subsequently, from these structures features are extracted for classification. Higaki et al. [6] propose an approach to characterize cytoskeletal structures also applying thresholding techniques and skeletonization, combined with subsequent feature extraction and clustering. Both approaches strongly rely on target structures being easily separable from the background of an image. The tool FilaQuant [3] for actin filament quantification builds on explicit fiber segmentation, however, aims to achieve larger robustness by including advanced preprocessing steps to improve image quality. Furthermore, the tool implements a more sophisticated segmentation algorithm adopting a coarse-to-fine strategy. This algorithm provides larger flexibility, but the results are very sensitive to a proper adjustment of several configuration parameters that need to be set in accordance to individual image and fiber characteristics.

To avoid an explicit segmentation of fiber structures in [7], [8] and [9] the use of Fourier transformations followed by an analysis of frequency orientations is suggested. This shifts the focus from local to more global structural patterns, hence, overcomes some of the drawbacks of explicit segmentation. But, the analysis of frequency orientations also works best if prominent fiber structures are present. The applicability of these methods to cell images lacking clear microfilament structures is limited.

The work of Cui et al. [4] also aims at large flexibility in actin fiber quantification. They propose a method based on a sequence of analysis steps. Initially cell areas are segmented into background, cytoplasm and white bright actin regions applying fuzzy clustering using intensity features. Subsequently, cytoplasm regions are further subdivided into stress fiber regions and substrate basically applying intensity thresholding. Finally several features are calculated to quantify differences between cells, e.g., texture measures characterizing the stress fiber regions in the cytoplasm, or the ratio of border bright actin to inner bright actin. Although this method does not explicitly require the adjustment of parameters, the algorithm still implicitly builds on a model of how actin is distributed within cells and its overall appearance, which may be problematic in some scenarios.

---

[1]MiToBo - a Microscope Image analysis ToolBox, website: http://www.informatik.uni-halle.de/mitobo/
[2]ImageJ, website: http://rsb.info.nih.gov/ij/

III. APPROACH

The overall goal of our work is to characterize and quantify relative differences in actin organization between individual cells and populations which have undergone different treatments. To allow for general applicability, this quantification is to be done without requiring any previous knowledge about actin structures and their individual characteristics.

To achieve this goal we propose a method that mainly relies on unsupervised pattern analysis techniques. The basic idea is given by the observation that in microscopy images of cells with immunostained actin structures usually different kinds of structural patterns can be identified within the cells. E.g., in some parts the cells may appear more or less homogeneous, elsewhere long and ordered fibers may dominate, or parts of the cells may appear grained (cf. Fig. 1 and Fig. 4). Cells of the same population, which have undergone the same treatment, usually exhibit similar actin structures except for natural variation. Accordingly, they usually share similar structural patterns in microscopy images. Contrary, if cells have undergone a treatment inducing changes in actin organization, also the patterns should diverge from those of control cells.

Our approach automatically identifies different kinds of structural patterns in the cells of given images, and uses the distributions of these patterns to discover notable similarities and differences in structural appearance between the cells. Since it is not possible to model all different kinds of patterns probably appearing in any experiment or any cell line in advance, the patterns for a given set of cells are extracted online in an unsupervised fashion.

The proposed approach comprises three stages of analysis, which are illustrated in Fig. 2. The input data is given by a set of images with immunostained actin structures. In addition, we currently assume that for each image a segmentation of individual cell regions, i.e. cell masks, is also provided from external. We plan to overcome this requirement in the near future by including an additional stage for automatic cell segmentation into the processing pipeline, which will result in a completely automatic workflow.

In the first stage of our approach (Subsec. III-A) textural features are extracted to characterize local structure in the images. This results in a set of feature vectors for each image and each cell, respectively. In the second stage (Subsec. III-B) these vectors for all given cells and all images are jointly clustered to identify groups of similar feature vectors, which refer to different kinds of structural patterns shared among different cells. Given these groups, for each cell a distribution of the structural patterns appearing in the cell is calculated. Finally in the third stage (Subsec. III-C) these distributions are analyzed applying PCA and hierarchical clustering to identify structural similarities and differences among the cells. Below we will outline the different stages of our approach in detail.

A. Stage I: Textural feature calculation

Initially features to characterize local structural patterns occurring within given cells are extracted. We are using Haralick features calculated from gray-level co-occurrence matrices to characterize local structure [10]. The basic idea of these matrices is to quantify image structure by generating statistics
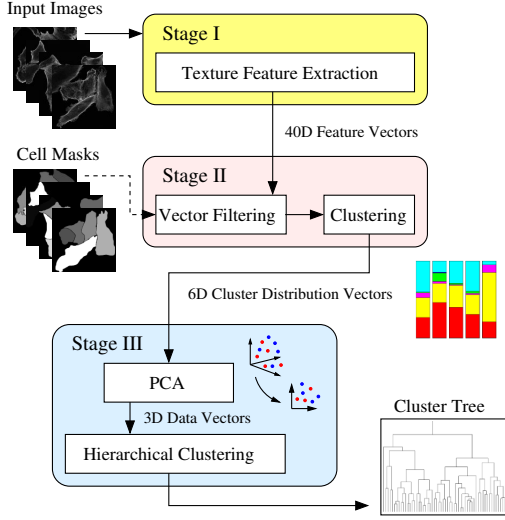
Fig. 2: Overview of our approach with three analysis stages.

over the existence of specific pairs of pixels in a local region of interest. In detail, for an image $I(\vec{p})$ with I_N quantized gray values and for a given region of interest R in the domain of I, for each pair of gray values (i,j) the fraction of all pixel pairs $[\vec{p}_1, \vec{p}_2]$ inside R is determined, for which $\vec{p}_1$ has a gray value of i, $\vec{p}_2$ has a gray value of j, the distance $dist(\vec{p}_1, \vec{p}_2)$ between the two pixels equals δ, and $\vec{p}_2$ lies in the angular direction α of $\vec{p}_1$, i.e. $dir(\vec{p}_1, \vec{p}_2) = \alpha$. Hence, formally the co-occurrence matrix $P_{\delta,\alpha}$ is a matrix of dimension $I_N \times I_N$, parametrized with distance δ and direction α, and calculated over a given region of interest R. Each entry $P_{\delta,\alpha}(i,j)$ for image intensities $i,j \in [0,1,\dots,I_N - 1]$ is calculated as follows:

$$
\begin{aligned}
P_{\delta,\alpha}(i,j) \quad &:= \quad \frac{p_{\delta,\alpha}(i,j)}{\sum_{i,j} p_{\delta,\alpha}(i,j)} \quad \text{with} \quad (1) \\
p_{\delta,\alpha}(i,j) \quad &= \quad card\{\, [\vec{p}_1, \vec{p}_2] \mid dist(\vec{p}_1, \vec{p}_2) = \delta \,\wedge \\
&\qquad dir(\vec{p}_1, \vec{p}_2) = \alpha \wedge I(\vec{p}_1) = i \,\wedge\, I(\vec{p}_2) = j \,\},
\end{aligned}
$$

i.e. $P_{\delta,\alpha}(i,j)$ yields the fraction of pixel pairs in R with the specified properties. Given the statistical data in two matrices calculated over different regions of interest, structural similarities between the two regions can be identified by matrix comparison. To simplify this comparison usually the information contained in each matrix is further condensed to several numerical values, i.e. texture measures, which are easier to handle than complete matrices [10].

In our approach each image is partitioned into a set of non-overlapping quadratic tiles with a certain tile size τ. For each tile and a given pixel distance δ, four co-occurrence matrices are calculated. The four matrices refer to the neighborhood relations between pixels in four different angular directions $\alpha \in \{0°, 45°, 90°, 135°\}$. Then for each matrix 10 customary texture features are calculated, resulting in a 10-dimensional feature vector for each tile. The features currently used are contrast, local homogeneity, entropy, correlation, auto-correlation, dissimilarity, cluster shade, cluster prominence and

maximum probability as defined in [11]. Note that overlapping tiles would yield an increased feature resolution. However, as experiments did not show a clear advantage of a higher resolution for structure quantification, for now the less time-consuming option with non-overlapping tiles is preferred.

Finally, for each tile the four 10-dimensional feature vectors, corresponding to the matrices of the four directions, are concatenated to form a 40-D feature vector. We explicitly do not average the features for the different directions as sometimes suggested. Since in this application structure orientation is obviously essential, we aim to keep a significant amount of anisotropic information in our feature vectors.

B. Stage II: Feature vector clustering

From Stage I for each tile of an image a feature vector results. Several of these tiles are linked to background regions, while others refer to areas where cells are present. As in subsequent stages only tiles linked to cell areas are of interest, at the beginning of Stage II feature vectors linked to background are excluded based on data from the provided cell masks.

The remaining vectors for all images and cells are clustered to discover similar structural patterns among the cells of all images. This is done by k-means clustering where K clusters are to be found. Subsequently each feature vector is labeled with the ID $k \in [1, K]$ of the cluster it is assigned to. From this for each cell a distribution of cluster IDs is calculated quantifying the fraction of its feature vectors belonging to each of the K clusters, i.e. showing a certain structural pattern. The cluster distribution for a cell C is given by

$$
p_C(i) = \frac{1}{T_C} \sum_{j=1}^{T_C} \delta_{i,t_j} \quad \text{with} \quad \delta_{i,j} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases} \quad (2)
$$

for $i = 1, \dots, K$, and with T_C being the total number of tiles in the cell, t_j being the cluster index of the j-th tile, and $\delta_{i,j}$ being the Kronecker delta. At the end of Stage II each cell is characterized by such a discrete cluster distribution (cf. Fig. 3).

C. Stage III: Analysis of cluster distributions

The approach presented here builds on the assumption that cells of a group, e.g., of a control or knock-down population, share similar structural patterns which are different from the ones to be found in other groups. After Stage II for each cell the occurrence of different kinds of patterns in the cell is quantified in terms of a discrete distribution. Consequently, in the last stage this set of distributions is now analyzed to identify subsets of cells sharing similar structural patterns.

Initially we apply a principal component analysis (PCA) to the K-dimensional distribution vectors of all cells in all images to ease exploration of the data in a lower dimensional subspace while at the same time preserving maximal information content in terms of data variance. To this end the eigenvalues and -vectors of the covariance matrix of the distribution vectors are calculated. The eigenvalues $\lambda_i, i = 1, \dots, K$, of the principal components are directly proportional to the amount of variance within the components. We restrict subsequent analysis steps to the first $D \leq K$ principal components with largest eigenvalues

covering more than 95% of data variance, i.e.

$$\frac{\sum_{i=1}^{D} \lambda_i}{\sum_{i=1}^{K} \lambda_i} \geq 0.95 \quad \text{with } \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_K. \quad (3)$$

The distribution vectors are projected into this D-dimensional subspace, and the resulting set of vectors is then clustered. Here hierarchical clustering, i.e. an unweighted average linkage scheme based on Euclidean distances, is used. Compared to k-means this allows for a more detailed investigation of pairwise vector relations, i.e. the resulting cluster tree is well-suited to carve out differences and similarities in structural appearance and actin organization among cells (cf. Fig. 6).

IV. EVALUATION AND DISCUSSION

The experimental evaluation of our method was performed with human melanoma-derived HT144 cells. Using RNAi the RNA-binding protein IGF2BP1 was depleted in some of these cells (denoted by *knock-down*) in comparison to a control population treated with a control siRNA (denoted by *control*). IGF2BP1 promotes mesenchymal cell properties and cell migration by controlling cellular signaling, the expression of transcription factors and F-actin organization in cells ([1], [2]). The F-actin cytoskeleton was labeled by Phalloidin staining and imaged on a Leica SP5X confocal microscope. 10 images of knock-down cells and 10 images of control cells were acquired, each with a size of 2048×2048 pixels. The cell areas in all images were labeled manually[3]. In total all images of the control population contained 35 labeled cells, and the images of the knock-down population 57, resulting in a total number of 92 cells available for the experimental evaluation.

Parameters

The co-occurrence matrices used in Stage I of our approach rely on several parameters (cf. Sec. III-A). Testing various combinations of tile sizes $\tau \in \{16, 32, 64\}$ and pixel distances $\delta \in \{1, 2, 4, 8\}$ on the data set, it turned out that tiles of size 32×32 yield the best compromise to calculate sufficiently local texture features while still having enough pixel pairs available in each tile for meaningful statistics. Initially matrices and feature vectors were extracted for all 4 distances. Given the image size of 2048×2048 pixels and a tile size of $\tau = 32$, for each image 4096 tiles were analyzed. Considering all 4 distances, this results in 4 sets of feature vectors per image each containing 4096 vectors of dimension 40. After eliminating vectors linked to background for each image and distance, on average 1100 vectors with a minimum of 400 and a maximum of 1850 remained to be considered subsequently.

The clustering in Stage II was performed independently for each of the four distances, in each run considering all 92 cells of all images. We used the k-means clustering algorithm of the Weka library with Euclidean distance metric and a-priori data normalization [12]. For the data set at hand a number of $K = 6$ clusters turned out to be reasonable to properly characterize structural patterns. The resulting sets of cluster distribution vectors were also used to restrict the final analysis to a single distance δ. Using the sum of their eigenvalues for the different sets as a measure of information content

---

[3]Some cells were excluded from the analysis, e.g., because their boundaries could not distinctly be identified or they were projected above image borders.
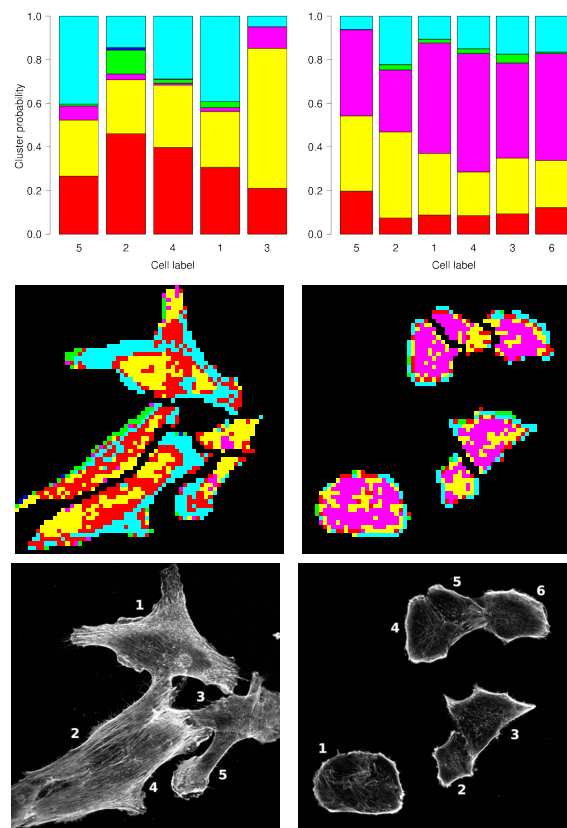
Fig. 3: Top: cell-wise cluster distributions for images of control (left) and knock-down (right) cells. Each color refers to a different cluster ID, each bar to a single cell. Middle: corresponding label images where each pixel represents a tile and its color refers to the cluster ID of its feature vector. Bottom: input images with cell labels shown as white numbers.

in terms of data variance, distance $\delta = 4$ appeared to be favorable, i.e. yielded maximal variance. Hence, in Stage III just the vector set extracted from the co-occurrence matrices with $\delta = 4$ was explored in detail. The hierarchical clustering in that stage was performed using MultiDendrograms [13].

Indeed the settings of some parameters are depending on the concrete data set at hand, e.g., the tile size τ is directly linked to image resolution. Nevertheless the settings used here are reasonable defaults also for other data sets, and the MiToBo operator supports individual adjustment of all parameters if this should be necessary in some scenarios.

Results of Stages I and II: cluster distributions

In Fig. 3 sample cluster distributions resulting from Stages I and II for one image from the control data set (left column) and one from the knock-down set (right column) are shown. Comparing the distributions of both images one can notice striking differences. Feature vectors assigned to the cluster pseudo-colored in red are mainly linked to control cells. Contrary, feature vectors assigned to the cluster colored magenta
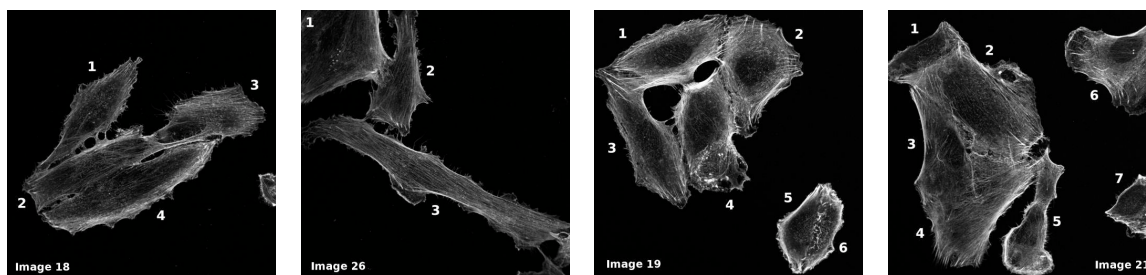
Fig. 4: Left, two control images are shown where some cells share significant similarity with knock-down cells; right, two images of the knock-down population are depicted containing cells with striking fiber structures. The numbers refer to cell labels.

almost exclusively appear in knock-down cells. Moreover, at this stage the different clusters can directly be linked to specific structural patterns. From the original images in the bottom row of Fig. 3 it can be seen that the magenta regions show very few ordered structures, e.g., compared to the cyan and red regions obviously related to areas with longer and more prominent fibers. However, not only between control and knock-down cells differences in cluster distributions can be observed, but also between the cells within a single image. This underlines the necessity of cell-wise evaluation to better account for natural variation, particularly in heterogeneous populations.

Results of Stage III: PCA and hierarchical clustering

At the beginning of Stage III principal component analysis is performed on the cluster distribution vectors. Analyzing the resulting eigenvalues shows that the first three principal components with largest eigenvalues cover approximately 97% of data variance while the first two components still cover $\approx 92.6\%$. In Fig. 5 the projections of the 6D distribution vectors of all 92 cells into the 2D subspace spanned by the first two principal components are shown. The data points related to control cells are shown in blue and the data points of knock-down cells in red. The majority of control and knock-down data points are located in distinct sections of the plot area, i.e. most control points are roughly located on the left while most knock-down points are roughly located on the right. Thus, already in the subspace defined by the first two principal components clear differences in structural appearance, i.e. actin organization, can be observed between both groups. Note that the components of these data points can no longer directly be linked to specific structural patterns as it was possible for the distribution vectors resulting from Stage II.

According to our selection criterion for the subspace dimensionality to be used in the final hierarchical clustering (cf. Eq. 3) we automatically selected the first three principal components as basis for the subspace and projected the 6D distribution vectors into this 3D space. In Fig. 6 the final tree resulting from clustering of the 3D data is shown. On the top level of this tree a separation into two main groups of cells (marked 'A' and 'B') can be observed. The left of these two clusters ('A') mainly refers to cells of the control set (named with 'C' at the beginning), i.e. 22 of the 29 data vectors are linked to control cells. Contrary, the cluster on the right ('B') and in particular the branches marked with 'D' and 'E' mainly refer to knock-down cells. In total 50 of the 63 data vectors in branch 'B' belong to that population. Branch 'C' contains
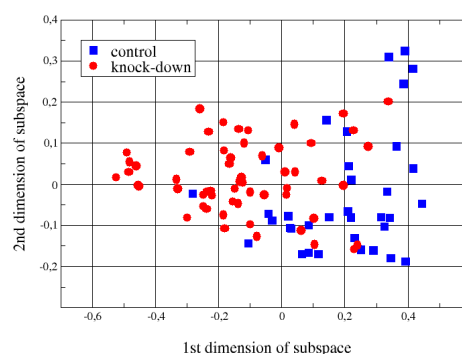
Fig. 5: Plot of the projections of all distribution vectors for all 92 cells into the 2D subspace spanned by the first two principal components with largest eigenvalues of the vector data set.

approximately as many control as knock-down data points, i.e. represents data vectors lying in the transition zone between both groups. Anyway, also here subclusters of cells belonging to either of the two groups can be identified. In general clear differences in actin organization between both groups can be discovered which allows for an objective quantification of structural differences instead of just a subjective visual assessment by manual inspection as it is often done.

The cluster tree also outlines the importance for properly considering natural variation. Within both branches 'A' and 'D' a couple of data points can be identified which seem to have been assigned to the wrong cluster given the original population they belong to. However, the assignment of most of these cells happened due to natural variation within cell answers on changes of cellular conditions. Most of the time not all cells are equally affected by a certain stimulus although the population as a whole might show a clear tendency. Cell division, spontaneously occurring in 10-20% of the cells, strongly affects actin organization. This holds for both populations equally since also among control cells natural variation in actin structures occurs without any treatment.

In Fig. 4 sample images of the control and knock-down population are shown containing some of the cells presumably being assigned to the wrong clusters. On the left, images 18 and 26 of the control set are shown from which, e.g., the cells indexed with 4 (in image 18) and with 1 (in image 26) have been assigned to the knock-down branch 'D' (see yellow leaf markings in Fig. 6). Cell 18-4 obviously shows
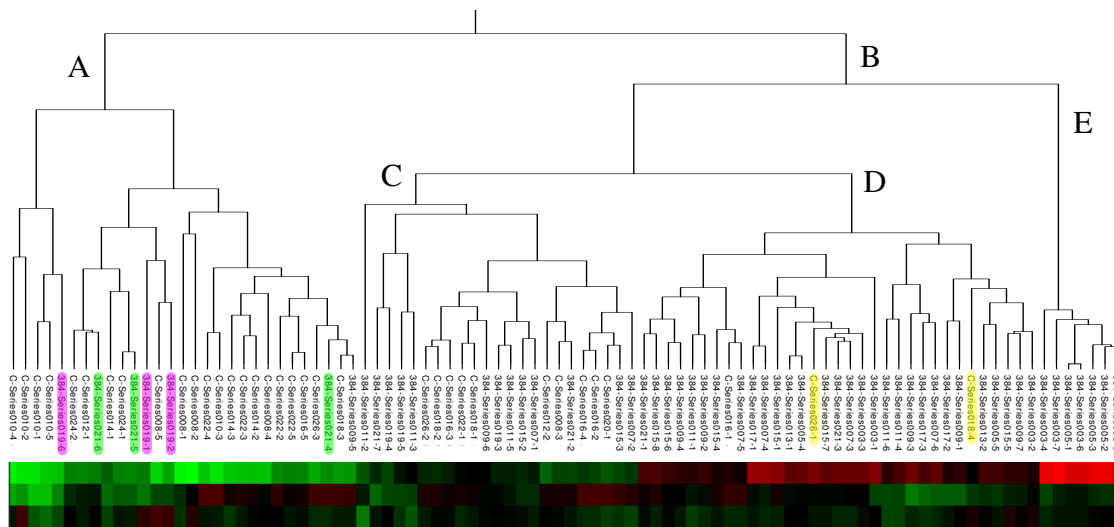
Fig. 6: Result tree of the hierarchical clustering in Stage III showing the cluster assignment for the data vectors of all 92 cells. The color bar at the bottom visualizes the 3D data vectors, negative values are shown in red and positive values in green, the brighter the color the bigger the value. For explanations about branch labels and leaf markings refer to the text.

significant bright actin along its borders which is more typical for knock-down cells (e.g., refer to the cells shown in Fig. 3, bottom right). Cell 26-1 lacks clear fibers, hence, also exhibits increased similarity with knock-down cells. Contrary, in the two knock-down images 19 and 21 on the right several cells exhibit clear fiber structures. Consequently, most of the cells from image 19 have been assigned to the control cluster 'A' on the left (magenta markings in Fig. 6), and also several of the cells in image 21 (green markings). Hence, the assignments are due to natural variation rather than misclassification which underlines the advantages of cell-wise analysis. In conclusion the results discussed here prove that our approach is well-suited to quantify differences in actin organization on the level of global tendencies as well as on single cell level.

V. CONCLUSION

We have presented an approach for quantifying differences in actin organization in cells. The experimental evaluation on sample data proves that our approach is able to quantify differences and similarities in the structure of the microfilament system in microscopy images of cell populations. Clear tendencies in actin organization can be identified, and we can also cope with single cells not following these tendencies due to natural variation. Since there is no previous knowledge used in our method and the analysis is performed unsupervised, our results add a significant degree of objectivity to the analysis of actin structures in images. Our approach also allows for considering larger sets of heterogeneous cell populations, yielding a more meaningful data basis in biomedical investigations. Currently we seek to further improve its usability by including an automatic cell segmentation stage which will overcome the need for external cell mask data. And also the set of texture features is to be optimized with regard to computational load and their discriminative power.

REFERENCES

[1] N. Stöhr *et al.*, "IGF2BP1 promotes cell migration by regulating MK5 and PTEN signaling," *Genes & Development*, vol. 26, no. 2, pp. 176–189, January 2012.

[2] A. Zirkel *et al.*, "IGF2BP1 promotes mesenchymal cell properties and migration of tumor-derived cells by enhancing the expression of LEF1 and SNAI2 (SLUG)," *Nucleic Acids Research*, May 2013.

[3] C. Matschegewski *et al.*, "Automatic actin filament quantification of osteoblasts and their morphometric analysis on microtextured silicon-titanium arrays," *Materials*, vol. 5, no. 7, pp. 1176–1195, 2012.

[4] C. Cui *et al.*, "Quantifying the astrocytoma cell response to candidate pharmaceutical from f-actin image analysis," in *Conf Proc IEEE Eng Med Biol Soc*, 2009, pp. 5768–71.

[5] Q. Yi and M. G. Coppolino, "Automated classification and quantification of f-actin-containing ruffles in confocal micrographs," *BioTechniques*, vol. 40, no. 6, p. 745756, June 2006.

[6] T. Higaki *et al.*, "Quantification and cluster analysis of actin cytoskeletal structures in plant cells: role of actin bundling in stomatal movement during diurnal cycles in arabidopsis guard cells," *The Plant Journal*, vol. 61, pp. 156–165, 2010.

[7] H. Guerin and D. Elliott, "Degeneration affects the fiber reorientation of human annulus fibrosus under tensile load," *Journal of Biomechanics*, no. 8, pp. 1410–1418, 2006.

[8] J. P. Marquez, "Fourier analysis and automated measurement of cell and fiber angular orientation distributions," *International Journal of Solids and Structures*, vol. 43, no. 21, pp. 6413 – 6423, 2006.

[9] E. Sander and V. Barocas, "Comparison of 2d fiber network orientation measurement methods," *Journal of Biomedical Materials Research A*, vol. 88, no. 2, pp. 322–331, February 2009.

[10] R. Haralick, K. Shanmugan, and I. Dinstein, "Texture features for image classification," *Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, pp. 610–621, November 1973.

[11] F. Albregtsen, "Statistical texture measures computed from gray level coocurrence matrices," Image Processing Laboratory, Department of Informatics, University of Oslo, Tech. Rep., November, 5th 2008.

[12] M. Hall *et al.*, "The WEKA data mining software: an update," *SIGKDD Explorations*, vol. 11, no. 1, 2009.

[13] A. Fernández and S. Gómez, "Solving non-uniqueness in agglomerative hierarchical clustering using multidendrograms," *Journal of Classification*, no. 25, pp. 43–65, 2008.

The IQD Family of Calmodulin-Binding Proteins Links Calcium Signaling to Microtubules, Membrane Subdomains, and the Nucleus[1][OPEN]

Katharina Bürstenbinder*, Birgit Möller, Romina Plötner, Gina Stamm, Gerd Hause, Dipannita Mitra, and Steffen Abel

Department of Molecular Signal Processing, Leibniz Institute of Plant Biochemistry, 06120 Halle (Saale), Germany (K.B., R.P., G.S., D.M., S.A.); Institute of Computer Science (B.M.), Biocenter (G.H.), and Institute of Biochemistry and Biotechnology (S.A.), Martin Luther University Halle-Wittenberg, 06120 Halle (Saale), Germany; and Department of Plant Sciences, University of California, Davis, California 95616 (S.A.)

ORCID IDs: 0000-0002-3493-4800 (K.B.); 0000-0002-7146-043X (B.M.).

Calcium (Ca^{2+}) signaling and dynamic reorganization of the cytoskeleton are essential processes for the coordination and control of plant cell shape and cell growth. Calmodulin (CaM) and closely related calmodulin-like (CML) polypeptides are principal sensors of Ca^{2+} signals. CaM/CMLs decode and relay information encrypted by the second messenger via differential interactions with a wide spectrum of targets to modulate their diverse biochemical activities. The plant-specific IQ67 DOMAIN (IQD) family emerged as possibly the largest class of CaM-interacting proteins with undefined molecular functions and biological roles. Here, we show that the 33 members of the IQD family in Arabidopsis (*Arabidopsis thaliana*) differentially localize, using green fluorescent protein (GFP)-tagged proteins, to multiple and distinct subcellular sites, including microtubule (MT) arrays, plasma membrane subdomains, and nuclear compartments. Intriguingly, the various IQD-specific localization patterns coincide with the subcellular patterns of IQD-dependent recruitment of CaM, suggesting that the diverse IQD members sequester Ca^{2+}-CaM signaling modules to specific subcellular sites for precise regulation of Ca^{2+}-dependent processes. Because MT localization is a hallmark of most IQD family members, we quantitatively analyzed GFP-labeled MT arrays in *Nicotiana benthamiana* cells transiently expressing GFP-IQD fusions and observed IQD-specific MT patterns, which point to a role of IQDs in MT organization and dynamics. Indeed, stable overexpression of select IQD proteins in Arabidopsis altered cellular MT orientation, cell shape, and organ morphology. Because IQDs share biochemical properties with scaffold proteins, we propose that IQD families provide an assortment of platform proteins for integrating CaM-dependent Ca^{2+} signaling at multiple cellular sites to regulate cell function, shape, and growth.

Calcium (Ca^{2+}), a general second messenger in all eukaryotes, is required for the execution of developmental programs and the coordination of numerous adaptive responses to external cues, complex processes necessitating a precise regulation of cell growth, and cell shape (Cárdenas, 2009; Steinhorst and Kudla, 2013).

Cellular Ca^{2+} oscillations are monitored by polydentate Ca^{2+} sensors, such as calmodulin (CaM) and calmodulin-like (CML) polypeptides, which differentially interact upon Ca^{2+} binding with a broad range of diverse proteins to modulate their biochemical activities (McCormack and Braam, 2003; McCormack et al., 2005). An extra layer of complexity is added by Ca^{2+}-free apo-CaM that regulates a largely different subset of cellular targets (Jurado et al., 1999). Fundamental processes affected by Ca^{2+} and CaM/CMLs include, among others, transcriptional reprogramming, regulation of metabolism, or the control of cell division and polarity (Hepler, 2005).

The microtubule (MT) cytoskeleton forms a highly dynamic network and plays a central role for coordinating cell growth. Cortical MT arrays are attached to the plasma membrane (PM) for structural support (Liu et al., 2015) and determine the direction of cell expansion by guiding cellulose synthase complexes (CSCs; Endler and Persson, 2011). Furthermore, the MT cytoskeleton mediates the intracellular transport of diverse cargoes (Lloyd and Hussey, 2001; Sedbrook and Kaloriti, 2008) and contributes to exocytosis (Zárský et al., 2009; Idilli et al., 2013; Kong et al., 2015; Zhu et al., 2015a). To generate the various MT arrays, networks of microtubule-associated

proteins (MAPs) control MT dynamics, stability, and organization (Gardiner, 2013; Struk and Dhonukshe, 2014). MAPs also connect CSCs to MTs and are thought to tether cortical MTs to the PM (Endler and Persson, 2011; Bringmann et al., 2012; Liu et al., 2016). Several studies implicated Ca^{2+}-CaM signaling in the control of cytoskeleton organization and dynamics (Hepler, 2016). For example, in yeast and animals, Ras GTPase-activating-like protein IQGAP (IQGAP) scaffold proteins, which recruit CaM via IQ motifs and feature a domain related to GTPase-activating proteins, are key regulators of the cytoskeleton (Shannon, 2012). IQGAPs contribute to the regulation of cell-to-cell contact and coordinate intracellular signaling from membranes to the nucleus (Smith et al., 2015). However, plant genomes do not encode IQGAPs, and the mechanisms of Ca^{2+}-mediated cytoskeletal organization are largely elusive (Hepler, 1992, 2005; Wang et al., 2011). During the colonization of terrestrial habitats, novel adaptive traits evolved in land plants (Graham, 1996), and the size of plant CaM/CML families expanded greatly to provide versatility for transducing complex Ca^{2+} signals into numerous cellular and environmental responses (McCormack et al., 2005; Zhu et al., 2015b). Likewise, the cytoskeleton acquired additional functions for adapting cell shape and cell growth, and a multitude of unique MAPs emerged for generating the various cytoskeletal arrays and controlling their organization and stability (Gardiner, 2013; Struk and Dhonukshe, 2014). Therefore, it is likely that signaling scaffolds analogous to IQGAP proteins evolved in plants for Ca^{2+}-CaM-dependent regulation of the cytoskeleton.

To elucidate CaM-mediated Ca^{2+} signaling in plants, extensive efforts have been made to identify CaM/CML-binding proteins (CaMBPs). More than 300 CaMBPs are currently known, which include transporters and channels, metabolic enzymes, transcription factors, myosins, and various proteins of undefined functions that mostly interact with either holo-CaM or apo-CaM (Reddy et al., 2011). To date, only one CaM-binding MAP has been reported, KINESIN-LIKE CaM-BINDING PROTEIN/ZWICHEL (KCBP/ZWI; Narasimhulu and Reddy, 1998), which functions in trichome development (Hülskamp et al., 1994; Oppenheimer et al., 1997; Tian et al., 2015) and root growth (Buschmann et al., 2015; Humphrey et al., 2015). During cell division, KCBP localizes to the cortical division zone and likely functions in phragmoplast guidance and in the spatial control of cytokinesis (Buschmann et al., 2015). The activity of KCBP/ZWI is regulated by Ca^{2+}-CaM (Kao et al., 2000), and CaM binding abolishes MT interaction of its motor domain (Deavours et al., 1998; Narasimhulu and Reddy, 1998). Because not all effects of CaM/CMLs on MT arrays may be explained by KCBP action, additional players likely are required for the CaM-dependent regulation of MT organization and dynamics (Hepler, 2016).

The IQ67 DOMAIN (IQD) family of mostly uncharacterized plant-specific CaMBPs may represent such candidates at the nexus of Ca^{2+} signaling and cell growth control. IQD proteins are defined by their characteristic IQ domain of 67 conserved amino acid residues, which harbors three different consensus CaM recruitment motifs in a unique and repetitive arrangement, including up to three copies of the IQ motif (Abel et al., 2013). The IQ67 domain is sufficient and required for CaM interaction in vitro, and the founding member, Arabidopsis (*Arabidopsis thaliana*) IQD1 (Levy et al., 2005), localizes to MT arrays and the cell nucleus (Bürstenbinder et al., 2013). The 33 IQD proteins of Arabidopsis are highly diverse in length (103–794 amino acids) and cluster in four major phylogenetic clades that differ mainly in the presence and distribution of family-specific motifs flanking the central IQ67 domain (Abel et al., 2005). Extensive *IQD* families (23–67 genes) have been annotated in rice (*Oryza sativa*; Abel et al., 2005) and recently in purple falsebrome (*Brachypodium distachyon*; Filiz et al., 2013), tomato (*Solanum lycopersicum*; Huang et al., 2013), soybean (*Glycine max*; Feng et al., 2014), poplar (*Populus trichocarpa*; Ma et al., 2014), and moso bamboo (*Phyllostachys edulis*; Wu et al., 2016). Thus, IQD proteins are possibly the largest class of CaMBPs in plants; however, their precise biological roles and biochemical functions remain to be elucidated (Abel et al., 2013; van der Knaap et al., 2014). Coexpression assays demonstrated that IQD1 colocalizes with CaM2 to MT arrays (Bürstenbinder et al., 2013), which suggests a role for IQD1 in linking Ca^{2+}-CaM signaling to the regulation of MT-associated processes (Abel et al., 2013). Additional support for IQD functions at MTs is indirectly provided by studies in tomato, which showed that duplication of the *IQD12/SUN* locus causes elongated fruits, altered cell shape, and twisted stems (Xiao et al., 2008; Wu et al., 2011). These phenotypes are reminiscent of plants with mutations in tubulin subunits (Ishida et al., 2007) and several MAPs (Buschmann et al., 2004; Sedbrook et al., 2004). Collectively, the limited evidence points to regulatory roles of IQD proteins in Ca^{2+}-CaM signaling and plant development.

Here, we present, to our knowledge, the first comprehensive characterization of the entire Arabidopsis IQD family. The 33 IQD proteins differentially localize to multiple subcellular sites, including MTs, the PM, and nuclear compartments. As demonstrated for eight representative IQD family members, the specific IQD localization patterns coincide with the subcellular patterns of IQD-dependent CaM recruitment. Misexpression of select IQD proteins alters the cell shape and pattern of MT organization. Thus, the prospect arises that the plant-specific IQD families provide an assortment of scaffold-like proteins to sequester and integrate CaM-dependent Ca^{2+} signaling at multiple subcellular sites to regulate cell function, shape, and growth.

RESULTS

IQD Family Members Differentially Localize to MTs, Membranes, and the Nucleus

To gain insights into the cellular sites of IQD function, we studied the subcellular distribution of all 33 Arabidopsis

IQD family members after transient expression in *Nicotiana benthamiana* leaves, which is a convenient system in which to monitor protein localization (Deeks et al., 2012). We expressed GFP-tagged IQDs under the control of the cauliflower mosaic virus (*CaMV*) *35S* promoter to facilitate their detection. The authentic GFP reporter was included as a reference, which uniformly labeled the cytoplasm and cell nucleus with the exception of its nucleolus (Fig. 1, bottom right). We confirmed the expression of all full-length IQD fusion proteins by immunoblot analysis using an anti-GFP antibody (Supplemental Fig. S1A). N-terminal GFP fusions of most IQD proteins were associated at least partially with cytoskeletal arrays (Fig. 1). Localization to the MT cytoskeleton was confirmed by coexpression with red fluorescent protein (RFP)-TUBULIN ALPHA5 (TUA5), a marker for MTs

(Gutierrez et al., 2009), as shown for GFP-IQD1 (Fig. 1), GFP-IQD13, and GFP-IQD16 (Supplemental Fig. S1B). About half of the GFP-tagged IQD family members labeled the PM (e.g. IQD12, IQD22, and IQD25) or the nuclear envelope (e.g. IQD8 and IQD27). In addition, several GFP-IQD fusions localized to the nucleus (e.g. IQD10, IQD28, and IQD32), nucleolus (e.g. IQD3, IQD4, and IQD20), or distinct nuclear bodies (e.g. IQD6, IQD9, and IQD24; Fig. 1, insets). Their translocation into the nucleus is likely an active process because the size of most GFP-IQD fusions exceeds the exclusion limit (greater than 50–60 kD) of nuclear pores (Nigg, 1997), and most nucleus-localized IQD proteins contain predicted nuclear localization signals (Supplemental Table S1). The smallest family member, IQD20, lacks a nuclear localization signal; thus, its GFP fusion (38.7 kD) is capable of passive diffusion into the nucleus.

Figure 1. Subcellular localization of Arabidopsis GFP-IQD fusion proteins in *N. benthamiana*. N-terminal GFP fusions of all Arabidopsis IQD family members (IQD1–IQD33) were transiently expressed under the control of the *CaMV 35S* promoter in *N. benthamiana* leaves. Colored bars above the images indicate the phylogenetic clades (Abel et al., 2005). MT colocalization of GFP-IQDs was confirmed by coexpression with RFP-TUA5 (shown for GFP-IQD1), and GFP alone was used as a reference (bottom right image). Micrographs of cells are projections of Z-stacks; insets are single-layer images of cell nuclei. Extra insets (GFP-IQD1 and GFP-IQD4) are single-layer images of MTs, which for GFP-IQD4 reveal MT localization only with increased laser intensities. Bars = 20 μm and 5 μm (insets).
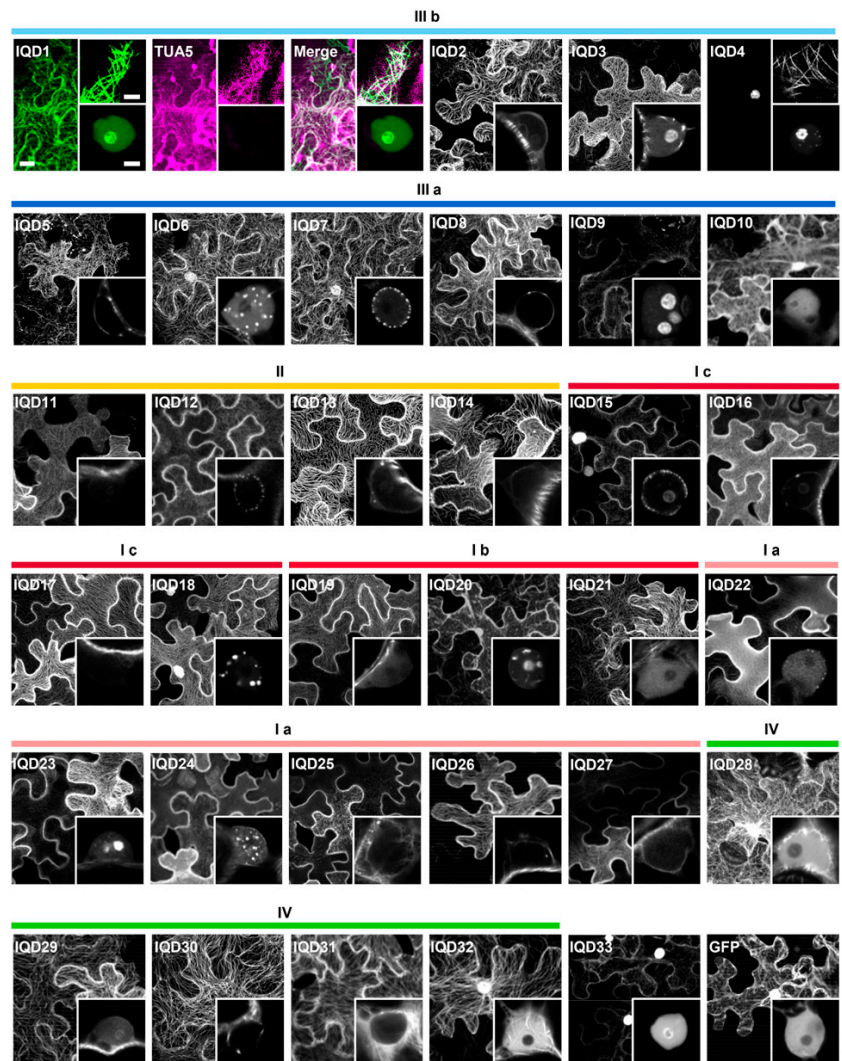
To validate the recorded subcellular localization patterns of N-terminal GFP-IQD fusion proteins, we selected six IQD members for independent localization studies, which represent the distinct localization patterns and the major phylogenetic clades of the Arabidopsis IQD family (i.e. IQD8, IQD14, IQD16, IQD20, IQD25, and IQD28; Fig. 1). We obtained similar results in the *N. benthamiana* system for the six selected C-terminal IQD-GFP fusion proteins, which we expressed under the control of either the *CaMV 35S* promoter (Fig. 2A) or the native *IQD* promoters (Fig. 2B). Transfection of Arabidopsis leaves with the corresponding *Pro-35S:GFP-IQD* constructs and analysis of subcellular GFP-IQD localization further supported the localization results (Fig. 2C). Thus, the subcellular localization of GFP-tagged IQDs is independent of GFP tag configuration, expression level, and experimental system. Expression under the control of native *IQD* promoters, however, resulted in weak fluorescence signals. In combination with in silico analyses (Toufighi et al., 2005), which indicate very low *IQD* mRNA levels (Supplemental Fig. S1C), our data suggest that *IQD* gene expression correlates with low IQD protein abundance. In conclusion, our data provide solid support for the Arabidopsis IQD family being a novel class of CaMBPs with presumed functions at MTs, at the PM, and in nuclear compartments.

A Subset of IQD Proteins Localizes to Distinct PM Subdomains

The targeting of several IQD proteins to MTs as well as to the PM prompted us to study in more detail the distribution of select IQDs displaying a prominent PM localization. Imaging of the upper surface of transiently transformed *N. benthamiana* epidermis cells revealed a nonuniform distribution of GFP-IQD proteins within the PM (Fig. 3A). IQD12 and IQD22 labeled filamentous structures that align along MTs, which is corroborated by coexpression with RFP-TUA5 (Fig. 3B). The accumulation of IQD12 and IQD22 in filamentous structures is abolished upon treatment with the MT-depolymerizing drug oryzalin (Fig. 3C). In addition, all four GFP-IQD fusion proteins mark distinct punctate structures (Fig. 3A). These patterns are reminiscent of membrane subdomains, which are largely immobile compartments within the PM (Saka et al., 2014), and thought to serve as platforms of signal transduction (e.g. during immune signaling or polarized growth; Malinsky et al., 2013; Jarsch et al., 2014). We observed similar localization patterns for two PM subdomain marker proteins of the plant-specific REMORIN (Rem) family, Rem6.6 and Rem6.7 (Jarsch et al., 2014), which we included as controls (Fig. 3A). The pattern of labeled PM domains differed and ranged from a more homogenous labeling (GFP-IQD12) to the decoration of distinct circular structures (GFP-IQD25 and GFP-Rem6.7) with an average size between ~0.11 and
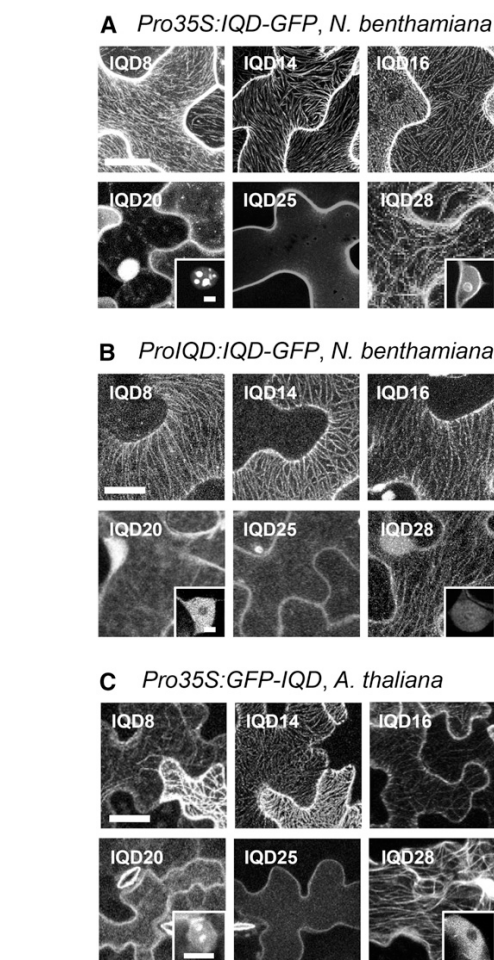
Figure 2. Independent verification of subcellular localization patterns of select Arabidopsis IQD family members. A and B, Subcellular localization of C-terminal GFP fusions of IQD proteins expressed under the control of the *CaMV 35S* promoter (A) or their endogenous regulatory elements (B) in *N. benthamiana* leaf epidermis cells. C, Subcellular localization of N-terminal GFP fusions of IQD proteins transiently expressed under the control of the *CaMV 35S* promoter in Arabidopsis leaf epidermis cells. Micrographs are projections of Z-stacks, and insets are single-layer images of nuclei. Bars = 20 μm and 5 μm (insets).

0.13 μm^2 (Fig. 3, A and D). To assess the lateral stability of GFP-IQD-labeled PM domains, we performed time-lapse experiments and generated kymographs (Fig. 3, E and F). As shown for GFP-IQD25, the labeled domains remained stationary for at least 20 min (Fig. 3E, arrowheads). Similarly, PM domains labeled by the other GFP-IQD fusions were temporally stable, as indicated by continuous vertical lines in the kymographs (Fig. 3F). The high lateral stability of IQD-labeled domains together with the pattern of PM localization suggest that IQD proteins are novel components of PM subdomains in planta.

Figure 3. IQD proteins label PM sub-domains in *N. benthamiana*. A, Imaging of the upper surface of *N. benthamiana* leaf epidermis cells expressing *Pro-35S: GFP-IQD* fusions or the PM subdomain markers *Pro-35S:GFP-Rem6.6* and *Pro-35S:GFP-Rem6.7*. B, Filamentous structures labeled by IQD12 and IQD22 aligned along MTs, as demonstrated by coexpression with RFP-TUA5. C, Depolymerization of MTs by oryzalin treatment abolishes the accumulation of IQD12 and IQD22 in filamentous structures. Bars in A to C = 5 μm. D, Quantification of domain size labeled by GFP-tagged IQD24, IQD25, Rem6.6, and Rem6.7. Data are medians of 10 independent images, and boxes range from first to third data quartiles. E, For kymographs, stacks of 10 to 14 images were acquired over 20 min in intervals of 2 min. Arrowheads indicate individual GFP-IQD25-labeled punctate structures within the PM, which remain stable over 20 min. Bar = 5 μm. F, Kymographs were created from three independent time-lapse movies (1, 2, and 3) of cells expressing GFP-tagged IQD12, IQD22, IQD24, or IQD25. Vertical lines indicate that the PM sub-domains are highly immobile. Bars = 20 μm (horizontal) and 20 min (vertical).
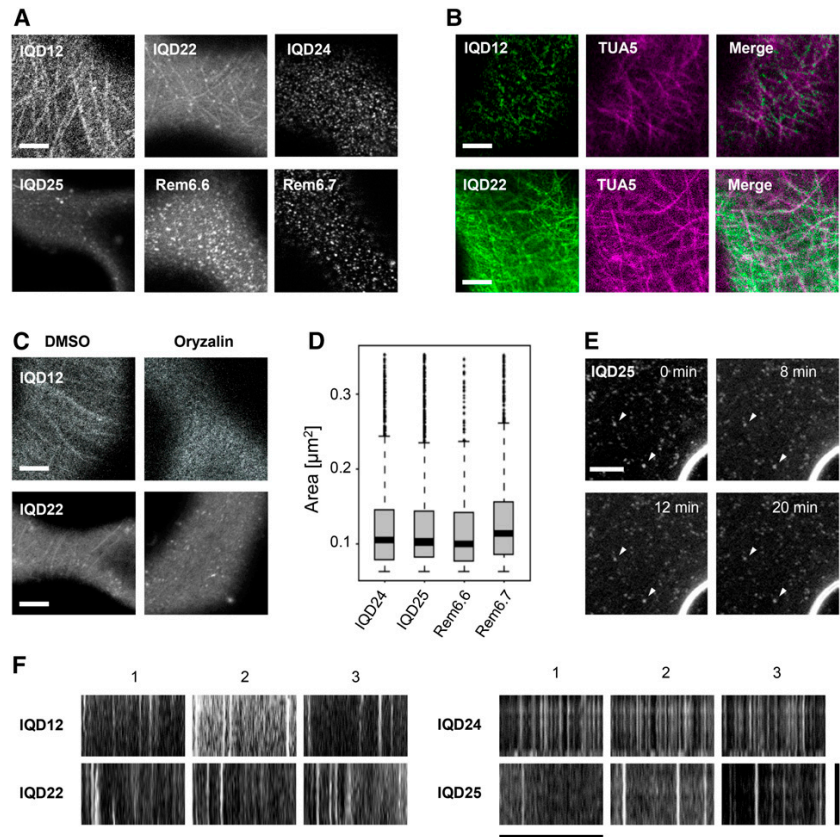
We generated transgenic *Pro-35S:GFP-IQD25* lines to confirm PM localization in Arabidopsis. As observed in *N. benthamiana* leaf cells, GFP-IQD25 decorated the periphery of root cells in a punctate pattern (Fig. 4A). Our data are in agreement with work by Jarsch et al. (2014), who reported the formation of PM subdomains in *N. benthamiana* and Arabidopsis. As expected, GFP-IQD25 fluorescence colocalized upon plasmolysis with the PM and not with the cell wall, which were stained with FM4-64 and propidium iodide (PI), respectively (Fig. 4B). PM localization of GFP-IQD25 was further corroborated by immunogold labeling using an anti-GFP antibody, which indicated enrichment of gold particles at the PM in tissue sections of root apices (Fig. 4, C and D). Interestingly, the shape and size of the rosette leaves were altered in plants overexpressing GFP-IQD25 or untagged IQD25 when compared with wild-type or GFP-expressing control lines, respectively (Fig. 4E). The cotyledons of *IQD25*-overexpressing plants also showed significant alterations in the shape of epidermal pavement cells (Fig. 4, F and G). However, we did not notice visible morphological changes in *iqd25* loss-of-function lines. *ProIQD25:GFP-GUS* expression was detectable in the shoot apical meristem and in the youngest leaves but not in older parts of cotyledons and leaves (Supplemental Fig. S2, A–E).

Thus, the observed phenotypes in *Pro-35S:GFP-IQD25* and *Pro-35S:IQD25* might be a result of ectopic expression. Alternatively, the lack of phenotypes in *iqd25* mutants possibly points to a high degree of functional redundancy within the IQD family.

Quantitative Analysis of GFP-IQD-Labeled MT Patterns

Because MT localization is a hallmark of the Arabidopsis IQD family (Fig. 1) and *IQD* overexpression alters plant development (Xiao et al., 2008; this study), we hypothesized that IQD proteins function in MT-dependent processes. To quantitatively record MT patterns, we adapted an approach used previously to measure actin structures in mammalian cells (Möller et al., 2014). We imaged epidermal cell outlines in the red channel by PI staining or by the expression of an RFP-fused PM-localized protein, PLASMODESMATA-LOCATED PROTEIN1 (PDLP1)-RFP (Amari et al., 2010), and traced GFP-IQD-decorated MT arrays in the green channel. Two-channel maximum projections of Z-stacks covering the upper half of the epidermal cell layer were analyzed, texture feature vectors were extracted from square windows based on local binary patterns (Ojala et al., 2002), and groups of MT patterns
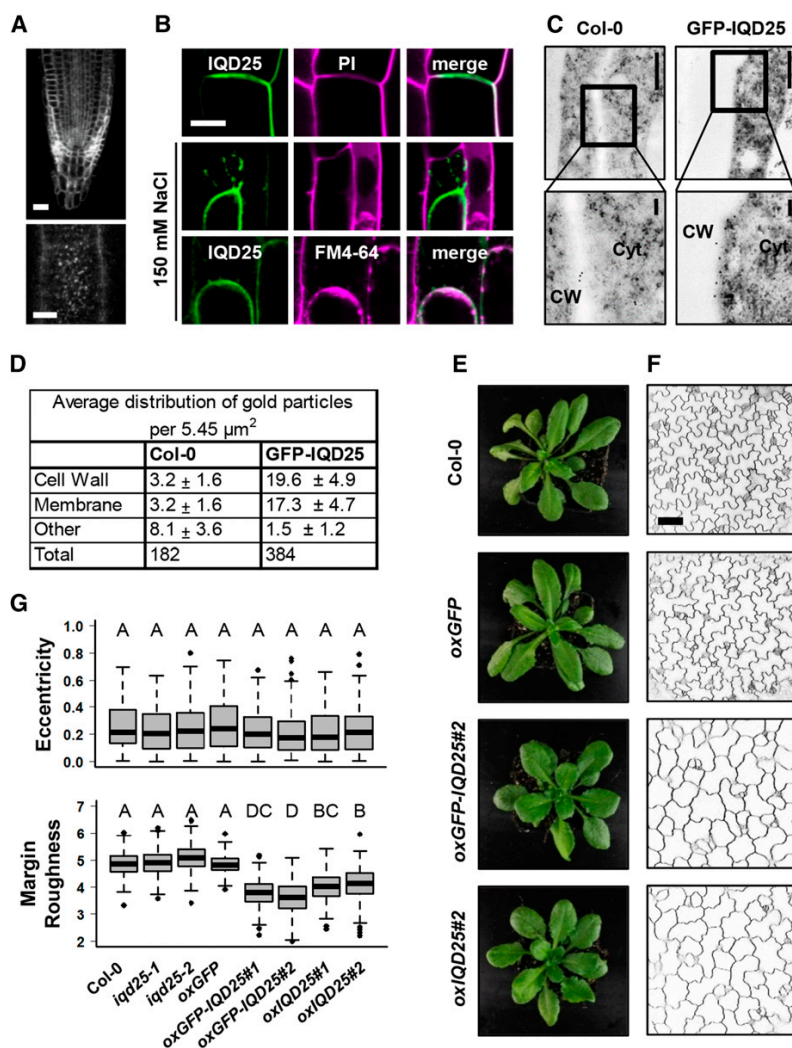
Figure 4. Subcellular localization and phenotypes in transgenic *Pro-35S:GFP-IQD25* Arabidopsis seedlings. A to C, Root cells of 4-d-old transgenic Arabidopsis seedlings expressing GFP-IQD25 under the control of the *CaMV 35S* promoter. A, Subcellular localization of GFP-IQD25 in a primary root tip (top) and surface imaging of root epidermis cells (bottom). Bars = 20 μm (top) and 5 μm (bottom). B, PM localization of GFP-IQD25. GFP-IQD25 localizes to the cell outline, as demonstrated by colocalization with the cell wall dye PI in root cells. After plasmolysis with 150 mM NaCl, GFP-IQD25 fluorescence is detached from PI-stained cell walls and colocalizes with FM4-64-stained PM. C, Localization of GFP-IQD25 by immunogold labeling and transmission electron microscopy. Bottom images are magnifications of the framed regions in the top images. CW, Cell wall; Cyt, cytosol. Bars = 0.5 μm (top) and 0.1 μm (bottom). D, Quantification of gold particles in 10 independent sections. A significant enrichment of gold particles at the PM and cell wall was observed in GFP-IQD25 when compared with the wild-type control (Columbia-0 [Col-0]). E and F, Phenotypes of wild-type, *Pro-35S:IQD25*, *Pro-35S:GFP-IQD25*, and *Pro-35S:GFP* transgenic seedlings. E, Shoots of 4-week-old plants grown on soil under long-day conditions. F, Single optical sections are shown for cotyledon epidermal cells (adaxial side) of 5-d-old seedlings grown under sterile conditions. Cell outlines were visualized with PI. Bar = 50 μm. G, Quantification of cellular elongation (eccentricity) and of the (ir)regularity/(non)smoothness of the cell contour (margin roughness). Results are medians from $n \geq 90$ cells and $n \geq 3$ seedlings, and boxes range from first to third quartiles. Different letters denote a significant statistical difference. $P < 0.005$ by one-way ANOVA.

were defined by cluster analysis (Fig. 5A). The relative distribution of windows belonging to each group was determined for each individual cell. Pairwise distances between the 23 most strongly MT-associated GFP-IQD fusions were visualized in a heat map (Fig. 5B). To validate our approach, we tested the robustness and reliability of the tool using different window sizes (16 × 16 and 32 × 32) and by applying sliding windows of sizes 8 to 32. All samples showed the same trends irrespective of the window settings. Thus, we conclude that our computational pattern-analysis tool is suitable to efficiently and robustly quantify MT array patterns. To assess MT patterns more broadly, we generated a similarity network (Fig. 5C). Our analysis revealed high similarity between MT networks labeled by IQD members of phylogenetic groups II (IQD13 and IQD14, encoded by a sister gene pair), IIIb, and IV, with the

exception of IQD11 and IQD28. A second cluster includes IQD7, IQD8 (encoded by a sister gene pair), and IQD6 (the most closely related member of group IIIa; Abel et al., 2005). In general, we noticed a tendency of phylogenetically related IQDs to decorate similar MT patterns. Notably, MT patterns observed by the overexpression of GFP-tagged IQD11 and IQD16 differed from all others and potentially reflect specialized functions.

Overexpression of *IQD16* Alters MT Organization and Cell Shape

To test our pattern-analysis tool and study the functions of IQD proteins in Arabidopsis, we initially focused on IQD16 because our data indicated MT localization of
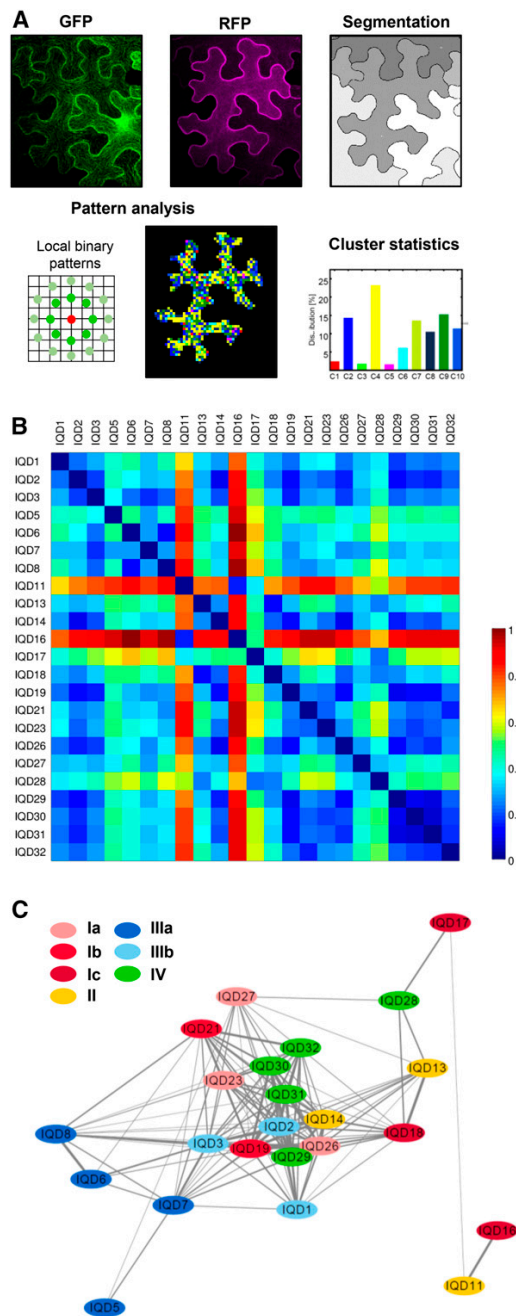
Figure 5. Quantification and network analysis of MT patterns in *N. benthamiana* epidermis cells. A, Workflow of MT pattern analysis. PI-stained cell outlines were imaged in the red channel for cell segmentation, and GFP-IQD-labeled MTs were recorded in the green channel. Texture features were extracted from local binary patterns, and groups of patterns were defined by cluster analysis. B, Heat map showing pairwise distances between MT patterns. The heat map was normalized to a range of [0,1], with blue colors representing high similarity and distances close to 0 and red colors representing high

IQD16 in a highly distinctive pattern (Figs. 1, 2, and 5, B and C). We did not observe phenotypic alterations for *iqd16* T-DNA loss-of-function lines, even after careful examination of tissues in which *ProIQD16:GFP-GUS* expression is detectable (Supplemental Fig. S3, A–G). However, plants overexpressing *IQD16* develop elongated aerial organs with significantly elongated cells (i.e. cotyledons, leaves, and hypocotyls) and display left-handed helical growth (Fig. 6, A–D). Phenotypic strength correlated with *Pro-35S:IQD16* expression (Fig. 6E) and suggests *IQD16* dosage-dependent effects on plant growth. *GFP-IQD16* overexpression lines display similar phenotypes (Fig. 6, A–D), indicating the functionality of the fusion protein. Helical growth aberrations and altered leaf shape were reported for tomato *sun* lines (Wu et al., 2011) and often are associated with defects in MT functions (Buschmann et al., 2004; Ishida et al., 2007). Moreover, the growth defects are reminiscent of plants overexpressing *LONGIFOLIA1/TON1 RECRUITING MOTIF2* (*TRM2*), which encodes a plant-specific MAP (Lee et al., 2006; Drevensek et al., 2012). Hence, the observed phenotypes point to roles of IQD16 in the regulation of MT arrays.

To analyze MT organization, we crossed the *GFP-MAP4* marker (Marc et al., 1998) into *Pro-35S:IQD16* lines. When compared with the parental wild-type (*GFP-MAP4*) line, light-grown plants overexpressing *IQD16* (*GFP-MAP4*) or *GFP-IQD16* displayed significantly elongated cells (Fig. 6, F and G), and the increased cell length of *IQD16* overexpression plants correlated with an altered orientation of cortical MTs (Fig. 6, H and I). Consistent with previous reports (Hamada et al., 2013), MTs in wild-type (*GFP-MAP4*) hypocotyl cells were distributed randomly. However, in *IQD16* (*GFP-MAP4*) and *GFP-IQD16*-overexpressing lines, MTs decorated by the GFP reporter were aligned preferentially in oblique arrays with angles between 40° and 70° (Fig. 6I). Thus, our experiments show that IQD16 localizes to cortical MTs in Arabidopsis and impacts MT orientation and cell shape.

Overexpression of *IQD11* and *IQD14* Differentially Modulates Plant Growth

We extended our analysis in Arabidopsis to validate the roles of additional IQD proteins at MTs (Fig. 7). We selected IQD11, which labeled MT patterns highly similar to IQD16 in the *N. benthamiana* system (Fig. 5), and IQD14, which decorates MT arrays in *N. benthamiana* (Fig. 1) and Arabidopsis (Fig. 2) in patterns different from IQD11 and IQD16 (Fig. 5). Consistent with the high similarity of MT patterns upon

dissimilarity and distances close to 1. The color bar shown encodes the similarity strength. C, Network analysis of MT patterns induced by the overexpression of GFP-IQD fusions. Nodes represent average MT patterns of individual IQD members. Node colors highlight the phylogenetic groups and subgroups, and the width of the connecting lines is proportional to the similarity between the nodes.
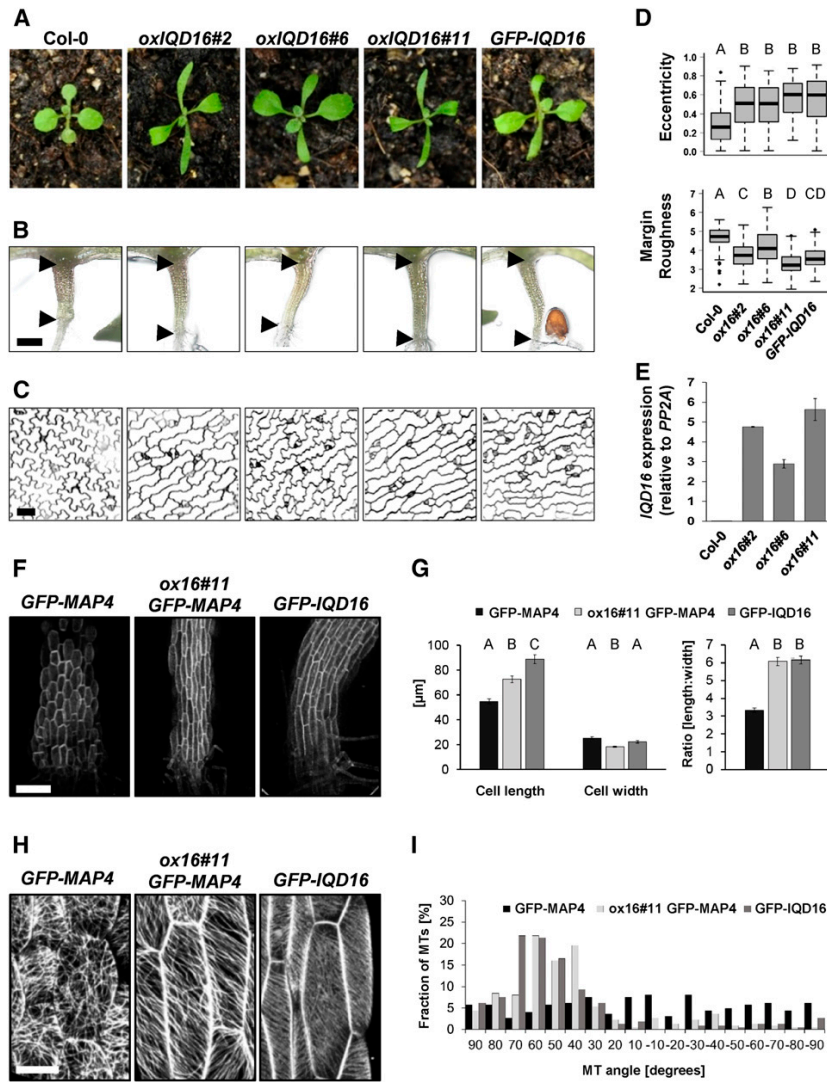
Figure 6. *IQD16* overexpression alters epidermal cell shape and cortical MT orientation. A to C, Phenotypes of Arabidopsis wild-type plants and *IQD16*-overexpressing plants grown under long-day conditions. Shown are representative images of shoots of the wild type (Col-0), three independent transgenic *Pro-35S:IQD16* lines (*ox16#2*, *ox16#6*, and *ox16#11*), and one transgenic *Pro-35S:GFP-IQD16* line (*GFP-IQD16*). A, Two-week-old plants grown on soil. B, Seven-day-old seedlings grown under sterile conditions. Arrowheads delimit hypocotyls, which are elongated in *IQD16*-overexpressing plants. Bar = 0.5 mm. C, Single optical sections of adaxial epidermal pavement cells in cotyledons of 5-d-old seedlings. Cell walls were visualized with PI. Bar = 50 μm. D, Quantification of the cellular eccentricity and margin roughness in cells shown in C. Results are medians from $n \geq$ 70 cells and $n \geq$ 3 seedlings, and boxes range from first to third quartiles. Different letters denote a significant statistical difference. $P < 0.005$ by one-way ANOVA. E, Quantitative reverse transcription (RT)-PCR analysis of *IQD16* transcript levels relative to *PP2A* in the three individual *Pro-35S:IQD16* transgenic lines shown in A in comparison with the wild type (Col-0). Results are averages of three replicates $\pm$ SE. F and G, Analysis of MT organization and cell shape in transgenic seedlings expressing *Pro-35S:GFP-MAP4* (*GFP-MAP4*), *Pro-35S:IQD16* and *Pro-35S:GFP-MAP4* (*oxIQD16#11 GFP-MAP4*), or *Pro-35S:GFP-IQD16* (*GFP-IQD16*). Seedlings of stably transformed Arabidopsis lines (5 d old) were grown on Arabidopsis salt medium under long-day conditions. F, Z-stack projections of hypocotyl epidermis cells. Bar = 100 μm. G, Epidermal hypocotyl cell size. Cell length and cell width were measured relative to the perpendicular axis, and the length-to-width ratio of individual cells was calculated (means $\pm$ SE, $n =$ 33 cells of three seedlings; Different letters denote a significant statistical difference. $P < 0.005$ by Student's t test). H, Z-stack projections of individual epidermal hypocotyl cells. Bar = 20 μm. I, Quantification of cortical MT orientation. Angles were measured relative to the perpendicular axis, and relative fractions were calculated ($n = 225$ MTs, with three independent experiments).

Figure 7. *IQD11* and *IQD14* over-expression lines display altered plant growth and MT orientation. A, Phenotypes of Arabidopsis wild-type and transgenic plants overexpressing *IQD11* and *IQD14*. Shown are representative images of shoots of 2-week-old plants grown under long-day conditions from the wild type (Col-0) and transgenic *Pro-35S:IQD11* (oxIQD11), *Pro-35S:YFP-IQD11* (YFP-IQD11), *Pro-35S:IQD14* (oxIQD14), and *Pro-35S:GFP-IQD14* (GFP-IQD14) lines. B, Subcellular localization of GFP-IQD16, YFP-IQD11, GFP-IQD14, GFP-MAP4, and GFP-ABD2 fusion proteins in seedlings treated with dimethyl sulfoxide (DMSO) or with the MT-depolymerizing drug oryzalin. Seedlings of transgenic Arabidopsis lines (5 d) were grown on Arabidopsis salt medium under long-day conditions. Micrographs are Z-stack projections of epidermal hypocotyl cells. Bar = 20 μm. C, Quantification of cortical MT orientation (Fig. 6). D, Epidermal pavement cell shape in the wild type and in transgenic *Pro-35S:YFP-IQD11* and *Pro-35S:GFP-IQD14* lines. Single optical sections show adaxial epidermal pavement cells in cotyledons from 5-d-old seedlings. Cell walls were visualized with PI. Bar = 50 μm. E, Quantification of the cellular eccentricity and margin roughness in cells. Results are medians from $n \geq 90$ cells and $n \geq 4$ seedlings, and boxes range from first to third quartiles. Different letters denote a significant statistical difference. $P < 0.001$ by one-way ANOVA.
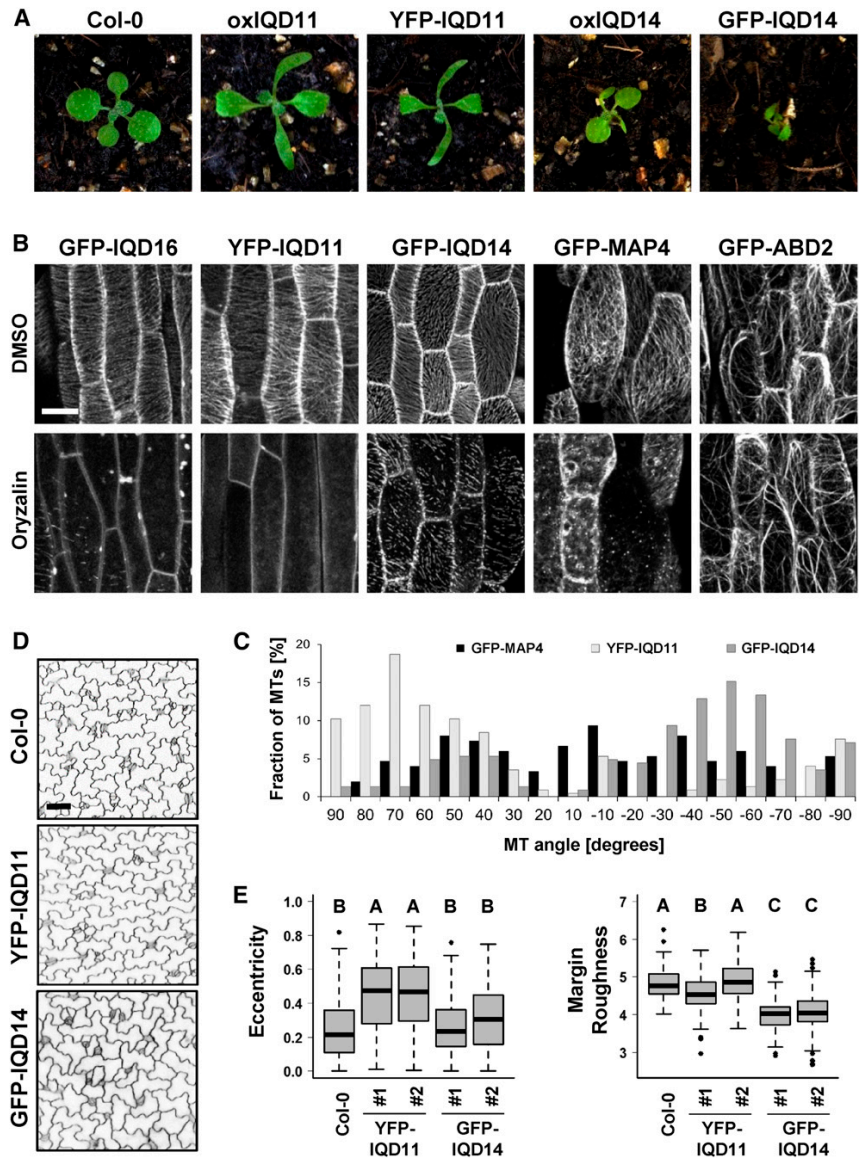
overexpression of *GFP-IQD16* and yellow fluorescent protein (YFP)-IQD11 in *N. benthamiana* (Fig. 5), the phenotypes of transgenic Arabidopsis plants overexpressing *IQD11* or *YFP-IQD11* resembled those of *IQD16* overexpression lines (Figs. 6 and 7A; Supplemental Fig. S4). YFP-IQD11 labeled MT arrays in patterns similar to GFP-IQD16 in epidermal hypocotyl cells of transgenic Arabidopsis seedlings (Fig. 7, B and C), and overexpression of YFP-IQD11 induced the elongation of epidermis pavement cells (Fig. 7, D and E). We observed distinct phenotypes upon overexpression (*Pro-35S*) of *IQD14* or *GFP-IQD14* (Fig. 7A; Supplemental Fig. S4). Transgenic plants displayed strongly induced organ twisting but no changes in leaf elongation (Fig. 7, A, D,

and E). Similar phenotypes were reported for *tortifolia* and *spiral* mutants (Furutani et al., 2000; Buschmann et al., 2004; Shoji et al., 2004). Helical growth defects commonly correlate with an increased transverse alignment of MTs (Ishida et al., 2007), which are visible upon the overexpression of YFP-IQD11, GFP-IQD14, and GFP-IQD16 (Figs. 6I and 7C). We confirmed the MT specificity of GFP-IQD16, YFP-IQD11, and GFP-IQD14 by the treatment of transgenic Arabidopsis seedlings with the MT-depolymerizing drug oryzalin (Fig. 7B, bottom row). We included transgenic GFP-MAP4 (Marc et al., 1998) and GFP-ABD2 (Sheahan et al., 2004; Wang et al., 2004) lines as controls for the MT and actin cytoskeleton, respectively. Oryzalin

treatment efficiently disrupted MT networks but had no effect on the actin cytoskeleton. Analysis of *ProIQD11:GFP-GUS* and *ProIQD14:GFP-GUS* expression revealed *IQD11* and *IQD14* promoter activity in the hypocotyl, cotyledons, leaf tissues, and petioles (Supplemental Fig. S4B). Thus, the phenotypes in combination with the observed MT localization suggest that elevated levels of IQD proteins may have diverse effects on MT organization and cell growth control in tissues with endogenous *IQD* expression. Moreover, our pattern-analysis tool proves to be a useful resource for the study of MT organization in complex shaped cells and provides, to our knowledge, the first experimental evidence for roles of IQD proteins in the regulation of MT networks and cell shape.

IQD Family Members Differentially Recruit CaM to Diverse Subcellular Sites

The differential subcellular localization of IQD family members (Fig. 1) prompted us to test whether IQD proteins recruit CaM to diverse subcellular sites. We selected eight IQD proteins representing the phylogenetic clades and different localization patterns of the Arabidopsis IQD family and performed bimolecular fluorescence complementation (BiFC) assays, which detect protein-protein interactions in planta with subcellular resolution (Gehl et al., 2009; Kudla and Bock, 2016). Coexpression in *N. benthamiana* of nYFP-IQD1 and cYFP-CaM2 reconstitutes YFP fluorescence at MTs and in the nucle(ol)us (Fig. 8), which is highly similar to the observed GFP-IQD1 localization (Fig. 1). Similarly, cYFP-CaM2 interacted with nYFP fusions of IQD8, IQD13, IQD16, and IQD32 at MTs. For nYFP-IQD20 and nYFP-IQD25, BiFC signals were detected in distinct PM subdomains, and nYFP-IQD33 interacted with cYFP-CaM2 in the nucleus (Fig. 8, insets). nYFP-TRM1 (Drevensek et al., 2012) served as a negative control, and the recovery of YFP fluorescence was not detectable with cYFP-CaM2, which confirmed the specificity of the BiFC assay. We independently verified the physical in vivo interactions of CaM2 and select IQDs by coexpression studies, which revealed GFP-IQD fusion protein-dependent relocalization of RFP-CaM2 (Supplemental Fig. S5). When coexpressed with GFP alone, RFP-CaM2 accumulated in the cytosol and nucleus, whereas coexpression with GFP-IQD1 induced the recruitment of RFP-CaM2 to MTs and the nucleolus, as reported previously (Bürstenbinder et al., 2013). In addition, coexpression with GFP-IQD8, GFP-IQD13, GFP-IQD16, or GFP-IQD32 caused the recruitment of RFP-CaM2 to MTs, and relocalization of RFP-CaM2 to the PM was evident upon coexpression with GFP-IQD25. Thus, IQD proteins recruit CaM2 and possibly other CaM/CMLs to MTs, PM subdomains, and nuclear compartments. Interaction of IQD proteins with CaM2 did not abolish IQD localization to MTs or the PM, suggesting that subcellular targeting of IQD proteins is independent of CaM recruitment.
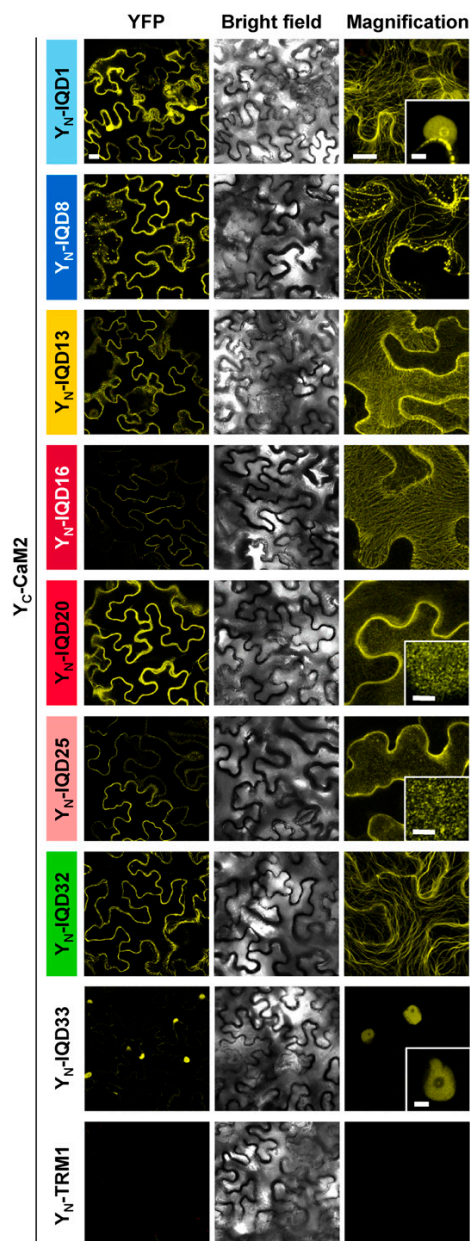
Figure 8. IQD proteins interact with CaM2 at MTs, at the PM, and in the nucleus. Single optical sections of BiFC signals (left column) and corresponding bright-field images (center column) as well as closeup Z-stack images of YFP fluorescence (right column) are shown for *N. benthamiana* epidermis cells after coexpression of IQD proteins N-terminally tagged with the N-terminal half of Venus (Y_N) and CaM2 N-terminally tagged with the C-terminal half of Venus (Y_C). Insets in the right column show nuclei or the membrane surface. Y_N-TRM1 was used as a negative control. Bars = 50 μm (left column), 20 μm (right column), and 5 μm (insets).

DISCUSSION

Plants evolved a remarkable repertoire of proteins for decoding information encrypted by cellular Ca^{2+} signatures, which are transiently generated in specific responses to numerous systemic or environmental stimuli. CaM/CML polypeptides are principal Ca^{2+} sensor relays that control the biochemical activities of diverse regulatory targets via complex Ca^{2+}-dependent and Ca^{2+}-independent interactions (Dodd et al., 2010; Kudla et al., 2010; Hashimoto and Kudla, 2011). Knowledge of the CaM/CML targets and of their biochemical functions and underlying biological processes will be imperative for deciphering CaM-mediated Ca^{2+} signal transduction in plants (Yang and Poovaiah, 2003). Plant-specific *IQD* gene families, first annotated in Arabidopsis and rice (Abel et al., 2005) and more recently in five additional species (Filiz et al., 2013; Huang et al., 2013; Feng et al., 2014; Ma et al., 2014; Wu et al., 2016), possibly encode the largest class of CaMBPs in plants. Although the roles of a few IQD members are beginning to emerge, the precise molecular functions and mechanisms of action are still elusive for any IQD protein (Abel et al., 2013).

We previously identified Arabidopsis IQD1 in a genetic screen for glucosinolate-related mutants (Levy et al., 2005) and noticed its colocalization with cortical MT arrays (Bürstenbinder et al., 2013). In this study, we report the subcellular distribution patterns of all 33 Arabidopsis IQD family members in transiently transfected *N. benthamiana* cells, which revealed that most GFP-IQD fusion proteins differentially localize to the PM, to the cortical MT network, and to distinctive sites of the cell nucleus (Fig. 1). We confirmed the localization patterns of select IQD proteins (Fig. 2), which were independent of GFP tag configuration (N or C terminal), promoter strength (*Pro-35S* and *ProIQD*), and expression system (*N. benthamiana* and Arabidopsis). The localization of GFP-fused IQD proteins to MT arrays, the PM, and the cell nucleus is in agreement with published proteome data. Analysis of MAP-enriched fractions prepared from Arabidopsis cell cultures identified a large set of known and novel MAPs. Interestingly, four IQD proteins of phylogenetic group IV (IQD28, IQD29, IQD31, and IQD32) were highly enriched in these fractions, and MT colocalization was confirmed for GFP fusions of IQD28 and IQD32 (Hamada et al., 2013). A phosphoproteome analysis of PM-associated proteins after elicitor treatment identified several IQDs in PM-enriched fractions, including IQD9, IQD11, and IQD14 (Benschop et al., 2007). Interrogation of the nuclear proteome detected IQD28 (Bigeard et al., 2014; Palm et al., 2016) and IQD32 (Bigeard et al., 2014), which independently confirms the targeting of at least two IQD proteins to multiple subcellular sites (Hamada et al., 2013).

While our work provides evidence for the nuclear, MT, and PM localization of most IQD family members, we are left to speculate about the modes of subcellular IQD targeting. In several MAPs, the binding to MTs is mediated by positively charged motifs, most likely via electrostatic interaction with acidic tail motifs of tubulin subunits (Smith et al., 2001; Mishima et al., 2007; Drevensek et al., 2012; Roll-Mecak, 2015). Similarly, membrane-binding motifs used to anchor proteins to the acidic phospholipid surface of membranes often consist of basic amino acids (Goldenberg and Steinberg, 2010; Scott et al., 2013), and nuclear localization signals require core basic residues for their activity (Kosugi et al., 2009). The high basic pI of IQD proteins (~10.3), which is a hallmark of the family (Abel et al., 2005), suggests that electrostatic interaction contributes to the MT and PM attachment of IQD proteins, and charge-plot analysis reveals the occurrence of basic patches in all 33 Arabidopsis IQD members (Supplemental Fig. S6). Membrane tethering of IQD proteins is possibly stabilized by *S*-acylation (or palmitoylation), which attaches fatty acids to internal Cys residues. In silico analyses using the GPS-Lipid and CSS-Palm tools (Ren et al., 2008; Xie et al., 2016) predicted the presence of *S*-acylation sites in several IQD family members (Supplemental Table S1). Experimental support for the lipidation of IQD proteins is provided by a large-scale proteomics data set, which reported *S*-acylation of IQD32 (Hemsley et al., 2013).

The four PM-localized GFP-IQD fusions that we examined in more detail (IQD12, IQD22, IQD24, and IQD25) labeled distinct and immobile PM subdomains, which suggests specific cellular functions for each IQD (Fig. 3). Studies in *Saccharomyces cerevisiae* (yeast) and Arabidopsis revealed that different populations of subdomains coexist in membranes (Spira et al., 2012; Jarsch et al., 2014). Notably, *S*-acylation appears to play a predominant role for protein anchoring at membrane subdomains in plants (Hemsley and Grierson, 2008), as has been demonstrated for Rho-like GTPases (Craddock et al., 2013) and for plant-specific scaffold proteins of the Rem family (Konrad et al., 2014; Konrad and Ott, 2015). Individual subdomains are thought to reflect specialized functions of PM-associated protein complexes, such as in signal transduction (Malinsky et al., 2013), because protein components of functional modules tend to colocalize in the same type of membrane subdomains. Such copartitioning was reported recently for the brassinosteroid receptor BRASSINOSTEROID INSENSITIVE 1 (BRI1) and the subdomain marker ARABIDOPSIS FLOTILLIN-LIKE PROTEIN 1 (AtFlot1) (Wang et al., 2015) as well as for the anion channel SLAC1-HOMOLOGUE 3 (SLAH3) and the Ca^{2+}-dependent protein kinase CPK21 (Demir et al., 2013). IQD-dependent recruitment of CaM to membrane subdomains (Fig. 7) is particularly intriguing, as it may locally restrict CaM-mediated Ca^{2+} signaling to specific sites within the PM and to spatially separate functionally distinct CaM signaling modules. Thus, the identification of IQD-interacting proteins within individual PM subdomains will be key for elucidating modules of CaM signaling pathways and assigning functions to individual IQD family members.

Lateral stability is a hallmark of membrane subdomains, and their stabilization within the PM requires scaffolding functions of the cytoskeleton (Kusumi et al., 2012). Colocalization with RFP-TUA5 revealed that the PM subdomains labeled by GFP-tagged IQD proteins partially align along MT arrays (Fig. 3). Similar patterns have been reported for Rem proteins (Raffaele et al., 2007; Jarsch et al., 2014), and targeting of AtRem6.6 to subdomains in the PM proved to be sensitive to MT-depolymerizing drugs (Jarsch et al., 2014), as was the case for IQD12 and IQD22 (Fig. 3). A function of the PM-MT nexus for subdomain formation is further supported by studies in yeast also in S. cerevisiae (Kusumi et al., 2012). In addition, the plant PM-MT interface plays important roles in cell wall biogenesis and cell shape regulation, mainly by guiding the deposition of nascent cellulose fibers. Cellulose biosynthesis requires the delivery of CSCs to the PM and the coordination of CSC trajectories along MT arrays (Liu et al., 2015). The dual localization of several IQD proteins to MTs and the PM, and the cell shape defects caused by *IQD* overexpression (Figs. 5–7), are consistent with a function of IQDs at the PM-MT connection. A striking transverse alignment of cortical MTs, as observed for *IQD*-overexpressing cells (Figs. 6 and 7), often correlates with a spiral pattern of cellulose deposition, which favors longitudinal cell expansion (Ivakov and Persson, 2013). Despite decades of research, it is not well understood how CSCs interact with MTs and how MTs are tethered to the PM. Our study identified IQD family members as putative candidates mediating such interactions. Elucidating how IQD proteins contribute to functional PM-MT connections (e.g. by affecting cellulose deposition, subdomain formation, or MT attachment to the PM) will be a challenging task for the future.

The differential effects of overexpressed IQD family members on MT arrays and the MT-related phenotypes of *IQD*-overexpressing Arabidopsis plants point to roles of IQDs in the organization of MT networks. Interestingly, we noticed that most GFP-IQD proteins uniformly decorated MT arrays, similar to members of the MAP65 family (Lucas et al., 2011), which function mainly in MT bundling (Chen et al., 2016). In contrast, most other MAPs either accumulate at the plus end of MTs, such as CLIP-ASSOCIATED PROTEIN (CLASP) (Kirik et al., 2007), END BINDING PROTEIN EB1 (Bisgrove et al., 2008), or SPIRAL1 (SPR1) (Sedbrook et al., 2004), where they regulate the dynamics of MT growth and shrinkage, or associate in punctate to patchy patterns preferably at newly formed MT crossover sites, as described for KATANIN 1(KTN1) (Lindeboom et al., 2013) and SPR2 (Shoji et al., 2004). Evenly distributed MT association suggests that IQD proteins may influence MT bundling. However, further work is required to uncover the mechanisms by which IQDs modulate MT organization. In vivo studies on IQD dynamics and cellular functions have been hampered by low levels of IQD protein abundance and a high degree of genetic redundancy within the IQD family.

Therefore, transient expression assays combined with our computational tool to quantitatively analyze MT patterns may prove beneficial for elucidating IQD functions. By applying this tool in a proof-of-concept study, we identified IQD16 and IQD11 as the most divergent IQD members in terms of MT patterning (Fig. 5) and confirmed their function in regulating MT orientation and cell shape in transgenic Arabidopsis plants (Figs. 6 and 7). Differential MT-modulating functions of individual IQDs are further corroborated by MT-related phenotypes in *IQD14*-overexpressing Arabidopsis seedlings (Fig. 7). Therefore, we conclude that MT arrays are sensitive to reorganization by (transient) overexpression of *IQDs*. Comparison of MT patterns caused by the overexpression of IQDs with the MT patterns formed upon the overexpression of MAPs with known functions (e.g. in bundling or cross-linking of MTs) could assist to predict molecular mechanisms of IQD functions. Thus, transient expression assays combined with our computational tool of MT array pattern analysis offers a useful resource with which to study potential regulators of MT organization in complex shaped cells. Functions in growth regulation, however, are not limited to MT-localized IQD family members, as revealed by altered cell shapes in plants overexpressing the PM-localized IQD25. Notably, we observed distinct and, in some cases, even opposing effects on growth and cell shape formation for the analyzed IQD members (i.e. increased/augmented/more pronounced elongation of leaves and cells in *IQD11*- and *IQD16*-overexpressing plants and rounding/more compact leaves and cells in *IQD25*-overexpressing plants; Figs. 4, 6, and 7). These results suggest specialized functions of individual IQDs in regulating the direction of cell elongation and possibly in cell polarity establishment.

We demonstrated in BiFC assays that select IQD members recruit CaM2 to diverse subcellular sites, which was confirmed independently by the relocalization of cytoplasmic CaM2 upon coexpression with each IQD protein (Fig. 8; Supplemental Fig. S5). Although we tested only one CaM isoform versus eight IQD proteins, it is likely that all IQD family members interact with a select set of the four CaM and 50 CML Ca^{2+} sensor polypeptides in Arabidopsis. We previously verified the predicted function of the conserved IQ67 domain for CaM binding and demonstrated in yeast two-hybrid assays the interaction of IQD1 with three CaM isoforms and of IQD20 with CaM2 and CML13 (Bürstenbinder et al., 2013). We are left to speculate on the Ca^{2+} state of recruited CaM because the IQ67 domain contains multiple Ca^{2+}-dependent and Ca^{2+}-independent CaM retention motifs and IQD20 binds in vitro to both apo-CaM and Ca^{2+}-CaM (Abel et al., 2005). The biochemical properties and functions of IQD proteins may be regulated directly upon binding to CaM and determined by its Ca^{2+} occupancy. Stimulus-specific activation of Ca^{2+} signaling pathways may specify and fine-tune the IQD affinity to MTs, the PM, and other interactors, such as IQD1-binding

KLCR1 (Bürstenbinder et al., 2013). Alternatively, as reported for fungal and mammalian IQGAP scaffold proteins, which orchestrate cellular signaling from the PM to the nucleus (Shannon, 2012; Smith et al., 2015), IQDs may act as reservoirs for CaM/CMLs. In that case, IQD proteins sequester apo-CaM/CMLs at specific subcellular sites, which upon Ca^{2+} sensing and loading are transferred to targets regulated by holo-CaM/ CMLs, or vice versa. Thus, the IQD family may provide an assortment of versatile platform proteins that facilitate and specify CaM/CML dynamics during Ca^{2+} signaling at the cell periphery, on the cytoskeleton, and in the cell nucleus. Interactions of IQDs and CaM/ CMLs in the nucleus may integrate nuclear calcium signaling (Charpentier and Oldroyd, 2013) and regulate a proposed function of select IQD proteins during mRNA maturation, export, and cytoplasmic transport (Abel et al., 2013; Bürstenbinder et al., 2013). These processes share nucleocytoplasmic protein factors and allow for efficient fine-tuning of protein targeting and activity (Marchand et al., 2012; Shahbabian and Chartrand, 2012).

Several properties of IQD family members point to their roles as scaffold or adaptor proteins. The central conserved IQ67 domain of most IQD proteins, which likely adopts an α-helical fold, is flanked on either side by extensive regions of predicted intrinsic disorder (Supplemental Fig. S6). Substantial intrinsic disorder is a hallmark of scaffolding proteins. The conformational flexibility and biochemical properties of natively unstructured regions, which often contain multiple, short linear peptide motifs for molecular interactions and/or posttranslational modifications (Tompa et al., 2014), specify and fine-tune the assembly of macromolecular complexes via induced and cooperative folding (Babu et al., 2011; Wright and Dyson, 2015). Because intrinsically disordered proteins are prone to engage in promiscuous molecular interactions via peptide motifs of low complexity, their steady-state concentrations are tightly regulated at multiple levels, and elevated gene expression often causes detrimental effects or disease (Vavouri et al., 2009; Babu et al., 2011). While almost nothing is known about the control of *IQD* gene expression and IQD protein stability, with the exception of *IQD22*, which is induced rapidly by GA_3 via DELLA-dependent regulation (Zentella et al., 2007), it is of note that IQD-related phenotypes have been reported only for *IQD*-overexpressing plants, which display altered secondary metabolism (Levy et al., 2005), cell shape (Figs. 6 and 7), and organ morphology (Xiao et al., 2008).

In conclusion, our comprehensive study of subcellular localization patterns of all 33 Arabidopsis IQD proteins, in combination with the computational analysis of MT arrays, in planta CaM interaction assays of select IQD members, and characterization of *IQD11*, *IQD14*, *IQD16*, and *IQD25* overexpression lines, provide strong evidence that plant-specific IQD families define a new class of largely MT-based CaMBPs with additional but distinctive roles in PM subdomains and nuclear compartments. We propose that IQD families provide a multifaceted toolbox of scaffold-like proteins for integrating CaM-dependent Ca^{2+} signaling and possibly other signal transduction pathways at multiple subcellular sites to regulate cell function, shape, and growth during plant development.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

Nicotiana benthamiana plants were grown in a greenhouse at 22°C to 24°C under long-day conditions (16 h of light/8 h of dark). Arabidopsis (*Arabidopsis thaliana*) ecotype Col-0 seeds were originally obtained from the Arabidopsis Biological Resource Center. Seeds of *iqd16-2* (SALK_053223), *iqd25-1* (SALK_058876), and *iqd25-2* (SALK_148613) T-DNA lines were obtained from the Nottingham Arabidopsis Stock Centre (Scholl et al., 2000). To visualize MT arrays, transgenic Arabidopsis seedlings expressing a GFP fusion of the MT-binding domain of mouse MAP4 under the control of the *CaMV 35S* promoter were analyzed (Marc et al., 1998). For growth under semicontrolled conditions, plants were stratified at 4°C for 2 d and cultivated under long-day conditions as described above. For sterile cultivation, Arabidopsis seeds were surface sterilized with chlorine gas and, after 2 d of stratification, grown vertically on square plates containing Arabidopsis salt medium and 1% (w/v) agar with cycles of 16 h of light and 8 h of dark (Lincoln et al., 1990).

Plasmid Construction

DNA sequence information of *IQDs*, *PDLP1*, *TRM1*, *Rem6.6*, *Rem6.7*, and *TUA5* was obtained from The Arabidopsis Information Resource. Open reading frames were amplified from the Arabidopsis Biological Resource Center or RIKEN cDNA clones or from cDNA of Arabidopsis Col-0 seedlings using gene-specific primers (Supplemental Tables S2 and Table S4). Full-length genomic loci and upstream promoter sequences were amplified from genomic DNA of Arabidopsis Col-0. Amplicons were ligated into Gateway-compatible entry and donor vectors (Supplemental Table S2), and insert fidelity was verified by DNA sequencing. The generation of CaM2 ENTR plasmid was described by Fischer et al. (2013). Inserted sequences were mobilized into pB7WGF2, pB7WGY2 (Karimi et al., 2002), or pGWB455 (Nakagawa et al., 2007) for the expression of the N-terminal GFP, YFP, or RFP fusion under the control of the *CaMV 35S* promoter, respectively, or into pB7FWG or pB7FWG2 (Karimi et al., 2002) for the expression of C-terminal GFP fusions under the control of the native promoter sequences or 35S promoter, respectively. For BiFC assays, coding sequences were mobilized into pDEST-VYNE(R)^{GW} and pDEST-VYCE(R)^{GW} to generate N-terminal fusions with the N- and C-terminal halves of VENUS, respectively. To generate *ProIQD*:*GFP-GUS* reporter constructs, the promoter region was mobilized into pBGWFS7 (Karimi et al., 2002).

Plant Transformation

For transient expression assays in *N. benthamiana*, leaves were coinfiltrated with *Agrobacterium tumefaciens* GV3101 pMP90RK harboring plasmids and the silencing suppressor p19 in a 1:1 ratio. Bacterial cultures were adjusted to an optical density at 600 nm of 0.5 using infiltration buffer, and *N. benthamiana* leaves were pressure infiltrated through the abaxial epidermis. *A. tumefaciens*-mediated transient transformation of Arabidopsis was performed in leaves of 4-week-old plants grown under short-day conditions according to the protocol described by Mangano et al. (2014). For stable transformation of Arabidopsis, the floral dip technique was used (Clough and Bent, 1998), and eight to 24 independent transgenic lines were identified in the T1 generation by Basta selection for each construct analyzed. To verify single-copy transgene insertion, T2 transgenic lines were analyzed for a 3:1 segregation ratio (Basta). GFP fluorescence and GUS expression were analyzed in homozygous T3 plants of two to four lines showing representative GUS expression patterns or growth phenotypes.

Quantitative Real-Time PCR and RT-PCR

Rosette leaves from three individual 4-week-old plants were pooled, and total RNA was extracted from approximately 100 mg of tissue using the RNeasy Plant

Mini Kit (Qiagen). After DNase treatment, first-strand cDNA was synthesized from 4 μg of RNA using oligo(dT) primers with the Revert Aid First Strand cDNA synthesis kit (Thermo Fisher). Quantitative real-time PCR was performed on the 7500 Fast Real-Time PCR system using Fast SYBR Green master mix (Applied Biosystems) with 2 μL of 1:10 diluted cDNA. The relative expression of *IQD16* was calculated relative to *PP2A*. For RT-PCR analyses, RNA was isolated from 5-d-old seedlings grown under sterile conditions, and cDNA was synthesized from 5 μg of total RNA as described above. *ACTIN2* was amplified from 2 μL of 1:10 diluted cDNA, and 2 μL of undiluted cDNA was used as a template for the amplification of *IQD16* and *IQD25*. Primers used for amplification and their sequences are given in Supplemental Tables S3 and S4.

Histochemical Analysis

For GUS staining, sample tissues were fixed in 80% (v/v) ice-cold acetone for 30 min and incubated for 2 to 8 h in GUS staining solution [50 mM sodium phosphate, pH 7.2, 0.5 mM $K_3Fe(CN)_6$, 0.5 mM $K_4Fe(CN)_6$, 2 mM 5-bromo-4-chloro-3-indolyl-β-glucuronic acid, 10 mM EDTA, and 0.1% (v/v) Triton X-100] at 37°C. Imaging of GUS staining patterns was performed using a Zeiss Axioplan 2 microscope.

Confocal Laser Scanning Microscopy

Unless stated otherwise, imaging was performed with a Zeiss LSM 700 inverted microscope using a 40× water-immersion objective. The excitation wavelength for GFP and YFP was 488 nm; emission was detected between 493 and 555 nm. RFP was excited with a 555-nm laser, and emission was detected using a long-pass filter (582–700). For colocalization assays, images were obtained in the sequential mode. Root cells of transgenic Arabidopsis seedlings were plasmolyzed by treatment with 150 mM NaCl as described by Müller et al. (2015). For PI or FM4-64 staining, cells were incubated for 1 to 5 min in 10 μM PI or 50 μM Synapto-Red C2, respectively. Seedlings were imaged after two wash steps in sterile water. For the analysis of epidermis pavement cell shapes, single optical sections of the adaxial side of FM4-64-stained cotyledons were acquired with a 20× objective. Surface images of epidermal cells were acquired with a 63× water-immersion objective. BiFC assays were carried out with a Zeiss LSM 710 inverted microscope according to the protocol described by Gehl et al. (2009). Imaging was performed with identical laser settings for all samples, and emission spectra of reconstituted fluorescence were recorded by lambda scan.

Oryzalin treatments were performed as described by Bürstenbinder et al. (2013). Briefly, a 2 mM stock solution of oryzalin was prepared in dimethyl sulfoxide. *N. benthamiana* leaf discs were incubated in an aqueous solution of 50 μM oryzalin for up to 90 min. For treatment of Arabidopsis seedlings, plants were incubated in a final concentration of 20 μM oryzalin for 1 to 3 h.

Image Analysis

For semiautomatic segmentation, cell outlines were labeled in the red channel by PI staining or by expressing PM-localized RFP-PDLP1 (Amari et al., 2010). In maximum projections, cell outlines appear as thin vessel-like structures with locally varying contrast. The contrast was optimized by smoothing of the image with a Gaussian filter ($\sigma = 1$) and applying a Mexican hat filter in 18 steps of 10° to emphasize vessel-like structures. The resulting images containing the maximal filter response for each pixel were binarized by local Niblack thresholding as implemented in MiToBo (Möller et al., 2016). Subsequently, postprocessing steps were applied to the binary images (i.e. component labeling, exclusion of too-small components, and skeletonization), which allowed the extraction of large parts of the cell boundaries. Remaining gaps were corrected by manual postprocessing, and cells exceeding the image border were excluded manually. Subsequently, texture analysis techniques were applied to automatically characterize MT organization as stated in the result section of the main text (see Fig. 5). The feature vectors of all cells were jointly clustered applying k means. For each cell, the ratio of windows belonging to the different k clusters was calculated, which results in a k dimensional cluster distribution vector. Each cluster refers to a specific structural pattern appearing in the cells; hence, each vector can be interpreted as a probability density distribution of these patterns appearing in the corresponding cell. From the set of vectors, pairwise distances or distances between average vectors for each cell line were extracted and visualized in heat maps or adjacency networks (Fig. 3, A and B). The plots shown in this study were generated by applying nonoverlapping windows with a size of 16 × 16 pixels from which rotation invariant local binary patterns were extracted. For clustering, k was set to 10 and Euclidean distances were used. These parameters were empirically determined in prestudies that revealed that changing of parameters (i.e. window size and sliding windows) did not affect the detected global tendencies. In total, the analysis was based on 130 images containing 184 cells. For each IQD family member, five to 14 individual cells were analyzed. Networks were visualized with Cytoscape (Shannon et al., 2003).

To assess the lateral mobility of membrane subdomains, stacks of 10 to 12 images were acquired over 20 min in intervals of 2 min. Lateral shift of stacks was corrected via the Fiji plugin Stackreg (Schindelin et al., 2012). Kymographs were generated with the Fiji command Reslice using a line width of 12.

Quantification of subdomains was performed with Fiji. Images were background subtracted (rolling ball, 70 pixels). The radius was set according to the heuristically determined maximal size of objects of interest. Dot structures were detected with the wavelet-based particle detector (Greß et al., 2010) included in the MiToBo toolbox (jmin = 3, jmax = 5, scale interval of s = 2). The correlation threshold was fixed to t = 0.6, and regions with sizes between 100 and 1,750 pixels were analyzed to calculate average domain sizes.

Epidermis pavement cell shape was analyzed in cotyledons of 5-d-old seedlings grown sterile under long-day conditions. Cell walls were visualized by PI staining. Cell outlines were labeled in single optical sections of epidermis pavement cells with Fiji, and margin roughness and eccentricity were calculated as measures of (non)smoothness of the contour and cellular elongation, respectively. As a measure of cell shape (ir)regularity, first an average angle of tangent orientation changes is computed along the cell contour based on tangent orientations of pairs of adjacent contour points. Next, the deviation of the calculated angle from the angle expected for a perfect circle with an equal number of sampling points is calculated, referred to as margin roughness. A higher margin roughness value corresponds to a higher degree of nonsmoothness and reflects the extent of cellular lobing. Eccentricity is defined as the ratio between the extension of a region along its main axis and along the corresponding perpendicular axis. The extension along each axis is extracted from second-order central moments of the cell region. An almost circular region has an eccentricity value of 0, which increases with increasing elongation.

Immunocytochemistry

For ultrastructural localization of GFP-tagged proteins, root tips were transferred into aluminum planchettes and high-pressure frozen with an HPM 10 apparatus (Bal-Tec). Subsequently, the material was cryosubstituted in 0.25% (v/v) glutaraldehyde (Sigma) and 0.1% (w/v) uranyl acetate (Chemapol) in acetone for 2 d at −80°C using cryosubstitution equipment (FSU; Bal-Tec). This was followed by embedding in HM20 (Polysciences Europe) at −20°C.

For immunolabeling of ultrathin sections, we used a polyclonal anti-GFP antibody (600-101-215; Rockland) detected by a rabbit anti-goat secondary antibody conjugated with 10-nm gold (G 5527; Sigma). Sections were poststained with uranyl acetate and lead citrate in an EM-Stain apparatus (Leica) and subsequently observed with an EM 900 transmission electron microscope (Carl Zeiss Microscopy). Micrographs were taken with the Variospeed SSCCD (TRS).

Statistical Analysis

Statistical analysis was performed in R (version 3.2.1) by one-way ANOVA with Tukey's honestly significant difference test.

Accession Numbers

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers At3g09710 (IQD1), At5g03040 (IQD2), At3g52290 (IQD3), At2g26410 (IQD4), At3g22190 (IQD5), At2g26180 (IQD6), At1g17480 (IQD7), At1g72670 (IQD8), At2g33990 (IQD9), At3g15050 (IQD10), At5g13460 (IQD11), At5g03960 (IQD12), At3g59690 (IQD13), At2g43680 (IQD14), At3g49380 (IQD15), At4g10640 (IQD16), At4g00820 (IQD17), At1g01110 (IQD18), At4g14750 (IQD19), At3g51380 (IQD20), At3g49260 (IQD21), At4g23060 (IQD22), At5g62070 (IQD23), At5g07240 (IQD24), At4g29150 (IQD25), At3g16490 (IQD26), At1g51960 (IQD27), At1g14380 (IQD28), At2g02790 (IQD29), At1g18840 (IQD30), At1g74690 (IQD31), At1g19870 (IQD32), At5g35670 (IQD33), At5g19780 (TUA5), At5g43980 (PDLP1), At3g02170 (TRM1), At1g13920 (Rem6.6), At5g61280 (Rem6.7), At3g18780 (ACTIN2), At1g13320 (PP2Ac), At2g41110 (CaM2).

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Expression analysis and subcellular localization of Arabidopsis *IQD* genes.

Supplemental Figure S2. Phenotypes of *iqd25* mutant lines and expression analysis of *ProIQD25::GFP-GUS*.

Supplemental Figure S3. Phenotypes of *iqd16* lines and expression analysis of *ProIQD16::GFP-GUS*.

Supplemental Figure S4. Expression analysis of *IQD11* and *IQD14* overexpression lines and of *ProIQD11:GFP-GUS* and *ProIQD14:GFP-GUS* lines.

Supplemental Figure S5. Coexpression assays of GFP-IQD fusions with RFP-CaM2.

Supplemental Figure S6. Predicted properties of Arabidopsis IQD proteins.

Supplemental Table S1. Prediction of nuclear localization signals, lipidation sites, and signal peptides in Arabidopsis IQD proteins.

Supplemental Table S2. Gene identifiers and plasmids.

Supplemental Table S3. Primer combinations used for quantitative PCR, RT-PCR, and genotyping.

Supplemental Table S4. Primer sequences.

LITERATURE CITED

Abel S, Bürstenbinder K, Müller J (2013) The emerging function of IQD proteins as scaffolds in cellular signaling and trafficking. Plant Signal Behav **8:** e24369

Abel S, Savchenko T, Levy M (2005) Genome-wide comparative analysis of the *IQD* gene families in *Arabidopsis thaliana* and *Oryza sativa*. BMC Evol Biol **5:** 72

Amari K, Boutant E, Hofmann C, Schmitt-Keichinger C, Fernandez-Calvino L, Didier P, Lerich A, Mutterer J, Thomas CL, Heinlein M, et al (2010) A family of plasmodesmal proteins with receptor-like properties for plant viral movement proteins. PLoS Pathog **6:** e1001119

Babu MM, van der Lee R, de Groot NS, Gsponer J (2011) Intrinsically disordered proteins: regulation and disease. Curr Opin Struct Biol **21:** 432–440

Benschop JJ, Mohammed S, O'Flaherty M, Heck AJR, Slijper M, Menke FLH (2007) Quantitative phosphoproteomics of early elicitor signaling in Arabidopsis. Mol Cell Proteomics **6:** 1198–1214

Bigeard J, Rayapuram N, Bonhomme L, Hirt H, Pflieger D (2014) Proteomic and phosphoproteomic analyses of chromatin-associated proteins from Arabidopsis thaliana. Proteomics **14:** 2141–2155

Bisgrove SR, Lee YR, Liu B, Peters NT, Kropf DL (2008) The microtubule plus-end binding protein EB1 functions in root responses to touch and gravity signals in *Arabidopsis*. Plant Cell **20:** 396–410

Bringmann M, Landrein B, Schudoma C, Hamant O, Hauser MT, Persson S (2012) Cracking the elusive alignment hypothesis: the microtubule-cellulose synthase nexus unraveled. Trends Plant Sci **17:** 666–674

Bürstenbinder K, Savchenko T, Müller J, Adamson AW, Stamm G, Kwong R, Zipp BJ, Dinesh DC, Abel S (2013) Arabidopsis calmodulin-binding protein IQ67-domain 1 localizes to microtubules and interacts with kinesin light chain-related protein-1. J Biol Chem **288:** 1871–1882

Buschmann H, Dols J, Kopischke S, Peña EJ, Andrade-Navarro MA, Heinlein M, Szymanski DB, Zachgo S, Doonan JH, Lloyd CW (2015) Arabidopsis KCBP interacts with AIR9 but stays in the cortical division zone throughout mitosis via its MyTH4-FERM domain. J Cell Sci **128:** 2033–2046

Buschmann H, Fabri CO, Hauptmann M, Hutzler P, Laux T, Lloyd CW, Schäffner AR (2004) Helical growth of the Arabidopsis mutant *tortifolia1* reveals a plant-specific microtubule-associated protein. Curr Biol **14:** 1515–1521

Cárdenas L (2009) New findings in the mechanisms regulating polar growth in root hair cells. Plant Signal Behav **4:** 4–8

Charpentier M, Oldroyd GE (2013) Nuclear calcium signaling in plants. Plant Physiol **163:** 496–503

Chen X, Wu S, Liu Z, Friml J (2016) Environmental and endogenous control of cortical microtubule orientation. Trends Cell Biol **26:** 409–419

Clough SJ, Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J **16:** 735–743

Craddock C, Lavagi I, Yang ZB (2012) New insights into Rho signaling from plant ROP/Rac GTPases. Trends Cell Biol **22:** 492–501

Deavours BE, Reddy AS, Walker RA (1998) Ca^{2+}/calmodulin regulation of the Arabidopsis kinesin-like calmodulin-binding protein. Cell Motil Cytoskeleton **40:** 408–416

Deeks MJ, Calcutt JR, Ingle EKS, Hawkins TJ, Chapman S, Richardson AC, Mentlak DA, Dixon MR, Cartwright F, Smertenko AP, et al (2012) A superfamily of actin-binding proteins at the actin-membrane nexus of higher plants. Curr Biol **22:** 1595–1600

Demir F, Horntrich C, Blachutzik JO, Scherzer S, Reinders Y, Kierszniowska S, Schulze WX, Harms GS, Hedrich R, Geiger D, et al (2013) Arabidopsis nanodomain-delimited ABA signaling pathway regulates the anion channel SLAH3. Proc Natl Acad Sci USA **110:** 8296–8301

Dodd AN, Kudla J, Sanders D (2010) The language of calcium signaling. Annu Rev Plant Biol **61:** 593–620

Drevensek S, Goussot M, Duroc Y, Christodoulidou A, Steyaert S, Schaefer E, Duvernois E, Grandjean O, Vantard M, Bouchez D, et al (2012) The *Arabidopsis* TRM1-TON1 interaction reveals a recruitment network common to plant cortical microtubule arrays and eukaryotic centrosomes. Plant Cell **24:** 178–191

Endler A, Persson S (2011) Cellulose synthases and synthesis in Arabidopsis. Mol Plant **4:** 199–211

Feng L, Chen Z, Ma H, Chen X, Li Y, Wang Y, Xiang Y (2014) The *IQD* gene family in soybean: structure, phylogeny, evolution and expression. PLoS ONE **9:** e110896

Filiz E, Tombuloglu H, Ozyigit II (2013) Genome-wide analysis of IQ67 domain (IQD) gene families in *Brachypodium distachyon*. Plant Omics **6:** 425–432

Fischer C, Kugler A, Hoth S, Dietrich P (2013) An IQ domain mediates the interaction with calmodulin in a plant cyclic nucleotide-gated channel. Plant Cell Physiol **54:** 573–584

Furutani I, Watanabe Y, Prieto R, Masukawa M, Suzuki K, Naoi K, Thitamadee S, Shikanai T, Hashimoto T (2000) The SPIRAL genes are required for directional control of cell elongation in Arabidopsis thaliana. Development **127:** 4443–4453

Gardiner J (2013) The evolution and diversification of plant microtubule-associated proteins. Plant J **75:** 219–229

Gehl C, Waadt R, Kudla J, Mendel RR, Hänsch R (2009) New GATEWAY vectors for high throughput analyses of protein-protein interactions by bimolecular fluorescence complementation. Mol Plant **2:** 1051–1058

Goldenberg NM, Steinberg BE (2010) Surface charge: a key determinant of protein localization and function. Cancer Res **70:** 1277–1280

Graham LE (1996) Green algae to land plants: an evolutionary transition. J Plant Res **109:** 241–251

Greß O, Möller B, Stöhr N, Hüttelmaier S, Posch S (2010) Scale-adaptive wavelet-based particle detection in microscopy images. *In* Proceedings of Workshop Bildverarbeitung für die Medizin. Springer, Heidelberg, Germany, pp 266–270

Gutierrez R, Lindeboom JJ, Paredez AR, Emons AM, Ehrhardt DW (2009) Arabidopsis cortical microtubules position cellulose synthase delivery to the plasma membrane and interact with cellulose synthase trafficking compartments. Nat Cell Biol **11:** 797–806

Hamada T, Nagasaki-Takeuchi N, Kato T, Fujiwara M, Sonobe S, Fukao Y, Hashimoto T (2013) Purification and characterization of novel microtubule-associated proteins from Arabidopsis cell suspension cultures. Plant Physiol **163:** 1804–1816

Hashimoto K, Kudla J (2011) Calcium decoding mechanisms in plants. Biochimie **93:** 2054–2059

Hemsley PA, Grierson CS (2008) Multiple roles for protein palmitoylation in plants. Trends Plant Sci **13:** 295–302

Hemsley PA, Weimar T, Lilley KS, Dupree P, Grierson CS (2013) A proteomic approach identifies many novel palmitoylated proteins in Arabidopsis. New Phytol **197:** 805–814

Hepler PK (1992) Calcium and mitosis. Int Rev Cytol **138:** 239–268

Hepler PK (2005) Calcium: a central regulator of plant growth and development. Plant Cell **17:** 2142–2155

Hepler PK (2016) The cytoskeleton and its regulation by calcium and protons. Plant Physiol **170:** 3–22

Huang Z, Van Houten J, Gonzalez G, Xiao H, van der Knaap E (2013) Genome-wide identification, phylogeny and expression analysis of *SUN, OFP* and *YABBY* gene family in tomato. Mol Genet Genomics **288:** 111–129

Hülskamp M, Misŕa S, Jürgens G (1994) Genetic dissection of trichome cell development in Arabidopsis. Cell **76:** 555–566

Humphrey TV, Haasen KE, Aldea-Brydges MG, Sun H, Zayed Y, Indriolo E, Goring DR (2015) PERK-KIPK-KCBP signalling negatively regulates root growth in Arabidopsis thaliana. J Exp Bot **66:** 71–83

Idilli AI, Morandini P, Onelli E, Rodighiero S, Caccianiga M, Moscatelli A (2013) Microtubule depolymerization affects endocytosis and exocytosis in the tip and influences endosome movement in tobacco pollen tubes. Mol Plant **6:** 1109–1130

Ishida T, Kaneko Y, Iwano M, Hashimoto T (2007) Helical microtubule arrays in a collection of twisting tubulin mutants of *Arabidopsis thaliana*. Proc Natl Acad Sci USA **104:** 8544–8549

Ivakov A, Persson S (2013) Plant cell shape: modulators and measurements. Front Plant Sci **4:** 439

Jarsch IK, Konrad SSA, Stratil TF, Urbanus SL, Szymanski W, Braun P, Braun KH, Ott T (2014) Plasma membranes are subcompartmentalized into a plethora of coexisting and diverse microdomains in *Arabidopsis* and *Nicotiana benthamiana*. Plant Cell **26:** 1698–1711

Jurado LA, Chockalingam PS, Jarrett HW (1999) Apocalmodulin. Physiol Rev **79:** 661–682

Kao YL, Deavours BE, Phelps KK, Walker RA, Reddy AS (2000) Bundling of microtubules by motor and tail domains of a kinesin-like calmodulin-binding protein from Arabidopsis: regulation by Ca^{2+}/calmodulin. Biochem Biophys Res Commun **267:** 201–207

Karimi M, Inzé D, Depicker A (2002) GATEWAY vectors for *Agrobacterium*-mediated plant transformation. Trends Plant Sci **7:** 193–195

Kirik V, Herrmann U, Parupalli C, Sedbrook JC, Ehrhardt DW, Hülskamp M (2007) CLASP localizes in two discrete patterns on cortical microtubules and is required for cell morphogenesis and cell division in Arabidopsis. J Cell Sci **120:** 4416–4425

Kong Z, Ioki M, Braybrook S, Li S, Ye ZH, Julie Lee YR, Hotta T, Chang A, Tian J, Wang G, et al (2015) Kinesin-4 functions in vesicular transport on cortical microtubules and regulates cell wall mechanics during cell elongation in plants. Mol Plant **8:** 1011–1023

Konrad SS, Popp C, Stratil TF, Jarsch IK, Thallmair V, Folgmann J, Marín M, Ott T (2014) S-Acylation anchors remorin proteins to the plasma membrane but does not primarily determine their localization in membrane microdomains. New Phytol **203:** 758–769

Konrad SSA, Ott T (2015) Molecular principles of membrane microdomain targeting in plants. Trends Plant Sci **20:** 351–361

Kosugi S, Hasebe M, Matsumura N, Takashima H, Miyamoto-Sato E, Tomita M, Yanagawa H (2009) Six classes of nuclear localization signals specific to different binding grooves of importin alpha. J Biol Chem **284:** 478–485

Kudla J, Batistic O, Hashimoto K (2010) Calcium signals: the lead currency of plant information processing. Plant Cell **22:** 541–563

Kudla J, Bock R (2016) Lighting the way to protein-protein interactions: recommendations on best practices for bimolecular fluorescence complementation analyses. Plant Cell **28:** 1002–1008

Kusumi A, Fujiwara TK, Chadda R, Xie M, Tsunoyama TA, Kalay Z, Kasai RS, Suzuki KG (2012) Dynamic organizing principles of the plasma membrane that regulate signal transduction: commemorating the fortieth anniversary of Singer and Nicolson's fluid-mosaic model. Annu Rev Cell Dev Biol **28:** 215–250

Lee YK, Kim GT, Kim IJ, Park J, Kwak SS, Choi G, Chung WI (2006) *LONGIFOLIA1* and *LONGIFOLIA2*, two homologous genes, regulate longitudinal cell elongation in *Arabidopsis*. Development **133:** 4305–4314

Levy M, Wang Q, Kaspi R, Parrella MP, Abel S (2005) *Arabidopsis* IQD1, a novel calmodulin-binding nuclear protein, stimulates glucosinolate accumulation and plant defense. Plant J **43:** 79–96

Lincoln C, Britton JH, Estelle M (1990) Growth and development of the *axr1* mutants of *Arabidopsis*. Plant Cell **2:** 1071–1080

Lindeboom JJ, Nakamura M, Hibbel A, Shundyak K, Gutierrez R, Ketelaar T, Emons AM, Mulder BM, Kirik V, Ehrhardt DW (2013) A mechanism for reorientation of cortical microtubule arrays driven by microtubule severing. Science **342:** 1245533

Liu Z, Persson S, Zhang Y (2015) The connection of cytoskeletal network with plasma membrane and the cell wall. J Integr Plant Biol **57:** 330–340

Liu Z, Schneider R, Kesten C, Zhang Y, Somssich M, Zhang Y, Fernie AR, Persson S (2016) Cellulose-microtubule uncoupling proteins prevent lateral displacement of microtubules during cellulose synthesis in Arabidopsis. Dev Cell **38:** 305–315

Lloyd C, Hussey P (2001) Microtubule-associated proteins in plants: why we need a MAP. Nat Rev Mol Cell Biol **2:** 40–47

Lucas JR, Courtney S, Hassfurder M, Dhingra S, Bryant A, Shaw SL (2011) Microtubule-associated proteins MAP65-1 and MAP65-2 positively regulate axial cell growth in etiolated *Arabidopsis* hypocotyls. Plant Cell **23:** 1889–1903

Ma H, Feng L, Chen Z, Chen X, Zhao H, Xiang Y (2014) Genome-wide identification and expression analysis of the IQD gene family in Populus trichocarpa. Plant Sci **229:** 96–110

Malinsky J, Opekarová M, Grossmann G, Tanner W (2013) Membrane microdomains, rafts, and detergent-resistant membranes in plants and fungi. Annu Rev Plant Biol **64:** 501–529

Mangano S, Gonzalez CD, Petruccelli S (2014) *Agrobacterium tumefaciens*-mediated transient transformation of *Arabidopsis thaliana* leaves. Methods Mol Biol **1062:** 165–173

Marc J, Granger CL, Brincat J, Fisher DD, Kao T, McCubbin AG, Cyr RJ (1998) A GFP-MAP4 reporter gene for visualizing cortical microtubule rearrangements in living epidermal cells. Plant Cell **10:** 1927–1940

Marchand V, Gaspar I, Ephrussi A (2012) An intracellular transmission control protocol: assembly and transport of ribonucleoprotein complexes. Curr Opin Cell Biol **24:** 202–210

McCormack E, Braam J (2003) Calmodulins and related potential calcium sensors of *Arabidopsis*. New Phytol **159:** 585–598

McCormack E, Tsai YC, Braam J (2005) Handling calcium signaling: Arabidopsis CaMs and CMLs. Trends Plant Sci **10:** 383–389

Mishima M, Maesaki R, Kasa M, Watanabe T, Fukata M, Kaibuchi K, Hakoshima T (2007) Structural basis for tubulin recognition by cytoplasmic linker protein 170 and its autoinhibition. Proc Natl Acad Sci USA **104:** 10346–10351

Möller B, Glaß M, Misiak D, Posch S (2016) MiToBo: a toolbox for image processing and analysis. J Open Res Softw **4:** e17

Möller B, Piltz E, Bley N (2014) Quantification of actin structures using unsupervised pattern analysis techniques. Proc Int Conf Pattern Recognition ICPR 14, IEEE, IEEE, 3251–3256

Müller J, Toev T, Heisters M, Teller J, Moore KL, Hause G, Dinesh DC, Bürstenbinder K, Abel S (2015) Iron-dependent callose deposition adjusts root meristem maintenance to phosphate availability. Dev Cell **33:** 216–230

Nakagawa T, Suzuki T, Murata S, Nakamura S, Hino T, Maeo K, Tabata R, Kawai T, Tanaka K, Niwa Y, et al (2007) Improved Gateway binary vectors: high-performance vectors for creation of fusion constructs in transgenic analysis of plants. Biosci Biotechnol Biochem **71:** 2095–2100

Narasimhulu SB, Reddy AS (1998) Characterization of microtubule binding domains in the *Arabidopsis* kinesin-like calmodulin binding protein. Plant Cell **10:** 957–965

Nigg EA (1997) Nucleocytoplasmic transport: signals, mechanisms and regulation. Nature **386:** 779–787

Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE T Pattern Anal **24:** 971–987

Oppenheimer DG, Pollock MA, Vacik J, Szymanski DB, Ericson B, Feldmann K, Marks MD (1997) Essential role of a kinesin-like protein in Arabidopsis trichome morphogenesis. Proc Natl Acad Sci USA **94:** 6261–6266

Palm D, Simm S, Darm K, Weis BL, Ruprecht M, Schleiff E, Scharf C (2016) Proteome distribution between nucleoplasm and nucleolus and its relation to ribosome biogenesis in Arabidopsis thaliana. RNA Biol **13:** 441–454

Raffaele S, Mongrand S, Gamas P, Niebel A, Ott T (2007) Genome-wide annotation of remorins, a plant-specific protein family: evolutionary and functional perspectives. Plant Physiol **145:** 593–600

Reddy AS, Ben-Hur A, Day IS (2011) Experimental and computational approaches for the study of calmodulin interactions. Phytochemistry **72:** 1007–1019

Ren J, Wen L, Gao X, Jin C, Xue Y, Yao X (2008) CSS-Palm 2.0: an updated software for palmitoylation sites prediction. Protein Eng Des Sel **21:** 639–644

Roll-Mecak A (2015) Intrinsically disordered tubulin tails: complex tuners of microtubule functions? Semin Cell Dev Biol **37:** 11–19

Saka SK, Honigmann A, Eggeling C, Hell SW, Lang T, Rizzoli SO (2014) Multi-protein assemblies underlie the mesoscale organization of the plasma membrane. Nat Commun **5:** 4509

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al (2012) Fiji: an open-source platform for biological-image analysis. Nat Methods **9:** 676–682

Scholl RL, May ST, Ware DH (2000) Seed and molecular resources for Arabidopsis. Plant Physiol **124:** 1477–1480

Scott AM, Antal CE, Newton AC (2013) Electrostatic and hydrophobic interactions differentially tune membrane binding kinetics of the C2 domain of protein kinase Cα. J Biol Chem **288:** 16905–16915

Sedbrook JC, Ehrhardt DW, Fisher SE, Scheible WR, Somerville CR (2004) The *Arabidopsis sku6/spiral1* gene encodes a plus end-localized microtubule-interacting protein involved in directional cell expansion. Plant Cell **16:** 1506–1520

Sedbrook JC, Kaloriti D (2008) Microtubules, MAPs and plant directional cell expansion. Trends Plant Sci **13:** 303–310

Shahbabian K, Chartrand P (2012) Control of cytoplasmic mRNA localization. Cell Mol Life Sci **69:** 535–552

Shannon KB (2012) IQGAP family members in yeast, Dictyostelium, and mammalian cells. Int J Cell Biol **2012:** 894817

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res **13:** 2498–2504

Sheahan MB, Staiger CJ, Rose RJ, McCurdy DW (2004) A green fluorescent protein fusion to actin-binding domain 2 of Arabidopsis fimbrin highlights new features of a dynamic actin cytoskeleton in live plant cells. Plant Physiol **136:** 3968–3978

Shoji T, Narita NN, Hayashi K, Asada J, Hamada T, Sonobe S, Nakajima K, Hashimoto T (2004) Plant-specific microtubule-associated protein SPIRAL2 is required for anisotropic growth in Arabidopsis. Plant Physiol **136:** 3933–3944

Smith JM, Hedman AC, Sacks DB (2015) IQGAPs choreograph cellular signaling from the membrane to the nucleus. Trends Cell Biol **25:** 171–184

Smith LG, Gerttula SM, Han S, Levy J (2001) Tangled1: a microtubule binding protein required for the spatial control of cytokinesis in maize. J Cell Biol **152:** 231–236

Spira F, Mueller NS, Beck G, von Olshausen P, Beig J, Wedlich-Söldner R (2012) Patchwork organization of the yeast plasma membrane into numerous coexisting domains. Nat Cell Biol **14:** 640–648

Steinhorst L, Kudla J (2013) Calcium: a central regulator of pollen germination and tube growth. Biochim Biophys Acta **1833:** 1573–1581

Struk S, Dhonukshe P (2014) MAPs: cellular navigators for microtubule array orientations in Arabidopsis. Plant Cell Rep **33:** 1–21

Tian J, Han L, Feng Z, Wang G, Liu W, Ma Y, Yu Y, Kong Z (2015) Orchestration of microtubules and the actin cytoskeleton in trichome cell shape determination by a plant-unique kinesin. eLife **4:** 4

Tompa P, Davey NE, Gibson TJ, Babu MM (2014) A million peptide motifs for the molecular biologist. Mol Cell **55:** 161–169

Toufighi K, Brady SM, Austin R, Ly E, Provart NJ (2005) The Botany Array Resource: e-northerns, expression angling, and promoter analyses. Plant J **43:** 153–163

van der Knaap E, Chakrabarti M, Chu YH, Clevenger JP, Illa-Berenguer E, Huang Z, Keyhaninejad N, Mu Q, Sun L, Wang Y, et al (2014) What lies beyond the eye: the molecular mechanisms regulating tomato fruit weight and shape. Front Plant Sci **5:** 227

Vavouri T, Semple JI, Garcia-Verdugo R, Lehner B (2009) Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. Cell **138:** 198–208

Wang L, Li H, Lv X, Chen T, Li R, Xue Y, Jiang J, Jin B, Baluška F, Šamaj J, et al (2015) Spatiotemporal dynamics of the BRI1 receptor and its regulation by membrane microdomains in living Arabidopsis cells. Mol Plant **8:** 1334–1349

Wang YC, Wang BC, Gilroy S, Chehab EW, Braam J (2011) CML24 is involved in root mechanoresponses and cortical microtubule orientation in Arabidopsis. J Plant Growth Regul **30:** 467–479

Wang YS, Motes CM, Mohamalawari DR, Blancaflor EB (2004) Green fluorescent protein fusions to Arabidopsis fimbrin 1 for spatio-temporal imaging of F-actin dynamics in roots. Cell Motil Cytoskeleton **59:** 79–93

Wright PE, Dyson HJ (2015) Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol **16:** 18–29

Wu M, Li Y, Chen D, Liu H, Zhu D, Xiang Y (2016) Genome-wide identification and expression analysis of the IQD gene family in moso bamboo (Phyllostachys edulis). Sci Rep **6:** 24520

Wu S, Xiao H, Cabrera A, Meulia T, van der Knaap E (2011) SUN regulates vegetative and reproductive organ shape by changing cell division patterns. Plant Physiol **157:** 1175–1186

Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E (2008) A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. Science **319:** 1527–1530

Xie Y, Zheng Y, Li H, Luo X, He Z, Cao S, Shi Y, Zhao Q, Xue Y, Zuo Z, et al (2016) GPS-Lipid: a robust tool for the prediction of multiple lipid modification sites. Sci Rep **6:** 28249

Yang T, Poovaiah BW (2003) Calcium/calmodulin-mediated signal network in plants. Trends Plant Sci **8:** 505–512

Zárský V, Cvrcková F, Potocký M, Hála M (2009) Exocytosis and cell polarity in plants: exocyst and recycling domains. New Phytol **183:** 255–272

Zentella R, Zhang ZL, Park M, Thomas SG, Endo A, Murase K, Fleet CM, Jikumaru Y, Nambara E, Kamiya Y, et al (2007) Global analysis of DELLA direct targets in early gibberellin signaling in *Arabidopsis*. Plant Cell **19:** 3037–3057

Zhu C, Ganguly A, Baskin TI, McClosky DD, Anderson CT, Foster C, Meunier KA, Okamoto R, Berg H, Dixit R (2015a) The Fragile Fiber1 kinesin contributes to cortical microtubule-mediated trafficking of cell wall components. Plant Physiol **167:** 780–792

Zhu X, Dunand C, Snedden W, Galaud JP (2015b) CaM and CML emergence in the green lineage. Trends Plant Sci **20:** 483–489

4.4 Quantitative and Comparative Analysis of Global Patterns of (Microtubule) Cytoskeleton Organization with CytoskeletonAnalyzer2D

Abstract:

The microtubule cytoskeleton plays important roles in cell morphogenesis. To investigate the mechanisms of cytoskeletal organization, for example, during growth or development, in genetic studies, or in response to environmental stimuli, image analysis tools for quantitative assessment are needed. Here, we present a method for texture measure-based quantification and comparative analysis of global microtubule cytoskeleton patterns and subsequent visualization of output data. In contrast to other approaches that focus on the extraction of individual cytoskeletal fibers and analysis of their orientation relative to the growth axis, CytoskeletonAnalyzer2D quantifies cytoskeletal organization based on the analysis of local binary patterns. CytoskeletonAnalyzer2D thus is particularly well suited to study cytoskeletal organization in cells where individual fibers are difficult to extract or which lack a clearly defined growth axis, such as leaf epidermal pavement cells. The tool is available as ImageJ plugin and can be combined with publicly available software and tools, such as R and Cytoscape, to visualize similarity networks of cytoskeletal patterns.

Link:

https://doi.org/10.1007/978-1-4939-9469-4_10

Chapter 5

Bioimage Analysis Software & Tools

5.1 Knowing what happened - Automatic Documentation of Image Analysis Processes

Abstract:

Proper archiving or later reconstruction and verification of results in data analysis requires thorough logging of all manipulative actions on the data and corresponding parameter settings. Unfortunately such documentation tasks often enforce extensive and error prone manual activities by the user. To overcome these problems we present Alida, an approach for fully automatic documentation of data analysis procedures. Based on an unified operator interface all operations on data including their sequence and configurations are registered during analysis. Subsequently these data are made explicit in XML graph representations yielding a suitable base for visual and analytic inspection. As example for the application of Alida in practice we present MiToBo, a toolbox for image analysis implemented on the basis of Alida and demonstrating the advantages of automatic documentation for image analysis procedures.

Link:

https://doi.org/10.1007/978-3-642-23968-7_1

248-258

A Framework Unifying the Development of Image Analysis Algorithms and Associated User Interfaces

Birgit Möller and Stefan Posch
Institute of Computer Science, Martin-Luther University Halle-Wittenberg,
Von-Seckendorff-Platz 1, 06099 Halle (Saale), Germany
{`birgit.moeller,stefan.posch`}`@informatik.uni-halle.de`

Abstract

Solving image analysis problems not only requires the development of suitable sets of algorithms to produce desired result data, but also demands for suitable user interfaces (UIs) to foster use in practice. Here we present our library Alida *which aims to promote the deployment of UIs by featuring their automatic generation from algorithm code.* Alida *supports command line and graphical UIs (GUIs), and ships with a graphical editor for designing more complex workflows. Enforcing only a small set of rules to obey* Alida *significantly reduces implementation overhead for developers and allows for focusing on algorithm rather than UI design. The suitability of* Alida*'s concept for real-life applications is shown by the library* MiToBo *for biomedical image analysis implemented based on* Alida.

1   Introduction

The process of automatically analyzing image data can generally be understood as a sequence of individual analysis steps applied to data. These steps are performed sequentially, in parallel, or in a nested fashion and transform given input data into application-specific result data. Often each single step of such analysis pipelines is associated with a specific processing unit or *operator* implementing functionality and performing the actual work on the data.

Consequently, the development of complex image analysis pipelines comprises two issues. On the one hand it subsumes the development of suitable operators, and on the other hand it requires their combination into pipelines through which the data is propagated to solve a task at hand. Besides these two fundamental issues the availability of suitable UIs on the programming as well as on the user level is equally important. Particularly the latter ones are inevitable to pave the path for newly developed algorithms into practice where non-expert users are highly interested in using the software to solve their problems.

Although a great consent is to be expected within the community regarding the necessity of handy user interfaces to be available, nevertheless their development is too often neglected in practice. The main reason for this is the significant amount of additional workload for interface design and implementation.

In this paper we present an approach for closing the gap between algorithm and pipeline design on the one side, and the development of UIs on the other. Our key contribution is to provide a development environment in terms of a Java library which inherently integrates the development of algorithms and UIs, and

which is publicly available under GPL license[1]. This library named *Alida* defines operators as key components and requests developers to implement these by following general rules. Each operator has to support inquiries for its input and output parameter data types and values. In addition operators have to be invoked by a unified procedure enabling generic handling of all operators.

These guidelines may seem to restrict developers in their design options, but they only enforce a minimal set of rules as will become obvious in Sec. 3. Moreover, by embedding new operators into the environment provided by *Alida* all operators are natively accessible through a unified API on the programming level, rendering their use within code quite easy. Finally, these rules lay the foundation for automatically generating graphical as well as command line UIs for all operators in a generic fashion (Sec. 4). They also enable the export of operators into more complex applications like graphical editors for pipelines in a straightforward manner. Accordingly, such an editor is also included in *Alida* and consequently extends the comfortable development and application of single operators towards user-friendly graphical pipeline design (Sec. 5).

2   Related Work

For many image analysis problems well-established algorithmic solutions have emerged over time. They are often collected in libraries like OpenCV [3] or ITK [6], or in commercial tools like Matlab. While this fosters wide-spread use, the lack of handy user interfaces still forms a barrier. Although OpenCV and Matlab provide basic components like windows and buttons for interaction, the implementation of GUIs is still left to the developer of a certain application.

For the development in Java libraries and toolkits like JAI [11] and ImageJ [1], which is a toolbox of image analysis algorithms widely-used withing the biomedical community, aim to simplify operator development providing an embedding framework. While JAI is focused on unifying parameter handling and operator invocation, the new version 2.0 of ImageJ[2] also includes mechanisms for automatic generation of GUIs. Our concept *Alida* surpasses this by featuring advanced concepts for handling operator parameters (cf. Sec. 4), handy command-line tools, and support for automatic process documentation [9].

For designing analysis pipelines various tools are available. While the graphical workflow editor Kepler[3] does not target at a specific field of application,

---

[1] http://www.informatik.uni-halle.de/alida/
[2] http://developer.imagej.net/
[3] https://kepler-project.org/

KNIME [2] is dedicated to data mining and natively supports image processing. However, extending these tools with new functionality is not always straightforward. Our library *Alida* is designed for data analysis in general, but we have also deployed the image processing library *MiToBo* based upon *Alida* and dedicated to biomedical imaging which proves the suitability of our concept in practice (Sec. 6). It is deeply integrated into ImageJ overcoming the need for non-expert users to adapt to a completely new environment for applying operators and developing their workflows.

3 Basic Concepts

As outlined in the introduction, *Alida* defines *operators* as the basic units for data analysis. Taking an object-oriented approach an operator is expected to extend the common super class `ALDOperator` and to implement its functionality by overriding specific methods.

All data to be processed, controlling manipulation, or to be returned as result from an operator are consistently denoted as *parameters*. The role of a parameter is specified via its *direction*, which may be IN, OUT, or INOUT. A typical example is a filter applied to an input image (direction IN) where the filtered image is returned in a newly allocated data structure as a parameter with direction OUT. If the filter acts destructively in place, this is described by a single parameter of direction INOUT. A parameter controlling the filter operation, e.g. a bandwidth, is provided as an IN parameter. Parameters of direction IN and INOUT may be either *required* or *optional* to simplify operator configuration. In addition *Alida* supports *supplemental* parameters which in contrast to required and optional ones must not influence processing results. Examples include flags to control debug messages or intermediate results to be returned from operators after execution.

For defining and accessing the parameters of an operator *Alida* makes use of Java's annotation mechanism. Parameters can easily be added by simply annotating member variables of the operator class. In addition, the annotations also allow to easily query an operator for all its parameters including their type, role and current value at run time using methods supplied by the common super class `ALDOperator`.

To make use of an operator's functionality an instance of the class needs to be created and its IN and INOUT parameters have to be set. Processing is invoked calling the generic `runOp()` method supplied by `ALDOperator`. This method subsumes an automatic parameter validation, e.g. it checks whether all required parameters are provided, and optionally enforces operator-specific constraints, e.g. admissible ranges of parameter values. Subsequently the operator's `operate()` method is called performing data manipulations, and finally the results are made available via the output parameters of the operator.

As outlined in the introduction image analysis problems usually require a combination of operators to be applied to data rather than only a single one. Consequently *Alida* natively supports the development of complete pipelines for data processing by providing the class `ALDWorkflow`. This class represents a processing pipeline as a set of operators. It allows for establishing links between OUT and IN parameters of different operators by which a flow of data and control and thus a pipeline can be realized.

A workflow in *Alida* also features entry and exit points for data into and out of the whole pipeline. These points have essentially the same role as parameters for operators, hence, they share exactly the same properties, e.g., have a direction and may be required or optional. This naturally implies to implement `ALDWorkflow` extending the super class `ALDOperator` of all operators. The `operate()` method of a workflow object invokes all included operators in topological order and forwards output data between operators according to the data flow. To facilitate graphical programming, the class `ALDWorkflow` also supplies methods to invoke only part of the processing pipeline, to save and load workflows, and offers an event mechanism for GUI components to register, thus facilitating a Model-View-Controller architecture (cf. Sec. 5).

4 Automatic User Interface Generation

The operator concept of *Alida* with its clearly defined specification of parameters and its standardized invocation procedure lays the basis for generic implementation of UIs. To facilitate generic configuration and execution of operators such UIs are required to allow for the input of parameter values, to invoke the operator, and finally to publish the results.

In *Alida* the *Model-View-Control* design pattern [4] is used to achieve maximal independence between the operators implementing the functionality, the I/O of data objects, and the graphical and textual UIs. As input and output mechanisms of individual data items are data type specific, I/O functionality is also completely encapsulated in data I/O *providers* hiding any data type specific knowledge from the generic viewers.

To endorse the development of new functionality *Alida* already includes various providers which facilitates I/O for a wide variety of Java objects out of the box and overcomes the need for programmers to implement I/O capabilities. Besides providers for all primitive data types and arrays *Alida* subsumes general purpose providers for all enumeration types, collections, and so-called *parametrized classes*. An arbitrary class may be declared as parametrized class, and any subset of its member variables as *class parameters*, both via annotations. This is sufficient for *Alida*'s general purpose providers to handle this class as an operator parameter. Likewise operator objects by itself may act as parameters of other operators. Only specialized classes like images or contour sets require additional providers to be implemented, but these can easily be added to the library with no need to modify the code of *Alida*.

4.1 Command line

Building on this infrastructure *Alida* features a command line operator runner (CLR) to invoke all operators via console or scripts. All input parameters are supplied as arguments by 'name=value' pairs. To ease the handling of class inheritance, parametrized classes and operators as parameters in a generic fashion, the CLR features a flexible parser for argument preprocessing. It allows for parsing CLR calls like the one shown in Fig. 1. The operator `SnakeOptimizerCoupled`

```
java OpRunner SnakeOptimizerCoupled initialSnakes=RoiSet.xml inImg=cell.tif
  outSnakes=snakesOut.xml snakeOptimizer='$SnakeOptimizerSingleVarCalc:{energySet=
  {energies=[$MTBSnakeEnergyCD_CVRegionFit:{lambda_in=1.0,lambda_out=5.0}],weights=[1.0]}}'
```

Figure 1. Example call of an operator from command line. The operator `SnakeOptimizerCoupled` called here among others takes an operator of type `SnakeOptimizerSingleVarCalc` as input parameter.

for multiple snake segmentation not only takes initial snakes and an image as input, but also an operator instance of `SnakeOptimizerSingleVarCalc` dealing with a single snake (cf. Fig. 2, right). The syntax of the individual value strings is defined by the specific I/O providers they are finally passed to and which, by convention, also allow values to be read from file. Analogously output parameters are specified as name–value pairs allowing to, e.g., redirect output into files, as an alternative to formatting the values onto standard out.

4.2 Graphical user interface

A GUI is supposed to support graphical configuration and execution of operators. Unlike the command line UI a graphical front-end to choose, configure and execute operators is also well-suited for interactive exploration of image processing, e.g. for online inspection of the effects of parameter changes during rapid-prototyping. The GUIs currently provided by *Alida* are implemented based on Swing, but this is not a fundamental restriction as other frameworks likewise web interfaces can be added in a straightforward way.

As soon as the user has selected an operator from the choice of available operators, a window to configure it and control its execution is automatically generated (cf. Fig. 2 on the right, which shows the GUI for the operator `SnakeOptimizerSingleVarCalc` introduced above). To this end the operator is first queried for the types of its input and output parameters. Subsequently for each parameter a corresponding graphical component is generated using the provider mechanism outlined above. All components are arranged in a frame whereas required, optional and supplemental parameters are grouped into different sections. Besides these components for operator configuration allowing for user inputs, the window also contains buttons for controlling operator execution and enabling interaction (if the operator supports that). In addition a menu-bar is available containing items, e.g., for accessing online help or saving and loading the configuration to and from file. After execution of an operator result data is displayed to the user again adopting the provider mechanism for generating graphical components for each output parameter.

The configuration and control window acts as controller and likewise observer of the underlying operator and its status. The configuration status of the operator is synchronized online with the window and vice versa. E.g., parameters that are required, but do not yet have suitable values are marked in red, and the color of the run button indicates whether the operator is ready for execution or not. On the other hand changes in the parameters made by the user are directly propagated to the operator and induce an instant update of its configuration and potentially also its status.

5 Graphical Programming

Designing more complex analysis pipelines featuring the combination of various operators can be facilitated in a comfortable, intuitive and user-friendly way by graphical programming editors. The underlying idea of these tools is to translate the design process into a graph editing task. Operators are represented by nodes, and edges in-between indicate the flow of data and control. Obviously *Alida* is an optimal foundation for such an editor. Its workflow concept is natively qualified as basis since it inherently defines a computational model of workflows. And also the clearly defined concept of operators renders it very easy to adopt operators as basis for configurable nodes within a graph data structure associated with a processing pipeline. They are handled in a generic fashion similar as in the context of UI generation discussed earlier.

The graphical editor *Grappa* included in *Alida* is built on these ideas (Fig. 2). It automatically includes all available operators into a menu (left of Fig. 2) from where the user can select operators for a workflow. For each operator a corresponding node is generated on the workbench showing the parameters of the operator as ports. These ports can be linked by edges resulting in a working cycle intuitive also for non-experts. For graphical node configuration each node is linked to a configuration window where the same components and mechanisms as for automatic GUI generation are used. Once pipeline design and configuration are completed the workflow graph can be executed either completely or in parts, e.g., only up to a certain node. After termination the results are displayed to the user again reusing mechanisms from GUI generation. Similar to the concepts of graphical operator configuration and control, the editor acts as controller and observer. The node color is updated according to the current status of the underlying operator in synchrony with changes in the configuration window which yields an intuitive visual guidance in pipeline configuration and execution.

6 Alida in Practice

The development of image analysis algorithms for biomedical applications natively requires deep links between algorithm developers and (non computer scientist) users. Algorithmic improvements and adaptations can most of the time best be done based on direct practical evaluations in the lab, and the frequent introduction of new imaging techniques and types of data both request for topical availability of appropriate software.

We meet these requirements by implementing our image analysis algorithms based on *Alida*, i.e. each algorithm is implemented as operator. The set of all operators is collected in the toolbox *MiToBo* which inherently supports execution of all available operators from command line as well as graphically. For straightforward usage the graphical front-end to select,
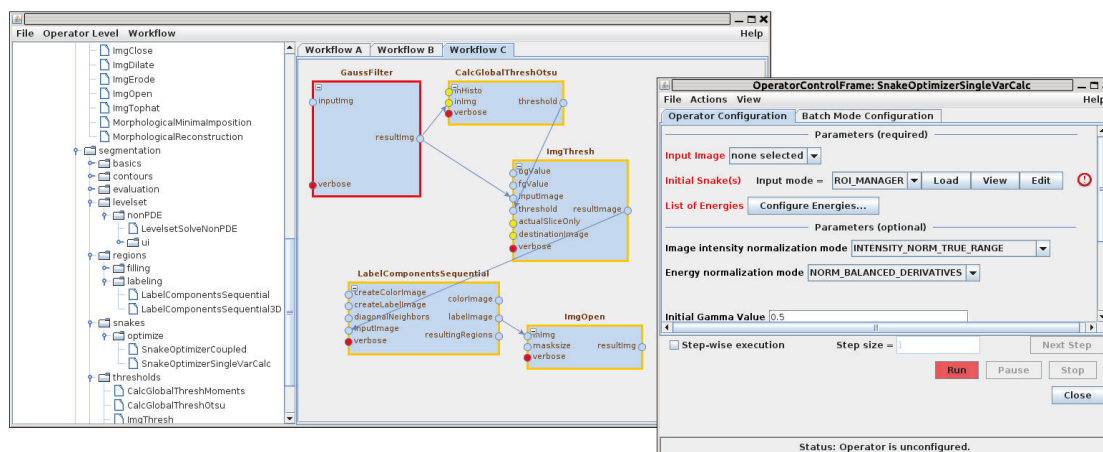
Figure 2. Screen-shot of the graphical editor for user-friendly pipeline design with its selection menu and workbench (left), and an automatically generated configuration and control window for an operator (right).

configure and execute operators is available as plugin[4] for ImageJ, a tool extensively used by our biomedical cooperation partners. This way algorithmic improvements and extensions can directly be released for practice in terms of new ImageJ plugins. By this users get software for current problems soon, and the developers earn topical feedback, in conclusion yielding great benefits for both sides.

Over the years a broad collection of algorithms for basic image processing tasks like morphological operations, filtering or labeling, as well as for specific areas of application, e.g., for sub-cellular particle detection [8], cell segmentation based on active contours [10, 7], or analysis of scratch assay images [5] have been developed. In addition the toolbox subsumes a plugin version of the editor *Grappa* (Sec. 5). Besides, non-expert users also benefit from some user-friendly built-in features of *Alida* not discussed here in detail, e.g., the automatic documentation of parameters used in analysis procedures simplifying communication between developers and users [9], or the option to filter as well the set of operators in the graphical selection menu as also the parameters of an operator displayed in the GUI.

7 Conclusions

Alida offers an integrated approach for unifying the development of image analysis algorithms and more complex analysis pipelines on the one hand and intuitive user interfaces on the other. Its underlying concept of operators lays the foundation for easy use of operators through a unified API on the programming level as well as for automatic generation of command line and graphical user interfaces on the user level. The concept results in a great flexibility which also becomes obvious in the graphical editor *Grappa* included in *Alida*. For its implementation *Alida*'s functionality with regard to operator execution and pipeline design had to be made available through appropriate graphical components, however, did not induce any conceptual changes in the underlying core. And the flexibility is not yet exhausted. It is obvious that the workflows of *Alida* being by themselves operators already inherently

support the concept of hierarchical workflows which is currently being transferred to and implemented in the graphical editor. Finally, a batch mode is currently under development for automatically running operators on sets of values for specified input parameters to further ease algorithm tuning for developers and algorithm usage for non-expert users.

References

[1] Abramoff, M.D., Magelhaes, P.J., Ram, S.J.: Image processing with ImageJ. Biophotonics Int 11(7), 36–42 (2004)

[2] Berthold, M.R., et al.: KNIME - the Konstanz Information Miner: version 2.0 and beyond. SIGKDD Explor. Newsl. 11(1), 26–31 (Nov 2009)

[3] Bradski, A.: Learning OpenCV: Computer Vision with the OpenCV Library. O'Reilly Media (2008)

[4] Fowler, M.: Patterns of Enterprise Application Architecture. The Addison-Wesley Signature Series (2003)

[5] Glaß, M., et al.: Cell migration analysis: Segmenting scratch assay images with level sets and support vector machines. Pattern Recognition 45(9), 3154–3165 (2012)

[6] Ibanez, L., Schroeder, W., Ng, L., Cates, J.: The ITK Software Guide, 2. edn. (November 2005)

[7] Möller, B., Posch, S.: Comparing active contours for the segmentation of biomedical images. In: IEEE Int. Symp. on Biomedical Imaging. pp. 736–739 (2012)

[8] Möller, B., et al.: Adaptive segmentation of particles and cells for fluorescent microscope imaging. In: VISIGRAPP 2010, Revised Selected Papers of Int. Joint Conf. on Comp. Vision, Imaging and Comp. Graphics. Theory and Appl. vol. 229, pp. 154–169 (2011)

[9] Möller, B., Greß, O., Posch, S.: Knowing what happened - automatic documentation of image analysis processes. In: Proc. of Int. Conf. on Comp. Vision Systems. LNCS, vol. 6962, pp. 1–10. Springer (2011)

[10] Möller, B., Stöhr, N., Hüttelmaier, S., Posch, S.: Cascaded segmentation of grained cell tissue with active contour models. In: Proc. of Int. Conf. on Pattern Recognition (ICPR). pp. 1481–1484 (2010)

[11] Sun Microsystems, Palo Alto, CA 94303, USA: Programming in Java Advanced Imaging (1999), Rel. 1.0.1

---

[4]http://www.informatik.uni-halle.de/mitobo/

SOFTWARE METAPAPER

Alida – Advanced Library for Integrated Development of Data Analysis Applications

Stefan Posch and Birgit Möller

Institute of Computer Science, Faculty of the Natural Sciences III, Martin Luther University Halle-Wittenberg, Halle (Saale), DE
Corresponding author: Stefan Posch
(stefan.posch@informatik.uni-halle.de)

Data analysis procedures can often be modeled as a set of manipulation operations applied to input data and resulting in transformed intermediate and result data. The Java library `Alida` is providing an advanced development framework to support programmers in developing data analysis applications adhering to such a scheme. The main intention of `Alida` is to foster re-usability by offering well-defined, unified, modular APIs and execution procedures for operators, and to ease development by releasing developers from tedious tasks. `Alida` features automatic generation of handy graphical and command line user interfaces, a built-in graphical editor for workflow design, and an automatic documentation of analysis pipelines. `Alida` is available from its project webpage http://www.informatik.uni-halle.de/alida, on Github and via our Maven server.

(1) Overview

Introduction

Automatic data analysis aims at cleaning, transforming, and modelling data to gain useful information in an application domain and for a specific problem statement. This process frequently requires the combination of various basic and advanced analysis steps into complex workflows, and several software tools for workflow design supporting this process on the user side are available [4, 11, 10, 1]. They for example target at distributed, grid and cluster computing, big data analytics, or on integrating data from different sources [2], and sometimes provide end-users with functionality for graphically combining analysis units into analysis workflows. However, solely applying and combining existing algorithms is not always sufficient to extract desired information from given data. Especially as progress in science and research is often linked to designing new experiments and acquiring new types of experimental data, sophisticated analysis requires the adaptation of existing or the development and investigation of new data analysis algorithms.

The development of such tools is usually performed by programmers in close collaboration with end-users from the application side, and a close interaction during the development process is essential. Consequently and

independent of the application domain, the programmer is required to not only develop the algorithms themselves, but he is also enforced to provide handy user interfaces and integrate the user as close as possible into the development process, e.g., by frequently releasing software updates. Workflow tools like KNIME [1] and Triana [10] in principal support the extension of their functionality programmatically. However, since they mainly focus on the end-user programmers have to cope, e.g., with restrictions on available data types and complex APIs.

To overcome these drawbacks and in contrast to these tools, `Alida` (*Advanced Library for Integrated Development of Data Analysis Applications,* [7, 9]) is a Java library which specifically targets at programmers rather than end-users of data analysis tools. It seeks to optimally support programmers in the process of developing and releasing new data analysis algorithms in close collaboration with end-users. To this end `Alida` defines a framework which allows programmers to easily implement new data analysis functionality in a modular fashion. It defines an API based on a very general model of data analysis where manipulation and transformation of input data into intermediate and final result data is performed by *operators* with a certain functionality. Every operator is fully specified by a set of parameters subsuming input data and configuration settings
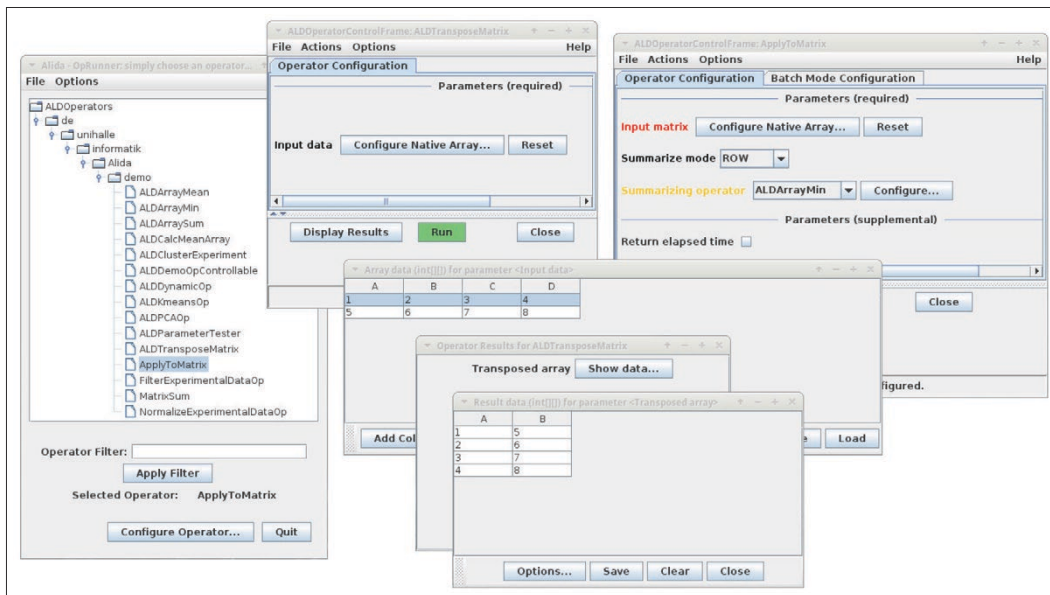
Figure 1: Screenshot of some automatically generated graphical user interfaces for configuration and execution of `Alida` operators.

for the functionality of the operator. During data analysis operators are applied sequentially, in parallel, or in a nested fashion to the input data and produce output data according to their configuration.

Based on this model `Alida` enforces only some few constraints on the implementation in order to release developers from reoccurring and tedious tasks like API design and user interface development. All operators share a common API for configuration and execution. On the one hand this facilitates reuse of operators on the code level and instant usage via the automatically generated command line user interface (see **Fig. 2**), e.g., for parameter optimization via a scripting language. On the other hand also graphical user interfaces are generated automatically (**Fig. 1**) fostering close end-user interaction and a tight feedback loop. Likewise all operators can automatically be included as potential building blocks in `Alida`'s built-in graphical workflow editor `Grappa` [3] (**Fig. 3**). Finally, since all operators are configured and executed by the same procedures automatic documentation of operator

configurations and consequently also complete analysis pipelines is supported [6, 8].

The basic concepts of the `Alida` framework and its implementation in the Java library have proven their practical suitability and relevance as fundament of MiToBo, a toolbox of basic, intermediate and advanced image processing and analysis operators and applications [5]. All of the more than 150 operators in MiToBo are implemented as `Alida` operators taking full benefit of the unified interfaces and execution procedures and particularly of the automatically generated user interfaces.

Implementation and Architecture

The abstract class `ALDOperator` lays the foundation for `Alida`'s object-oriented design for data analysis. It is designed to enable `Alida`'s capabilities to automatically generate user interfaces, for graphical programming, and automatic documentation.

All operators to be implemented in `Alida` are required to extend this class. All data to be processed by an operator,

```
rigel:alida [161] java \
?     de.unihalle.informatik.Alida.tools.ALDOpRunner ApplyMatrix \
?     matrix='[[1,2,3],[4,5,6]]' \
?     summarizeOp='$ALDArrayMean:' summarizeMode=ROW \
?     summaries=-
summaries = [2.0,5.0]
rigel:alida [162]
```

Figure 2: Sample call of an operator from command line. The operator `ApplyToMatrix` is executed to apply an operator to a 2D array supplied on the command line. The parameter `summarizeOp` expects as value an `ALDOperator` which summarizes a 1D array, and here `ALDArrayMean` is specified computing the mean. The parameter `summarizeMode` is of enumeration type, and in this case row-wise summarization requested. The output is sent to standard output, but can be redirected to a file as well.
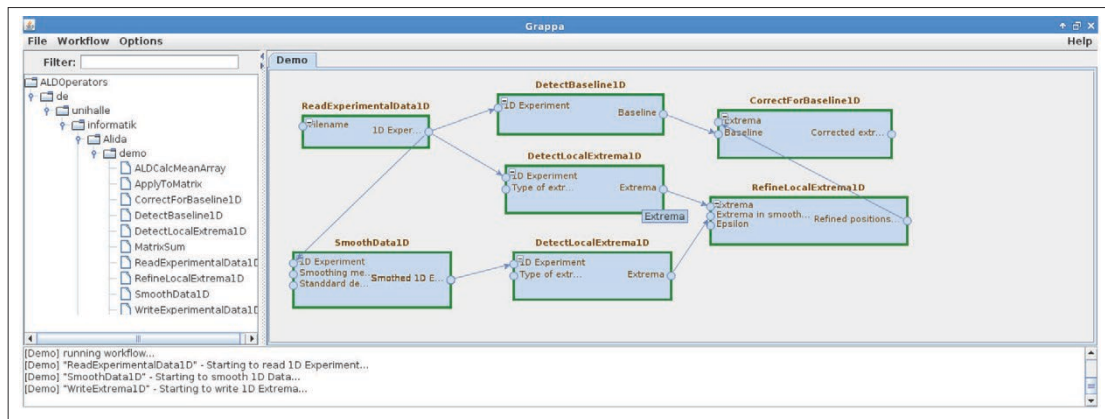
Figure 3: Screenshot of the graphical workflow editor `Grappa`, showing a demo workflow shipped with `Alida`. To the left a partial tree of available operators is displayed to choose from.

controlling its manipulation, or to be returned as result are consistently denoted as *parameters* in `Alida`. For each parameter a member variable is defined and Java's annotation mechanism is used to declare these members as parameters and specify their various properties. Java's reflection mechanism is exploited to implement methods for querying an operator for its parameters including data types and properties, as well as generic getter and setter methods for all parameters. The abstract method `operate()` of ALDOperator contains the data processing functionality and needs to be overridden by each operator implementation. The abstract class `ALDOperator` implements the method `runOp()` which is the only admissible way to invoke an operator. This allows to keep track of all operator invocations.

For all data processing algorithms implemented as `Alida` operators graphical and command line interfaces are instantly available to the users. To automatically generate these interfaces an operator needs to be queried for its parameters and their properties as stated above. In addition it is necessary to query values for parameters from the user, to instantiate parameter objects from these values, and to present output parameter values to the user, e.g., graphically or via console. As this depends on the specific data type and the set of potential parameter data types is unknown in advance, `Alida` incorporates a mechanism to link this I/O knowledge to specific data types. This is facilitated via so-called data I/O *providers* which provide the functionality for a given data type or set of data types and register to `Alida`'s framework using Java's annotations. Currently, `Alida` features general purpose providers for all primitive data types, enumeration types, arrays, collections, and so-called *parameterized classes*. An arbitrary class may be declared as parameterized class, and any subset of its member variables declared as *class parameters*, both via annotations. This is sufficient for `Alida`'s general purpose provider to handle this class as an operator parameter if providers for the class parameters exist.
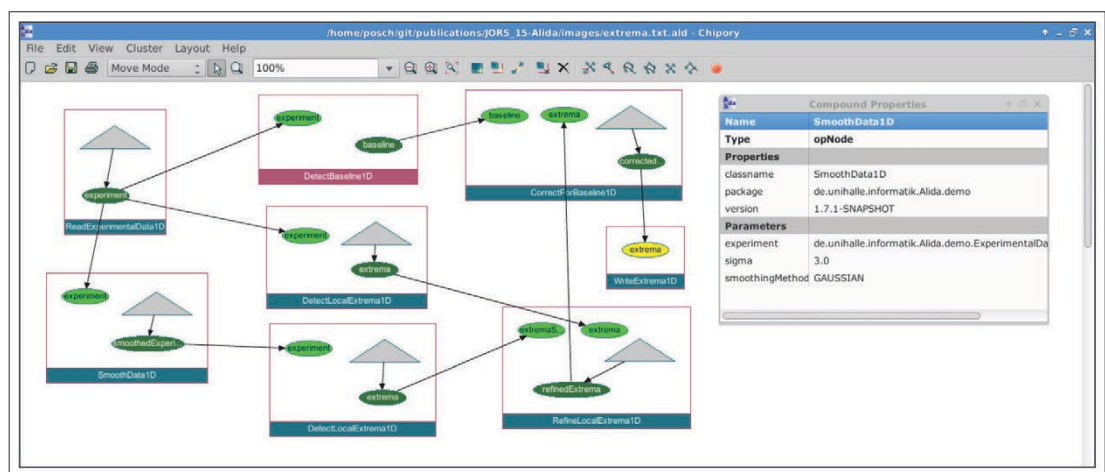
Figure 4: The processing graph for the workflow in Fig. 3. Each operator invocation is represented by a blue or red rectangle. A red rectangle indicates that an operator was *collapsed* to hide nested operator calls. Light and dark green ellipses are input and output ports respectively of an operator, gray triangles depict data ports representing newly generated data. To the right the information for the operator `SmooothData1D` is shown including the values of input parameters and software version.

Likewise operators may act as parameters of other operators. If necessary additional providers may easily be added without the necessity to modify `Alida`'s core. **Figs. 1** and **2** show examples for graphical respectively command line UIs automatically generated by `Alida`.

`Alida` extends the operator concept towards combining operators into more complex workflows. A workflow is defined as a combination of operators to be executed sequentially, in parallel, or in a nested fashion. This concept is implemented as the class `ALDWorkflow` which extends `ALDOperator`. The graphical programming editor `Grappa` is included in `Alida` to interactively design workflows in an intuitive fashion. The data processing pipeline is naturally modelled as a graph, where operators are represented by nodes, and the parameters of different operators are connected by edges to describe the flow of data. All data processing algorithms implemented as an `Alida` operator are right away available as operator nodes in `Grappa` and form the building blocks for workflows (see **Fig. 3** for an example). When connecting parameters of different nodes the validity is verified. For example, an input parameter may have at most one incoming edge, and the data types of parameters connected by an edge need to be compatible. Data propagated along an edge may be converted on user request if an appropriate converter is implemented. For example `Alida` includes functionality to convert an array to a collection. The set of converters may be extended in analogy to data I/O providers. In general the `operate()` method of a workflow object invokes all operators of the workflow in topological order and forwards output data between operators according to the data flow. In addition partial execution of the workflow is supported.

`Alida` also includes automatic process documentation of an analysis procedure which is supposed to contain all information necessary to recover the results from the same input data at a later point in time. Since each operator execution is realized invoking the generic `runOp()` method, the processing pipeline can be understood as a subgraph of the dynamic call graph of the analysis process. This call graph may also be interpreted as a hierarchical graph where each invocation of an operator is represented by a node. Besides the input data provided by the data flow between operators, in addition all control settings and also metadata like software versions are fully automatically retrieved during processing and represented. At any point in time the relevant portion of this processing graph may be retrieved and made explicit in terms of XML representations. This representation may be stored for archival purposes to, e.g., extract relevant information for publication. `Alida` also includes Chipory (see `Alida`'s homepage) to graphically display the processing graph and to inspect, e.g., parameter settings (see **Fig. 4** for an example).

Quality Control

The `Alida` library is actively developed since 2010 and has reached a mature state. The core has converged to a stable status and new features are integrated very diligently. The core functionality of `Alida` and particularly the components of the graphical user interfaces

are mainly tested manually, partially relying on test operators specifically designed to test a certain functionality. Feedback may be submitted via a bug tracking system and using Github's pull requests. In addition, since `Alida` forms the base of the Microscope Image Analysis Toolbox MiToBo (http://www.informatik.uni-halle.de/mitobo), its development is also significantly triggered and supported by feedback, bug reports and feature wishes from the users of MiToBo [5]. This significantly adds to the robustness and stability of the `Alida` library. The tests and the use of MiToBo subsuming `Alida` have been performed on different operating systems (64-bit Linux, Windows XP and 7, OS X) and with different Java versions.

(2) Availability

Project Homepage
http://www.informatik.uni-halle.de/alida.

Operating System
Alida runs on different versions of Linux, OS X, and Windows.

Programming Language
Java, version 1.8.

Additional system requirements
None.

Dependencies
The `Alida` distribution is shipped with all libraries required to make use of `Alida`'s complete functionality. For own developments based on `Alida` a Maven server[1] hosts the latest artifacts keeping track of dependencies automatically.

List of contributors
Birgit Möller
Stefan Posch

Software location
Archive
Name: Zenodo Research Archive
Title: Alida – Advanced Library for Integrated Development of Data Analysis Applications: v2.7
URL: https://zenodo.org/record/47586
Persistent identifier: https://doi.org/10.5281/zenodo.47586
Licence: GNU General Public License, Version 3
Publisher: Stefan Posch, Birgit Möller
Artifact Version: 2.7
Date published: 15/03/2016

Code repository
Name: Github
Title: alida
Identifier: https://github.com/alida-hub/alida
Licence: GNU General Public License, Version 3
Publisher: Birgit Möller, Stefan Posch
Date published: 14/06/2015

Maven repository server
Name: Apache Archiva Repository Server
Title: de.unihalle.informatik.Alida
Identifier: https://moon.informatik.uni-halle.de/archiva/
Licence: GNU General Public License, Version 3
Publisher: Birgit Möller, Stefan Posch
Date published: snapshots and releases are continuously published

Language
English

(3) Reuse Potential

The overall target of the `Alida` library is to provide a framework for developing modular, easy-to-use and particularly reusable data analysis software. Consequently, re-usability is the inherent key paradigm which coins the design and implementation of `Alida` in all respects. Moreover, this re-usability is not restricted to a specific research field, rather `Alida` is suitable for developments in all domains where data analysis coincides with `Alida`'s concept of operators performing data manipulations.

Besides this conceptual perspective `Alida` also aims to foster re-usability from a technical point-of-view. As mentioned above Maven is used to automatically resolve dependencies. Moreover on the website of `Alida` in the 'Downloads' section a template Maven project[2] can be found which is readily configured for immediate use in own projects. Finally, the open source strategy of `Alida` and the GPL licensing concept inherently guarantee a high degree of flexibility and adaptivity of `Alida` which renders it easy to even adjust the core functionality to new areas and fields of application if necessary.

Notes
[1] https://moon.informatik.uni-halle.de/archiva/#welcome.
[2] http://www2.informatik.uni-halle.de/agprbio/alida/downloads/alida-project-template-maven-1.2-src.zip.

Competing Interests
The authors have no competing interests to declare.

References
1. **Berthold, M R, Cebron, N, Dill, F, Gabriel, T R, Kötter, T, Meinl, T, Ohl, P, Thiel, K** and **Wiswedel, B** 2009 KNIME – The Konstanz Information Miner: version 2.0 and beyond. *ACM SIGKDD Explorations Newsletter,* 11(1): 26–31, DOI: https://doi.org/10.1145/1656274.1656280
2. **Curcin, V** and **Ghanem, M** 2008 Scientific workflow systems – can one size fit all? In *2008 Cairo International Biomedical Engineering Conference,* pages 1–9, Dec 2008. DOI: https://doi.org/10.1109/cibec.2008.4786077
3. **Kirchner, S, Posch, S** and **Möller, B** 2012 Graphical programming in Alida and ImageJ 2.0 with Grappa. In *Proc. of ImageJ User & Developer Conference,* pages 138–143, Mondorf-les-Bains, Luxembourg, October.
4. **Ludäscher, B, Altintas, I, Berkley, C, Higgins, D, Jaeger, E, Jones, M, Lee, E A, Tao, J** and **Zhao, Y** 2006 Scientific workflow management and the Kepler system. *Concurrency and Computation: Practice and Experience,* 18(10): 1039–1065,. DOI: https://doi.org/10.1002/cpe.994
5. **Möller, B, Glaß, M, Misiak, D** and **Posch, S** 2016 MiToBo – A Toolbox for Image Processing and Analysis. *Journal of Open Research Software,* 4(1): p. e17, DOI: https://doi.org/10.5334/jors.103
6. **Möller, B, Greß, O** and **Posch, S** 2011 Knowing What Happened – Automatic Documentation of Image Analysis Processes. In Crowley, J, Draper, B and Thonnat, M. (Eds.), *Proceedings of 8th International Conference on Computer Vision Systems,* volume 6962 of LNCS, pages 1–10, Sophia Antipolis, France, Springer.
7. **Möller, B** and **Posch, S** 2013 A Framework Unifying the Development of Image Analysis Algorithms and Associated User Interfaces. In *Proc. of 13th IAPR International Conference on Machine Vision Applications,* pages 447–450, Kyoto, Japan.
8. **Posch, S** and **Möller, B** 2012 Automatic Generation of Processing Histories using Alida. *In Proc. of ImageJ User & Developer Conference,* pages 218–221, Mondorf-les-Bains, Luxembourg, October.
9. **Posch, S** and **Möller, B** 2016 Design and Implementation of the Alida Framework to Ease the Development of Image Analysis Algorithms. *Pattern Recognition and Image Analysis,* 26(1): 181–189, DOI: https://doi.org/10.1134/S105466181601020X
10. **Taylor, I, Shields, M, Wang, I** and **Harrison, A** 2007 *The Triana Workflow Environment: Architecture and Applications,* pages 320–339. Springer, London.
11. **Wolstencroft, K, Haines, R** and **Fellows, D** et al 2013 The Taverna workflow suite: designing and executing workflows of web services on the desktop, web or in the cloud. *Nucleic Acids Research,*. DOI: https://doi.org/10.1093/nar/gkt328

SOFTWARE METAPAPER

MiToBo – A Toolbox for Image Processing and Analysis

Birgit Möller[1], Markus Glaß[2], Danny Misiak[2] and Stefan Posch[1]

[1] Institute of Computer Science, Faculty of the Natural Sciences III, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany
 birgit.moeller@informatik.uni-halle.de

[2] Institute of Molecular Medicine, Medical Faculty, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany

Corresponding author: Birgit Möller

MiToBo is a toolbox and Java library for solving basic as well as advanced image processing and analysis tasks. It features a rich collection of fundamental, intermediate and high-level image processing operators and algorithms as well as a couple of sophisticated tools for specific biological and biomedical applications. These tools include operators for elucidating cellular morphology and locomotion as well as operators for the characterization of certain intracellular particles and structures.

MiToBo builds upon and integrates into the widely-used image analysis software packages ImageJ and Fiji [11, 10], and all of its operators can easily be run in ImageJ and Fiji via a generic operator runner plugin. Alternatively MiToBo operators can directly be run from command line, and using its functionality as a library for developing own applications is also supported. Thanks to the Alida library [8] forming the base of MiToBo all operators share unified APIs fostering reusability, and graphical as well as command line user interfaces for operators are automatically generated. MiToBo is available from its website http://www.informatik.uni-halle.de/mitobo, on Github, via an Apache Archiva Maven repository server, and it can easily be activated in Fiji via its own update site.

(1) Overview
Introduction
The Microscope Image Analysis Toolbox MiToBo is an extendable collection of basic as well as more sophisticated image processing operators. Additionally, it offers several pipelines for analyzing specific microscopy based experimental setups [1, 2, 5, 6].

Although the focus is on biomedical image processing, it is also applicable to other image analysis tasks. Since it integrates into the popular image processing platforms ImageJ and Fiji, respectively, it automatically benefits from the capability of these programs to process a huge variety of input image types. MiToBo provides an elaborated infrastructure for development and rapid prototyping. Besides its usage in biomedical research, e.g. [4, 12, 13], it is also actively used for teaching. Its implementation focuses on easy and unified (re-)usability as well as modularity. MiToBo operators can be configured to be executed via a graphical user interface as well as from command line. Furthermore, the integrated graphical programming platform Grappa [3] can be used to create complex analysis pipelines by simply combining existing operators inside a workbench area.

Algorithms and Applications Basic image processing algorithms implemented in MiToBo comprise, for example, linear and non-linear filters, morphological operations, image arithmetics, connected component labeling or several automatic thresholding methods.

Among the more complex operators are segmentation algorithms like snakes or level set methods, which are highly configurable, e.g., by offering different energy functionals or initialization methods [7]. Furthermore, MiToBo contains operators for contrast enhancement, illumination correction or feature extraction as well as operators to segment cell nuclei or boundaries, for example. Some processing pipelines tailored for the analysis of specific biomedical experiments are also included in MiToBo and its number increases constantly. The ScratchAssayAnalyzer [2] is designed for quantifiying the development of the cell-free area in scratch/wound closure assays. With the ActinAnalyzer2D [6] the distribution of actin microfilament patterns inside cells can be quantified and compared. The NeuronAnalyzer2D [5] segments fluorescently labeled neurons, characterizes the cell morphology and extracts profiles from labeled molecules within the cell. The MTBCellCounter [1] is dedicated to semi-automatic

269

labeling and counting of spot-like structures and small cells. The migratory behavior of fluorescently labeled cells can be analyzed using the CellMigrationAnalyzer. The MiToBo Cell Image Analyzer (MiCA) [9] is conceptualized for segmenting cell boundaries and sub-cellular particles in fluorescence microscopy images. In total, the current release (MiToBo 1.7) contains more than 150 distinct operators.

General features MiToBo's capacities to supply easy usability, generic user interfaces and graphical programming for image processing applications are mainly taken from Alida (see Section Implementation). Building on Alida's concept, each image data analysis action is realized in terms of an operator where each operator implements a common interface definition. As furthermore their input and output parameters are accessible in a standardized manner each operator can be invoked in a uniform way. This is exploited to automatically generate command line and graphical user interfaces for each operator implemented in MiToBo without any additional burden on the programmer (see **Figs. 1** and **2** for examples). The graphical user interface includes the option for batch mode operation which, e.g., allows to apply an operation to all files of a given directory or for a range of parameter values.

While developing image analysis pipelines and also with regard to the reproducibility of image analysis results it is usually of interest to thoroughly document parameter settings and analysis steps. To simplify such logging procedures MiToBo offers an automatic process documentation where the invocation of an operator not only produces image analysis results, but at the same time registers within the framework along with all input and output objects as well as parameters settings. These data acquired during an image analysis process and the order of operator calls form a directed graph containing all relevant information for later reconstruction or verification of the analysis procedure. This processing graph is made explicit in terms of an XML representation which can be visually explored with appropriate graphical frontends like Chipory (see the Alida home page) or might be stored in data bases for archival purposes. **Fig. 3** gives an example processing graph for an image processing pipeline.

In most cases several image analysis operations, e.g., filters, thresholding, or segmentation algorithms, have to be applied in a sequential, parallel or nested fashion to achieve desired image processing results. The graphical programming editor Grappa included in MiToBo facilitates the design of such processing pipelines or workflows
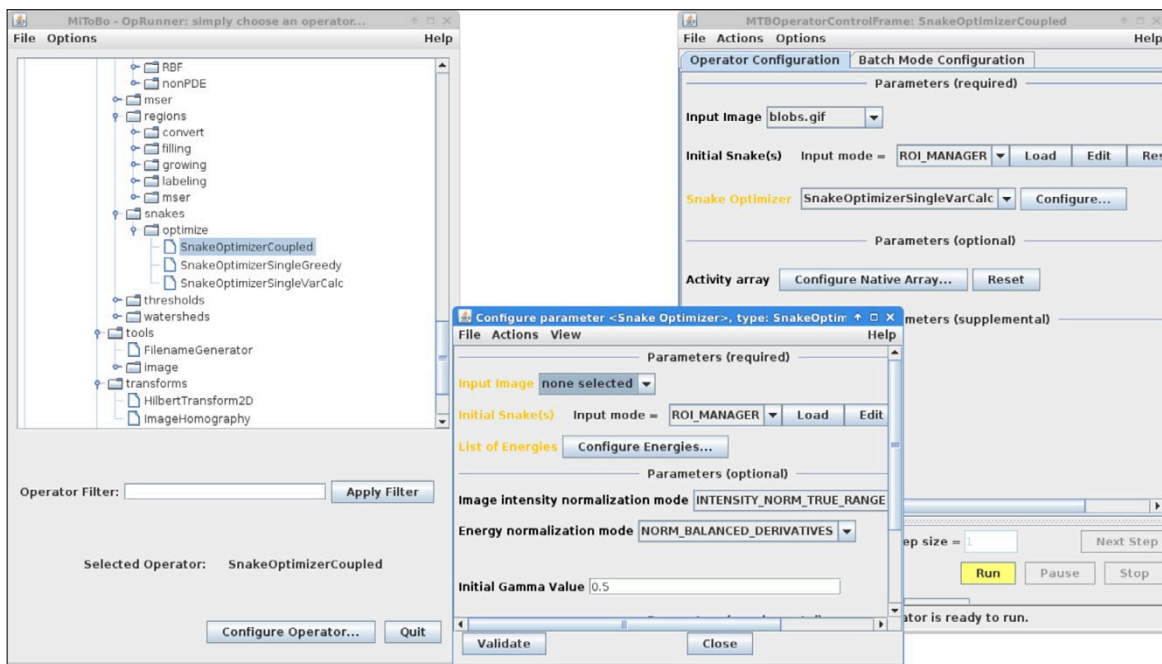
Figure 1: The main window of Alida's operator runner which allows to select any of the available MiToBo image processing operators (top left), and two automatically generated control and configuration windows for operators (right and bottom).

```
java de.unihalle.informatik.Alida.tools.ALDOpRunner GaussFilter\
    inputImg=cell.tif sigmaX=1.4 resultImg=cell-g3.tif
```

Figure 2: Example call of an image analysis operator from command line. The operator `GaussFilter` is executed to apply a Gaussian filter with standard deviation of 1.4 to the input image `cell.tif`, generating the filtered result image `cell-g3.tif`.
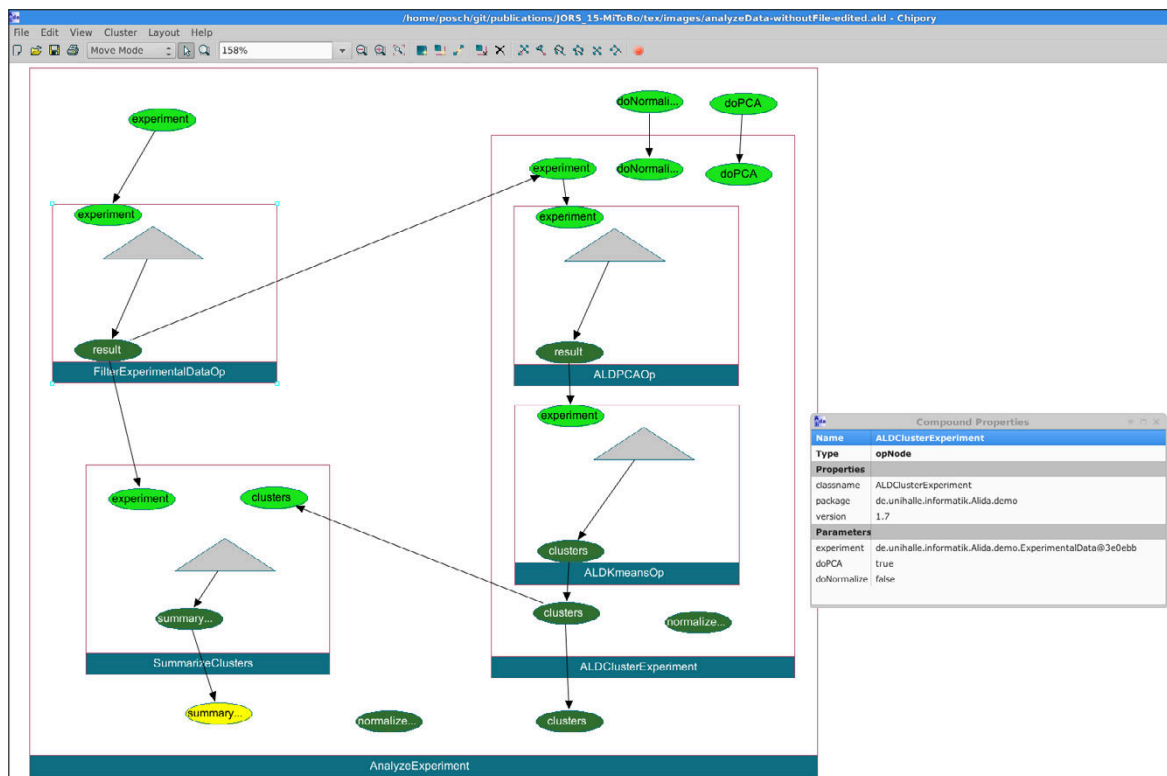
Figure 3: Example processing graph representing the history of operations for producing the image data object shown as yellow ellipse. Each operator invocation is represented by a blue rectangle. Light and dark green ellipses are input and output ports of operators, gray triangles depict data ports representing newly generated data. The rightmost window displays the parameters and software version for the operator `ALDClusterExperiment` during execution.

in an intuitive way. All image processing algorithms implemented as an Alida operator are right away available as operator nodes in Grappa and form the building blocks for workflows. Their input and output parameters may be connected interactively by user-friendly mouse actions where compatibility of data types is enforced and automatic data type conversion is applied if appropriate converts are available (see **Fig. 4** for an example).

Implementation

The MiToBo library, implemented in Java, uses and strongly builds on Alida, an *Advanced Library for Integrated Development of Data Analysis Applications*, which defines and implements a concept for designing libraries and toolkits in data analysis (see [8] and http://www.informatik. uni-halle.de/alida).

On the other hand MiToBo builds upon ImageJ, a Java toolkit for image processing and analysis [11] widely used, but not confined to biomedical applications. ImageJ provides the user with flexible means to interactively process, modify, and display images, with a large variety of basic built-in image processing operations, and also with a huge collection of optional plugins downloadable from the web. While MiToBo is compatible with ImageJ it significantly improves data access and exchange between more complex processing modules exploiting Alida's operator

concept. Furthermore, MiToBo extends ImageJ's image data types and provides more sophisticated data structures for, e.g., regions and contours. MiToBo's facilities are available within ImageJ either via the MiToBo OpRunner plugin or as dedicated ImageJ plugins. The former allows to configure and run each data analysis procedure implemented as an Alida operator (c.f. Subsection General features and **Fig. 1**).

Quality Control

The MiToBo core library and the plugin collection currently consist of about 500 Java class files. Approximately 10% of them regularly undergo automatic tests based on JUnit test classes, while the bigger part of the testing is done by manual verification. MiToBo is developed on different operating systems (e.g., Ubuntu Linux 12.04 and 14.04, Microsoft Windows 7 and 8). While the officially supported Java version is still 1.6 (for compatibility reasons with ImageJ/Fiji), developers and users also run MiToBo frequently with additional Java versions (OpenJDK and Oracle Java 1.7/1.8) to ensure a maximum degree of portability.

MiToBo's operators and applications are usually developed in close collaboration with users mainly from Computer Science, Biology and Medicine. They particularly provide feedback on software usability and the plausibility of computed results. Thanks to the update
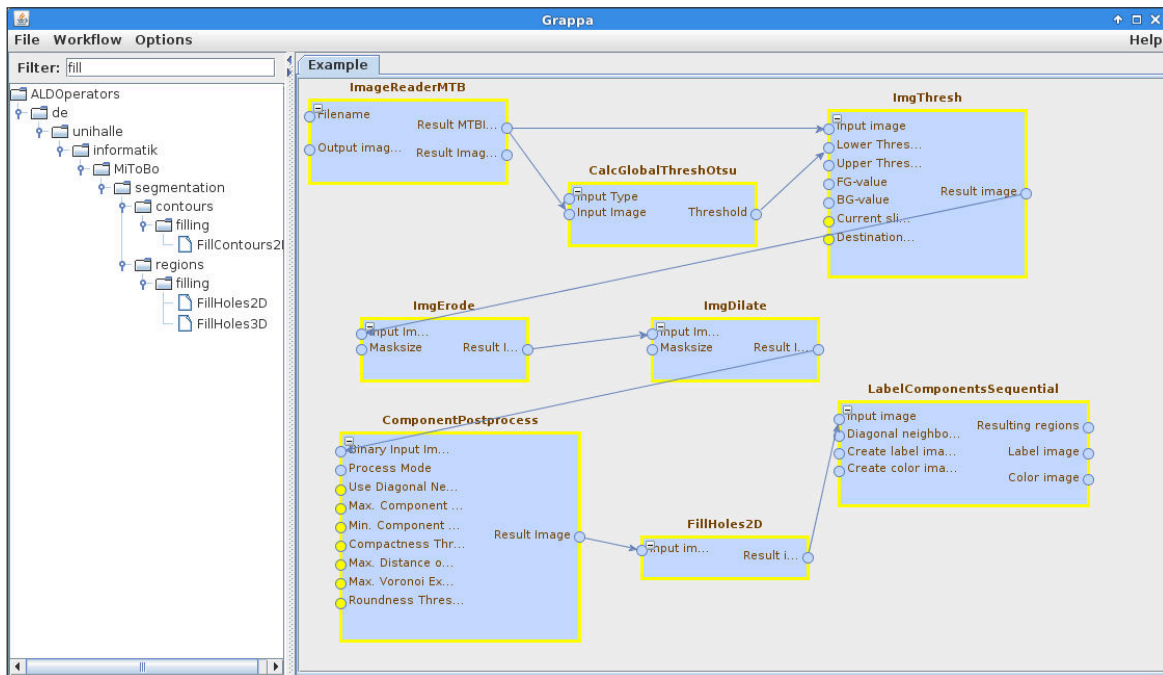
Figure 4: Screen shot of Alida's graphical programming editor Grappa. The operator chooser panel on the left allows to choose from all available operators. The main panel shows a workflow composed of various opertors connected by edges defining the data flow. Additional parameters may be set via the same automatically generated GUIs available within the operator runner.

mechanisms integrated in Fiji, such feedback is promptly available resulting in rather short prototyping, software development and bugfix cycles.

(2) Availability

Project Homepage
http://www.informatik.uni-halle.de/mitobo

Operating System
MiToBo and ImageJ/Fiji run on different versions of Unix, Linux, Mac OS, and Windows.

Programming Language
Java, version 1.6

Additional system requirements
None.

Dependencies
MiToBo builds upon ImageJ/Fiji and its operators can best be run in ImageJ or Fiji, but usage as a library and on command line is also possible. For usage as a library or with ImageJ all required dependencies are shipped as part of the MiToBo distribution. In Fiji MiToBo is available via an update site which automatically resolves the dependencies (http://sites.imagej.net/MiToBo/). To activate MiToBo in Fiji the entry of MiToBo's update site in Fiji's built-in list of available sites just needs to be checked and the updater run (for further details refer to the Fiji documentation at http://fiji.sc/Update_Sites).

List of Contributors
Markus Glaß
Oliver Greß
Danny Misiak
Birgit Möller
Stefan Posch

Software Locations:
· Archive:
   Name: Zenodo Research Archive
   Title: MiToBo – A Microscope Image Analysis Toolbox: v1.7
   URL: https://zenodo.org/record/31364
   Persistent identifier: http://dx.doi.org/10.5281/
          zenodo.31364
   Licence: GNU General Public License, Version 3
   Publisher: Birgit Möller, Stefan Posch, Markus Glaß,
          Danny Misiak
   Artifact Version: 1.7
   Date published: 09/24/2015

· Code repository:
   Name: Github
   Titles: mitobo / mitobo-plugins
   Identifiers: https://github.com/mitobo-hub/mitobo
          https://github.com/mitobo-hub/
          mitobo-plugins
   Licence: GNU General Public License, Version 3
   Publisher: Birgit Möller, Stefan Posch, Markus Glaß,
          Danny Misiak
   Date published: 06/28/2015

· Maven repository server:
Name: Apache Archiva Repository Server
Title: de.unihalle.informatik.MiToBo
Identifier: https://moon.informatik.uni-halle.de/
archiva/
Licence: GNU General Public License, Version 3
Publisher: Birgit Möller, Stefan Posch, Markus Glaß,
Danny Misiak
Date published: snapshots and releases are continu-
ously published

Language
English.

(3) Reuse Potential

One of the explicit goals of MiToBo is reusability. As pointed out in Section 1 MiToBo includes implementa-tions of algorithms and operators with varying complex-ity. Most algorithms and concepts are independent of concrete applications. Expert as well as non-expert users can easily apply them to image processing and analy-sis problems in a large variety of domains, not neces-sarily restricted to microscopy images or biomedical scenarios. In addition, also the more specialized opera-tors built on top of these basic ones and being integral part of MiToBo demonstrate the reuse potential. They cover applications from biology, biomedicine and ecol-ogy. Complementary the set of different user interfaces which Alida and MiToBo offer, i.e., graphical and com-mand line interfaces, support efficient reuse of MiToBo not only with regard to concrete image processing problems, but also in different usage scenarios ranging from rapid-prototyping and development with tied user-developer feedback loops to scripting and high-through-put processing.

Finally, reusability of the MiToBo toolbox and its opera-tors is not restricted to the user side. Also for developers reusing operators and implemented algorithms is easy. Based on Alida's concepts for generic implementation of functionality and user interfaces [8] all operators provide unified interfaces for reusing them on the code level in a modular fashion. Likewise new operators and also data types can easily be added to MiToBo's core and take imme-diate advantage of the rich set of built-in features in Alida and MiToBo. To ease the first steps with MiToBo for own developments a pre-configured Maven project contain-ing a demo operator can be found on MiToBo's project website.

Competing Interests
The authors declare that they have no competing interests.

References

1. **Franke, L, Storbeck, B, Erickson, J L, Rödel, D, Schröter, D, Möller, B** and **Schattat, M H** 2015 The 'MTB Cell Counter' a versatile tool for the semi-automated quantification of sub-cellular phenotypes in fluorescence microscopy images. A case study on plastids, nuclei and peroxisomes. *Journal of Endocyto-biosis and Cell Research*, 26: 31–42.
2. **Glaß, M, Möller, B, Zirkel, A, Wächter, K, Hüttelmaier, S** and **Posch, S** 2012 (September) Cell migration analysis: segmenting scratch assay images with level sets and support vector machines. *Pattern Recognition*, 45(9): 3154–3165. DOI: http://dx.doi.org/10.1016/j.patcog.2012.03.001
3. **Kirchner, S, Posch, S** and **Möller, B** 2012 (October) Graphical programming in Alida and ImageJ 2.0 with Grappa. In *Proc. of ImageJ User & Devel-oper Conference.* Mondorf-les-Bains, Luxembourg, pp. 138–143.
4. **Komaravolu, R K, Adam, C, Moonen, J-R A J, Harmsen, M C, Goebeler, M** and **Schmidt, M** 2015 Erk5 inhibits endothelial migration via KLF2-depend-ent down-regulation of PAK1. *Cardiovascular Research*, 105(1): 86–95. DOI: http://dx.doi.org/10.1093/cvr/cvu236
5. **Misiak, D, Posch, S, Lederer, M, Reinke, C, Hüttelmaier, S** and **Möller, B** 2014 (March) Extraction of protein profiles from primary neurons using active contour models and wavelets. *Journal of Neuroscience Methods*, 225: 1–12. DOI: http://dx.doi.org/10.1016/j.jneumeth.2013.12.009
6. **Möller, B, Piltz, E** and **Bley, N** 2014 (August) Quan-tification of actin structures using unsupervised pat-tern analysis techniques. In *Proceedings of Int. Conf. on Pattern Recognition (ICPR).* Stockholm, Sweden, IEEE, pp. 3251–3256. DOI: http://dx.doi.org/10.1109/icpr.2014.560
7. **Möller, B** and **Posch, S** 2012 (May) Comparing ac-tive contours for the segmentation of biomedical images. In *Proc. of IEEE International Symposium on Biomedical Imaging: From Nano to Macro(ISBI), IEEE Catalog No.: CFP12BIS-CDR ISBN: 978-1-4577-1856-4*, pp. 736–739, Barcelona, Spain. DOI: http://dx.doi.org/10.1109/isbi.2012.6235653
8. **Möller, B** and **Posch, S** 2013 (May) A framework uni-fying the development of image analysis algorithms and associated user interfaces. In *Proc. of 13th IAPR In-ternational Conference on Machine Vision Applications (MVA).* Kyoto, Japan, pp. 447–450.
9. **Möller, B, Stöhr, N, Hüttelmaier, S** and **Posch, S** 2010 (August) Cascaded segmentation of grained cell tissue with active contour models. In *Proceedings of International Conference on Pattern Recognition.* Istan-bul, Turkey, IEEE, pp. 1481–1484. DOI: http://dx.doi.org/10.1109/icpr.2010.366
10. **Schindelin, J, Arganda-Carreras, I, Frise, E, Kaynig, V, Longair, M, Pietzsch, T, Preibisch, S, Rueden, C, Saalfeld, S, Schmid, B,** et al. 2012 Fiji: an open-source platform for biological-image analysis. *Nature methods*, 9(7): 676–682. DOI: http://dx.doi.org/10.1038/nmeth.2019
11. **Schneider, C A, Rasband, W S** and **Eliceiri, K W** 2012 NIH Image to ImageJ: 25 years of image analy-sis. *Nature Methods*, 9(7): 671–675. DOI: http://dx.doi.org/10.1038/nmeth.2089

12. **Wolf, A, Rietscher, K, Glaß, M, Hüttelmaier, S, Schutkowski, M, Ihling, C, Sinz, A, Wingenfeld, A, Mun, A** and **Hatzfeld, M** 2013 Insulin signaling via Akt2 switches plakophilin 1 function from stabilizing cell adhesion to promoting cell proliferation. *Journal of Cell Science*, 126(Pt 8): 1832–1844. DOI: http://dx.doi.org/10.1242/jcs.118992

13. **Zirkel, A, Lederer, M, Stöhr, N, Pazaitis, N** and **Hüttelmaier, S** 2013 IGF2BP1 promotes mesenchymal cell properties and migration of tumor-derived cells by enhancing the expression of LEF1 and SNAI2 (SLUG). *Nucleic Acids Research*, 41(13): 6618–6636. DOI: http://dx.doi.org/10.1093/nar/gkt410

MiCA - Easy Cell Image Analysis with Normalized Snakes

Birgit Möller and Stefan Posch

Institute of Computer Science, Martin Luther University Halle-Wittenberg, Germany

Abstract— **Quantitative analysis of microscopy cell images requires accurate detection of cell boundaries, nuclei and sub-cellular structures. Accordingly, tools for integrated cell image analysis are required which not only provide a variety of different segmentation and detection algorithms, but at the same time support easy usage also by life scientists. In this paper we present our integrated cell analysis tool MiCA offering different segmentation techniques, e.g., based on wavelets or snakes. Iterative optimization of snake energies depends on a variety of parameters that require thorough adjustment for optimal results. To facilitate easy use of these techniques MiCA provides a new energy normalization scheme for snakes allowing for intuitive interpretation of energy parameters and, thus, simplified cell image analysis. The high quality of the result data of MiCA is proven on two sample data sets by qualitative assessments and ground-truth comparisons.**

I. INTRODUCTION

One fundamental building block of nowadays research in cell biology are fluorescent microscope images, allowing for qualitative – but more importantly – also for quantitative characterization of sub-cellular processes. Nuclei, cytoplasm or various structures within cells are fluorescently labeled, and quantitative evaluation of their size, number, spacial distribution, shape and further properties allow for biological and biomedical investigations.

Most of the time quantitative evaluation of microscope image data is accomplished manually, which induces a certain degree of subjectivity into the results. Furthermore this impedes high-throughput experiments which are necessary for statistically meaningful analysis of larger populations. Consequently, to increase flexibility and impact of biomedical research automated analysis procedures are indispensable. However, these are not only required to yield accurate segmentation of all structures under consideration, but also need to allow for easy handling by life scientists.

In this paper we present our integrated cell image analysis tool 'MiCA', the MiToBo[1] Cell Image Analyzer, meeting these requirements. MiCA implements various techniques to segment cell boundaries and cell nuclei, as well as cell structures like stress granules (SGs), processing bodies (PBs), or focal contacts. For the segmentation of nuclei and structures algorithms based on morphological operators and wavelets yield satisfactory results. However, detection of cell boundaries requires more elaborate approaches based on explicit active contours. These snakes are based on iterative numerical optimization and depend on various parameters

which have to be carefully chosen for optimal results. But, parameter adjustment typically requires deep knowledge of the underlying energy models and optimization techniques, hampering easy use of this approach for non-experts.

To simplify this search and tuning process for identifying well-suited parameters we propose a new normalization scheme for active contour energies implemented within MiCA. The normalization scheme introduces a more intuitive way of parameter interpretation and also aims at numerical stability in iterative optimization. In detail, balancing between different components of the energy becomes much easier, and snake behavior gets more predictable. Hence, MiCA offers new ways for easy analysis of cell images and is particularly well-suited for life scientists.

The benefits of MiCA and its integrated normalization scheme are demonstrated on two sets of experimental data for which in part ground-truth is available. Data set D1 consists of 8 epifluorescence multi-channel microscopy images, each containing DAPI labeled nuclei and 2 additional channels with fluorescently marked SGs and PBs (see [1], [2] for details). Data set D2 subsumes 8 images of a cell population where focal contacts are labeled. Focal contacts are macromolecular assemblies establishing links between actine proteins in a cell's cytosceleton and its matrix for transmission of regulatory signals. To visualize focals the protein vinculin is fluorescently labeled, which is located at focal adhesion sites (see Fig. 6). We prove the high quality of the results provided by MiCA, i.e. cell areas and assignments of sub-cellular structures to individual cells, by qualitative assessments, and by quantitative evaluations if ground-truth expert labelings of the data are available.

The remainder of the paper is organized as follows. After reviewing related work we discuss active contours and the new normalization scheme in Sec. III. As the biological analysis of microscope images from cells also subsumes the segmentation of nuclei and different types of sub-cellular structures, related techniques based on morphological operations and wavelets are briefly summarized in Sec. IV. Sec. V presents the freely available MiCA ImageJ plugin, Sec. VI and Sec. VII discuss results and give a conclusion.

II. RELATED WORK

Automatic segmentation of cells in fluorescence microscopy images and the detection of fluorescently labeled sub-cellular structures are examples of general segmentation problems in computer vision. As typically the case this general techniques have to be customized to the special properties and needs of the application considered.

Cells imaged as "punctuated" patterns in 2D or 3D are segmented in [3] proposing a multi-resolution and multi-scale approach. It basically employs a voting scheme combining intensity-based soft thresholding to discriminate between background and cells with a spatial low-pass filter to separate individual cells from each other. Segmentation and tracking of cells are treated in [4] employing level sets and the watershed transform. Initially, cells are segmented with a region-based level set procedure, where the classical Chan-Vese model [5] is modified by replacing the external energy by a log-Gaussian likelihood with unknown variances. The watershed transform is used to cope with lumped cells. For subsequent tracking a multi-phase level set approach is used incorporating a coupling term. A similar problem is treated in [6] where a particle filter is employed for tracking. Snakes are used for post-processing using the classical snake energy [7] where the external energy is derived from the gradient vector flow field. In [8] an integrated approach for cell segmentation and particle detection is presented. In contrast to the work presented in this article cell segmentation relies on cell specific labeling with HCS Cellmask. Initial contours are found by k-means clustering and refined using snakes combining region fitting and gradient-based data terms with a shape prior to favor elongated cell contours. Sub-cellular particles are detected adopting the wavelet-based technique [9], being also the basis for our cell analyzer. In general, the automatic detection of such spot-like particles often relies on global and local thresholding techniques like Otsu's global method or the local Niblack operator ([10], [11]). Further techniques include sampling from an image intensity density estimated via h-dome transform and subsequent sample clustering [12].

In [13] a method to segment nuclei is proposed extending the Chan-Vese model. The spatial intensity distribution of each nucleus is modeled by a Gaussian distribution, where the parameters of each cell are estimated from the current segmentation in each iteration of a multi-phase level set approach. The results derived for a sequence of real 3D data rely on an initial segmentation by a biologist expert underscoring the sensitivity of this spatial energy functional.

Solving microscope image analysis tasks not only includes algorithm development, but also the need to make them available to life scientists in a comfortable way. Applications like CellProfiler[2] provide integrated frameworks with a large collection of different algorithms, but may also require a significant amount of time for users to learn how to exploit the subset of features relevant to their problems. Contrary, ImageJ[3] provides a platform easily adaptable to user-specific needs by a vast amount of freely available plugins. It is widely used in biomedical research, but plugins are often quite specialized and finding a suitable set of plugins to solve a concrete problem might be challenging. MiCA chooses a way in between by restricting functionality to cell image analysis, but by this preserves high usability.

[2]http://www.cellprofiler.org/
[3]http://rsbweb.nih.gov/ij/

III. Normalized Active Contours

Segmenting cell boundaries requires elaborate techniques, particularly as cellular membranes are difficult to stain directly, and also as the amount of antibodies and fluorochromes which can be applied reliably in one experiment is limited. Hence, we adopt one of the channels used to label other compounds of interest for cell segmentation, but, as a result need to cope with fuzzy cell boundaries. The segmentation is accomplished applying snakes as introduced in [1]. Here we enhance the approach towards more intuitive parameter tuning with a new snake normalization scheme.

A. Snake Basics

We adopt explicit active contours for the segmentation of cell boundaries. Each cell contour is modeled as a curve $c(s)$ in 2D which is discretized by a set $\{\vec{p}_i | i = 1, \ldots, L\}$ of 2D points. The curve is iteratively evolved given an underlying energy functional and is, hence, commonly denoted as *snake*. The energy specifies properties of the desired object contour, e.g., intensity profile, gradient strength, or a certain shape, and is to be minimized during segmentation.

In its most general form the continuous energy functional $E[c(s); \vec{\Theta}], s \in [0, 1]$, for an explicit active contour $c(s)$ can be formulated as a set of K different energies E_k, each integrating along the contour. All energies sum up to the complete energy functional:

$$E[c(s); \vec{\Theta}] = \sum_{k=1}^{K} \int_{s=0}^{1} E_k[c(s); \vec{\theta}_k] ds . \tag{1}$$

Each single energy E_k depends on a set of individual parameters $\vec{\theta}_k \subseteq \vec{\Theta}$. Note that most region-based energies like Chan-Vese [5] or binary flows [14] defined as integrals over regions can be transformed into contour integrals and subsumed by (1) as well.

The discretized form of this functional can be iteratively optimized applying gradient descent techniques. In detail, for each energy the Gateaux derivative is calculated, and their sum is considered in the iterative gradient optimization scheme which for a single discretized snake point $\vec{p}_i$ results in (see [7] for further details)

$$\vec{p}_i^{(t)} = \vec{p}_i^{(t-1)} - \gamma \cdot \left[\sum_{k=1}^{K} \frac{\partial}{\partial \vec{p}_i^{(t)}} \sum_{j=1}^{L} E_k[\vec{p}_j^{(t)}; \vec{\theta}_k] \right] . \tag{2}$$

$\gamma \in \mathbb{R}$ denotes the stepsize of the gradient descent, and t the iterations. Obviously, the motion of a snake point $\vec{p}_i$ is directly influenced by γ, however, also modifications of energy parameters $\vec{\theta}_k$ induce changes is snake dynamics. Hence, γ and Θ are deeply linked to each other, rendering proper parameter adjustment as one of the main problems when adopting active contours for object segmentation.

B. Normalization and Parameter Tuning

Different energies and their derivatives typically have different ranges of values where the difference is often

several orders of magnitude. Hence, finding suitable energy and optimization parameters is cumbersome and requires thorough experience. Also numerical instabilities frequently occur, often resulting in uncontrollable snake behavior or a lack of convergence.

To ease the search for proper parameters we propose to normalize all derivatives to a common interval of possible values. This results in a twofold benefit. On the one hand the numerical stability of the optimization is improved. On the other hand it becomes significantly simpler to determine appropriate parameters $\vec{\theta}_k$ as these can be interpreted more intuitively. In essence it can be attributed to the property of the proposed energy normalization to shift the focus of parameter tuning from the need to consider absolute energy values to a view of weighting the various energies E_k. This relieves the user from the need to have detailed knowledge about the characteristics of all individual energies.

In our implementation all energy derivative components are linearly normalized to an admissible range of $R_E = [-1, 1]$. Given an original interval $v \in [v_{min}, v_{max}]$ of possible values the linear normalization T is given as:

$$T(v) = \frac{(v - max(|v_{min}|, |v_{max}|))}{2 \cdot max(|v_{min}|, |v_{max}|)} \cdot 2 + (-1) . \quad (3)$$

Note that the normalization is defined to map the interval $[-max(|v_{min}|, |v_{max}|), max(|v_{min}|, |v_{max}|)]$ to R_E to ensure that a value of zero is mapped to zero. If we assume a stepsize of $\gamma = 1$ and consider one iteration step, this normalization confines the influence of each energy E_k to move a snake point within a square of radius one pixel. Thus, the normalization connects the stepsize specified by γ to pixel units allowing easy adaptation of snake speed.

At the same time, given such normalized energy derivatives, balancing of different energy terms against each other also becomes very intuitive. In particular, adding weights for all energy terms in (1), we have

$$E[c(s); \Theta] = \sum_{k=1}^{K} \omega_k \int_{s=0}^{1} E_k[c(s); \vec{\theta}_k] ds , \quad (4)$$

where $\omega_k \in \mathbb{R}_0^+, \sum_{k=1}^{K} \omega_k = 1$. This allows the adaptation of the parameter set Θ to be conceptually divided into two abstraction levels. As each energy E_k typically models a certain aspect of the desired contour, such as fidelity to data or global shape, its parameters $\vec{\theta}_k$ are adjusted according to the given application focussing on this aspect. Then all energy terms and, thus, all aspects modeled are balanced using the weights ω_k at the second level of abstraction. Examples illustrating the approach are given in the next subsection for the segmentation of cell boundaries.

C. Energies for Cell Segmentation

For the segmentation of cells in fluorescent microscope images we apply the following energy functional, defined over a set of N different snakes $c_n(s)$ according to the number of cells present in the image (more details of the non-normalized approach are described in [1]). For each cell a Chan-Vese fitting term E_{cv} and a regularizer E_r are used with an additional coupling term E_{cp} resulting in a total of $2N + 1$ individual terms:

$$E[c_1(s), \ldots, c_N(s)] = \omega_{cp} E_{cp}(c_1(s), \ldots, c_N(s))$$
$$+ \sum_{n=1}^{N} (\omega_{cv} E_{cv}[c_n(s)] + \omega_r E_r[c_n(s)])$$

$$E_{cv}[c_n(s)] = \lambda_{in} \int_{R_{in}(c_n)} \left(I(x,y) - \mu_n^{in}\right)^2 d\Omega$$
$$+ \lambda_{bg} \int_{\Omega \backslash (\bigcup_{n=1}^{N} R_{in}(c_n(s)))} \left(I(x,y) - \mu^{bg}\right)^2 d\Omega$$

$$E_r[c_n(s)] = \frac{1}{2} \int_0^1 \beta \cdot \| c_n(s)'' \|^2 ds$$

$$E_{cp}[c_1(s), \quad \ldots \quad , c_N(s)] = \rho \sum_{m=1}^{N} \sum_{n=m+1}^{N} \int_{O_{m,n}} 1 d\Omega$$

$O_{m,n}$ denotes the overlap of the interiors of snakes c_m and c_n. The normalization is based on the individual ranges of each energy's derivative. For the regularizer E_r, possible values lie in the range of $[-4\beta, 6\beta]$ (cf. [7]). The Chan-Vese energy term E_{cv} derives as follows,

$$\frac{\partial}{\partial \vec{p}_i} \sum_{j=1}^{L} E_k[\vec{p}_j; \vec{\theta}_k] =$$
$$\left(\lambda_{in} \cdot [I(\vec{p}_i) - \mu_{in}]^2 - \lambda_{out} \cdot [I(\vec{p}_i) - \mu_{out}]^2\right) \cdot \vec{n}_c$$

where $\vec{n}_c$ is the local normal vector of $c(s)$ at position $\vec{p}_i$. Here the range of possible values is given by $[-\lambda_{out}\Delta I^2, \lambda_{in}\Delta I^2]$, where ΔI is the dynamic intensity range of the image. Finally, the derivative values of the snake overlap penalizer term E_{cp} lie between 0 and ρN.

D. Algorithm outline

Segmenting cells with active contours requires a proper initialization. As each cell usually contains one nucleus we use results from prior nucleus detection (cf. Subsec. IV). The contour of each nucleus is expanded to the neighboring cell cytoplasm area to ensure that initial contours are not located in the significantly darker nucleus regions (cf. Fig. 6, left). Subsequently iterative snake optimization is run until convergence, i.e. until a maximum number of iterations is reached or the change in area size enclosed by the snake falls below a threshold between two iterations.

IV. NUCLEUS AND STRUCTURE DETECTION

In addition to the normalized snake approach for cell boundary detection MiCA also provides techniques for segmenting nuclei and sub-cellular structures.

a) Nucleus Detection. The detection of cell nuclei is typically of twofold use. First nuclei are used to derive initial snake contours for the cells to be used in subsequent snake segmentation. Secondly in some applications it is necessary to distinguish between sub-cellular particles within the cytosol and within the nucleus for biological evaluation.

For the segmentation of nuclei we use DAPI staining. A global Otsu thresholding is applied with subsequent post-processing. This comprises morphological opening with a square structuring element of size 5 pixels, removal of very small components, which are sized smaller than 500 pixels, and removal of holes enclosed in nuclei regions.

For most cells the nucleus is detected correctly, however, sometimes nuclei of neighboring cells are merged. To separate these we apply a simple yet efficient procedure adapted from [15], briefly outlined below. As nuclei show an approximately elliptical shape distances between the centroid and the border vary smoothly. If, however, two or more nuclei are merged into one component a distinct increase of these distances will be noticed. For nucleus separation the distances are considered as a cyclic 1D signal and subjected to a top hat transform with a structuring element of size 0.28π. Subsequently thresholding with $\theta_{\text{dist}} = 20$ is performed. If a distance is above the threshold this indicates a merged neighboring nucleus in the corresponding direction. To eliminate noise the cyclic binary vectors representing these hypotheses are morpholocially opened and dilated with masks of size 10 and 5, respectively, yielding the final directions of contact to a neighboring nucleus. To separate these nuclei the border is linearly interpolated and the procedure applied recursively to handle cases of three or more merged nuclei.

b) Particle Detection. To detect spot-like sub-cellular structures of varying size we have extended the wavelet-based approach [9] for scale-adaptivity in [16] which is shortly summarized in the following. The image $I_0(x,y)$ is successively smoothed yielding images $I_1(x,y), s \in \{1, \ldots S\}$. Wavelet coefficients $W_s(x,y)$ are derived as

$$W_s(x,y) = I_s(x,y) - I_{s-1}(x,y), \qquad (5)$$

and the amplitude-scale-invariant Bayes estimator is applied for denoising yielding coefficients $\tilde{W}_s(x,y)$. Due to the wavelet transform applied, particles are represented by wavelet coefficients in adjacent scales, thus, adjacent scales are combined to a *correlation image*

$$c_{[a,b]}(x,y) = \prod_{s=a}^{b} \tilde{W}_s(x,y). \qquad (6)$$

This correlation image is globally thresholded and the resulting connected components yield the final particles detected. The interval of scales $[a,b]$ used to correlate the wavelet coefficients defines the scales at which the particles of interest are represented and is chosen fixed in [9]. If all particles share the same characteristics, one interval is appropriate, otherwise, however, one single interval either includes irrelevant scales or excludes important ones.

To overcome these issues we apply the wavelet-based detection to a set of – usually overlapping – intervals $[a_n, b_n]$, to accommodate for different scales of particles.

In general, this results in overlapping and, thus, competing particle hypotheses.

These are represented as nodes in hypothesis trees with edges connecting competing hypotheses (cf. Fig. 1 for a simple example for two intervals). The root of each detection tree refers to the coarsest particle hypothesis available at a given location. To select the correct hypothesis we
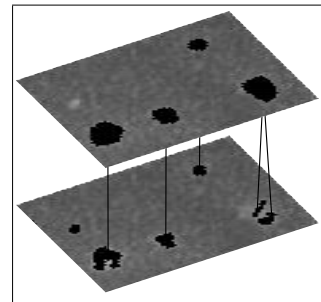
Fig. 1. Particle hypotheses for two adjacent intervals (top: coarse, bottom: fine) and resulting hypothesis trees.

employ the concept of meaningful events (see [17]). For each particle hypothesis we compute a meaningfulness which can be understood as the p-value of the image event under the null hypothesis H_0 of being caused by chance. These p-values are used to prune the decision trees where we normalize the p-values to account for different size of support of different particle hypotheses.

V. MiCA - INTEGRATED CELL ANALYSIS

Our cell analysis tool MiCA is freely available as part of MiToBo, a Java library of data types, algorithms and ImageJ plugins for advanced image analysis[4]. The MiCA plugin features a graphical user interface (Fig. 2) that allows to configure the different algorithms to be applied to either a given image or all images of a directory in batch mode. In detail, morphology-based nucleus detection including separation of conglomerates, wavelet- as well as morphology-based detection of structures and snake-based segmentation of cell boundaries are currently included.

For cell segmentation different kinds of energies are provided which can freely be combined to yield the final energy functional for optimization (cf. Sec. III). As outlined above, individual weights can be assigned to each of the selected energies introducing large flexibility for processing various types of cell images. Once the different segmentation results are available MiCA provides statistics like number and size of structures per image or per cell, respectively. Besides, masks of detected entities overlayed with extracted cell boundaries can be requested and saved for visual inspection of the segmentation results. All figures extracted are summarized in a results table which can be saved to disc in TSV text format for further evaluation.

For user convenience MiCA supports the full functionality of Alida[5] for process documentation. In particular, each result image stored on disc is accompanied by a history file which summarizes the input images used and the processing steps applied including all parameter settings and software versions in terms of an XML graph. By this complete and automatic process documentation and easy reconstruction of results in the long run are guaranteed (see [18] for details).
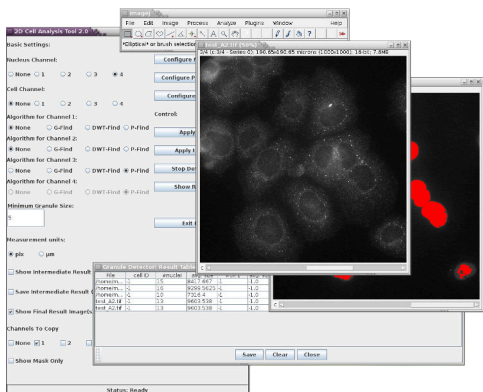
278

Fig. 2. Screenshot of the MiCA plugin GUI.

VI. Experimental Results

In this section we show results for both data sets described in Sec I. For quantitative evaluation of data set D1 we focus on cell segmentation to demonstrate the effectiveness of normalization, and on the biological interpretation.

In the 8 images 146 cells are projected, with 19 missing the nucleus as it is outside the field of view. Thresholding yields 122 components and the separation successfully performs 10 splits of merged nuclei, leaving 5 components subsuming multiple nuclei each.
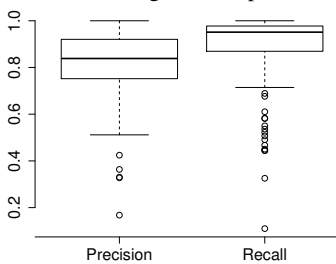
Fig. 3. Boxplots of the precision and recall of cell area for data set D1.

For evaluation of the snake segmentation, ground-truth labeling of the corresponding cell area was manually acquired. If an automatically segmented nucleus subsumes two or three true nuclei, the ground-truth cell area was defined as the union of the true cell areas of the merged nuclei. This procedure yields a total of 134 ground-truth cells.

Cell boundary segmentation is accomplished using normalized snakes with parameters $\omega_{cp} = \frac{15}{91}, \omega_{cv} = \frac{75}{91}, \omega_r = \frac{1}{91}, \lambda_{in} = 1$ and $\lambda_{out} = 100$. Note that β and ρ are absorbed by ω_{cp} and ω_r, respectively. The resulting segmentation is evaluated using precision and recall rates of pixels correctly assigned to cell area or background. The performance is summarized in box plots in Fig. 3. For all 134 cells a mean precision of 0.81 is achieved with a median of 0.84. For recall the mean is 0.88 with a median value of 0.95. For the outliers noticeable in the box plots two different reasons exist. First, in few cases a snake initialized for one nucleus expands beyond the boundary of the corresponding cell and occupies large fractions of the neighboring cell's area due to very low contrast at the boundary. Secondly, as noted above there are 19 cells without nucleus in the image. Therefore no snake can be initialized for these cells and in many cases the orphan cell area is included by a neighboring cell due to lack of distraction by the coupling term.
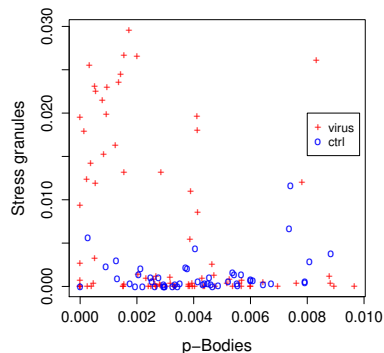
Fig. 4. Scatter plot of area fraction for both types of particles and both cell groups in data set D1.

For biological interpretation the fraction of the cell area occupied by PBs and SGs is relevant as cells exposed to viral infections (5 images) are compared to control cells (3 images). As ground-truth data are very difficult to collect for sub-cellular particles like PBs and SGs we relied on visual inspection of biological experts which rate the segmentation results for PBs and SGs as distinctly suited for biological interpretation (data not shown).

The area fractions for the 84 cells of the virus-infected test cells and the 50 cells of the control set are given in Fig. 4. The cells of the control (blue) contain a varying amount of PBs while with a few exceptions SGs remain absent. In contrast, for a part of cells exposed to the virus (red) SG-formation is induced, while the amount of PBs is reduced. This allows to hypothesize a viral infection to have been successful in only a fraction of all infected cells. Thus, widely-used average quantification over complete images is biased and does not reflect the situation correctly. In contrast, evaluation on a cell level as feasible with an automated approach allows to derive biological conclusions on a sound quantitative basis.

Fig. 6 (left) shows a sample clip for one image of data set D2. The images of D2 differ in their intensity characteristics from the ones in D1, hence, snake parameters need to be adapted. Also, to segment the complete cell areas including all focals mainly located along the sometimes quite dark exterior cell boundaries, preprocessing is required. The

Fig. 5. Mean area fraction of focals for all conditions of data set D2. The error bars have a height of two-times the estimated standard deviation and give an approximate 95% confidence interval for the mean estimated.

images are smoothed with a Gaussian of 4 pixels standard deviation prior to snake segmentation to diffuse intensity information and to adopt the bright focals for boundary identification (Fig. 6, middle).

The segmentation itself is then performed with parameters $\omega_{cp} = \frac{45}{120}, \omega_{cv} = \frac{75}{120}, \omega_r = 0, \lambda_{in} = 1$ and $\lambda_{out} = 150$. The stronger penalty for inhomogenity in background and the omission of regularization is supposed to cause the
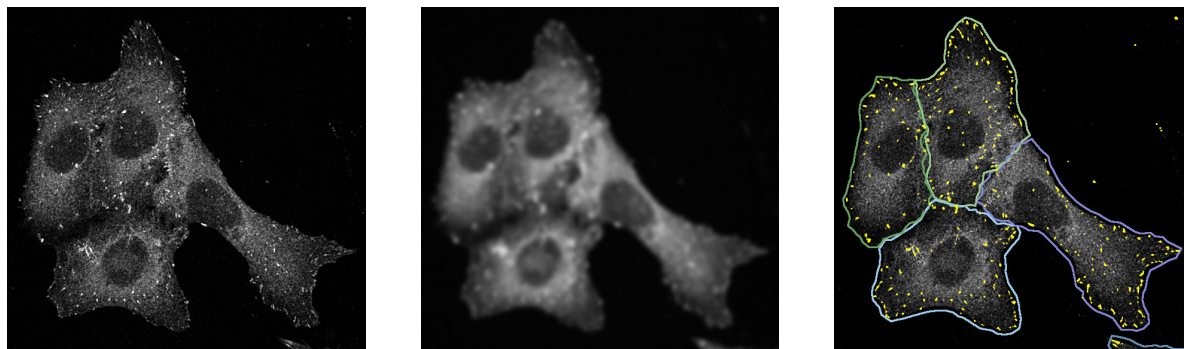
Fig. 6. Examples of segmentation results for data set D2: (left) contrast-enhanced clip from one original image; (middle) smoothed version; (right) segmented focal contacts (in yellow) and cell boundaries.

snake to enclose even outmost focal contacts of the cells.

The sample segmentation on the right of Fig. 6 proves this working quite well in many cases. It is clearly visible that the individual cell boundaries are accurately extracted, and that a large majority of detected focals is enclosed by the cell boundaries which is an indispensable prerequisite for biological data reliability. The area fraction of focals is given in Fig. 5 averaged over all cells for each of the three conditions monitored in data set D2. The first condition refers to cells with one type of mRNA silenced (3 images, denoted as *Silenced*) which should show reduced focal contacts compared to cells from the control of 2 images. For the third condition (3 images, denoted *Inhibitor*) a pharmacological inhibitor was added to silenced cells which should prevent focal contacts to be reduced.

Fig. 5 shows a significant difference between the mean area fraction of focals comparing treatment (*Silenced*) and the control cells confirming the hypothesized effect of silencing. Contrarily, in cells subsequently exposed to the pharmacological inhibitor the amount of focal contacts recovers approximately to the levels detected in control cells. This analysis again demonstrates the potential of MiCA to automatically derive biologically meaningful statistics from fluorescent microscope images.

VII. Conclusions and Future Work

The ImageJ plugin 'MiCA' provides in particular users from the life sciences with a useful collection of algorithms for cell image analysis. Relevant parts of a cell can be segmented within an integrated framework with high quality, resulting in quantitative data suitable for biomedical investigations. Future work mainly aims at extending MiCA's flexibility by integrating additional normalized snake energies and improving nucleus detection and separation.

References

[1] B. Möller, N. Stöhr, S. Hüttelmaier, and S. Posch, "Cascaded segmentation of grained cell tissue with active contour models," in *Proc. Int. Conf. on Pattern Recognition*, 2010, pp. 1481–1484.

[2] B. Möller, O. Greß, N. Stöhr, S. Hüttelmaier, and S. Posch, "Adaptive segmentation of cells and particles in fluorescent microscope images," in *Proc. VISAPP*, 2010, vol. 2, pp. 97–106.

[3] G. Srinivasa, M.C. Fickus, Y. Guo, A.D. Linstedt, and J. Kovacevic, "Active mask segmentation of fluorescence microscope images," *Trans. IP*, vol. 18, pp. 1817–1829, 2009.

[4] O. Dzyubachyk, W. Niessen, and E. Meijering, "Advanced level-set based multiple-cell segmentation and tracking in time-lapse fluorescence microscopy images," in *ISBI*, 2008, pp. 185–188.

[5] T.F. Chan and L.A. Vese, "Active contours without edges," *Trans. IP*, vol. 10, no. 2, pp. 266–277, 2001.

[6] H. Shen, G. Nelson, S. Kennedy, D. Nelson, J. Johnson, D. Spiller, M.R.H. White, and D.B. Kell, "Automatic tracking of biological cells and compartments using particle filters and active contours," *Chemometrics and Intelligent Laboratory Systems*, vol. 82, no. 1-2, pp. 276–282, 2006.

[7] M. Kass, A. P. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[8] A. Dufour, V. Meas-Yedid, A. Grassart, and JC Olivo-Marin, "Automated quantification of cell endocytosis using active contours and wavelets," in *Proc. Int. Conf. on Pattern Recognition*, 2008.

[9] A. Genovesio, T. Liedl, V. Emiliani, W.J. Parak, M. Coppey-Moisan, and J.C. Olivo-Marin, "Multiple Particle Tracking in 3-D+ Microscopy: Method and Application to the Tracking of Endocytosed Quantum Dots," *Trans. IP*, vol. 15, no. 5, 2006.

[10] JB Xavier, A. Schnell, S. Wuertz, R. Palmer, DC White, and JS Almeida, "Objective threshold selection procedure (OTS) for segmentation of scanning laser confocal microscope images.," *Journal of microbiological methods*, vol. 47, no. 2, pp. 169, 2001.

[11] S. Bolte and F. P. Cordelieres, "A guided tour into subcellular colocalization analysis in light microscopy," *Journal of Microscopy*, vol. 224, no. 3, pp. 213–232, 2006.

[12] I. Smal, W. Niessen, and E. Meijering, "A new detection scheme for multiple object tracking in fluorescence microscopy by joint probabilistic data association filtering," in *ISBI*, 2008, pp. 264–267.

[13] A. Gelas, K. Mosaliganti, A. Gouaillard, L. Souhait, R. Noche, N. Obholzer, and SG Megason, "Variational level-set with gaussian shape model for cell segmentation," in *International Conference on Image Processing (ICIP)*, 2009, pp. 1089–1092.

[14] A. Yezzi, A. Tsai, and A. Willsky, "A fully global approach to image segmentation via coupled curve evolution equations," *J. of Visual Comm. and Image Representation*, vol. 12, pp. 195–216, 2002.

[15] L. Opitz, A. Schliep, and S. Posch, "Joint analysis of in-situ hybridization and gene expression data," in *Advances in Data Analysis: Proc. of the 30th Ann. Conf. of the Gesellschaft für Klassifikation e.V.* 2007, pp. 577–584, Springer.

[16] O. Greß, B. Möller, N. Stöhr, S. Hüttelmaier, and S. Posch, "Scale-adaptive wavelet-based particle detection in microscopy images," in *Proc. of the Workshop "Bildverarbeitung für die Medizin" (BVM)*. 2010, pp. 266–270, Springer.

[17] A. Desolneux, L. Moisan, and J.-M. Morel, "A grouping principle and four applications," *Trans. PAMI*, vol. 25, no. 4, pp. 508–513, 2003.

[18] B. Möller, O. Greß, and S. Posch, "Knowing what happened - automatic documentation of image analysis processes," in *Int. Conf. on Computer Vision Systems*, 2011, accepted.

5.6 User-friendly assessment of pavement cell shape features with PaCeQuant: Novel functions and tools

Abstract:

Leaf epidermis pavement cells develop complex jigsaw puzzle-like shapes in many plant species, including the model plant *Arabidopsis thaliana*. Due to their complex morphology, pavement cells have become a popular model system to study shape formation and coordination of growth in the context of mechanically coupled cells at the tissue level. To facilitate robust assessment and analysis of pavement cell shape characteristics in a high-throughput fashion, we have developed PaCeQuant and a collection of supplemental tools. The ImageJ-based MiToBo plugin PaCeQuant supports fully automatic segmentation of cell contours from microscopy images and the extraction of 28 shape features for each detected cell. These features now also include the *Largest Empty Circle* criterion as a proxy for mechanical stress. In addition, PaCeQuant provides a set of eight features for individual lobes, including the categorization as type I and type II lobes at two- and three-cell junctions, respectively. The segmentation and feature extraction results of PaCeQuant depend on the quality of input images. To allow for corrections in case of local segmentation errors, the LabelImageEditor is provided for user-friendly manual postprocessing of segmentation results. For statistical analysis and visualization, PaCeQuant is supplemented with the R package PaCeQuantAna, which provides statistical analysis functions and supports the generation of publication-ready plots in ready-to-use R workflows. In addition, we recently released the FeatureColorMapper tool which overlays feature values over cell regions for user-friendly visual exploration of selected features in a set of analyzed cells.

5.7 rhizoTrak: a flexible open source Fiji plugin for user-friendly manual annotation of time-series images from minirhizotrons

Abstract:

Background and aims Minirhizotrons are commonly used to study root turnover which is essential for understanding ecosystem carbon and nutrient cycling. Yet, extracting data from minirhizotron images requires extensive annotation effort. Existing annotation tools often lack flexibility and provide only a subset of the required functionality. To facilitate efficient root annotation in minirhizotrons, we present the user-friendly open source tool rhizoTrak.

Methods and results rhizoTrak builds on TrakEM2 and is publicly available as Fiji plugin. It uses treelines to represent branching structures in roots and assigns customizable status labels per root segment. rhizoTrak offers configuration options for visualization and various functions for root annotation mostly accessible via keyboard shortcuts. rhizoTrak allows time-series data import and particularly supports easy handling and annotation of time-series images. This is facilitated via explicit temporal links (connectors) between roots which are automatically generated when copying annotations from one image to the next. rhizoTrak includes automatic consistency checks and guided procedures for resolving inconsistencies. It facilitates easy data exchange with other software by supporting open data formats.

Conclusions rhizoTrak covers the full range of functions required for user-friendly and efficient annotation of time-series images. Its flexibility and open source nature will foster efficient data acquisition procedures in root studies using minirhizotrons.

Appendix A

Classification scheme

Predominant contribution: own contribution more than 70%

Major contribution: own contribution between 30 and 70%

Minor contribution: own contribution less than 30%

Chapter 2: Keypoints, Spots, and Particles

- B. Möller and S. Posch,
  Robust Features for 2-D Electrophoresis Gel Image Registration,
  Electrophoresis, pp. 30(23):4137-4148, 2009

  Major contribution to the design of the study and the experiments.

  Predominant contribution to the implementation of the spot detectors.

  Major contribution to the experimental evaluation.

  Major contribution to writing and revising the manuscript.

- O. Greß, B. Möller, N. Stöhr, S. Hüttelmaier and S. Posch,
  Scale-adaptive Wavelet-based Particle Detection in Microscopy Images,
  Bildverarbeitung für die Medizin, CEUR Workshop Proceedings, pp. 266-270, 2010

  Major contribution to the concept of the algorithm.

  Minor contribution to the experimental evaluation.

  Minor contribution to writing and revising the manuscript.

- N. Bley, M. Lederer, B. Pfalz, C. Reinke, T. Fuchs, M. Glaß, B. Möller,
  and S. Hüttelmaier,
  Stress granules are dispensable for mRNA stabilization during cellular stress,
  Nucleic Acids Research 43(4), e26. doi: 10.1093/nar/gku1275

  Major contribution to the implementation of the software used for granule detection.

Chapter 3: Object Segmentation

- B. Möller, Y. Poeschl, R. Plötner, K. Bürstenbinder,
  PaCeQuant: A Tool for High-Throughput Quantification of Pavement Cell Shape Characteristics,
  Plant Physiology, pp. 175(3):998-1017, 2017

  > Major contribution to concept and design of the research.
  >
  > Predominant contribution to concept and implementation of the image analysis workflow.
  >
  > Minor contribution to the experimental evaluation.
  >
  > Major contribution to writing and revising the manuscript.

- B. Möller, Y. Poeschl, S. Klemm, and K. Bürstenbinder,
  Morphological Analysis of Leaf Epidermis Pavement Cells with PaCeQuant,
  Plant Cell Morphogenesis, Methods in Molecular Biology, Vol. 1992,
  Chap. 22, pp. 329-349, 2019 (invited book chapter)

  > Predominant contribution to the implementation of PaCeQuant.
  >
  > Major contribution to writing and revising the manuscript.

- B. Möller and K. Bürstenbinder,
  Semi-automatic Cell Segmentation from Noisy Image Data for Quantification of Microtubule Organization on Single Cell Level,
  Proc. of 16th Intern. Symp. on Biomedical Imaging (ISBI), pp. 199-203, 2019

  > Predominant contribution to concept and implementation of image analysis algorithms.
  >
  > Major contribution to the experimental evaluation.
  >
  > Predominant contribution to writing and revising the manuscript.

- B. Möller and S. Posch,
  Comparing active contours for the segmentation of biomedical images,
  Proc. of Intern. Symp. on Biomedical Imaging (ISBI), pp. 736-739, 2012

  > Major contribution to the design of the study and the experiments.
  >
  > Predominant contribution to the implementation of the snake segmentation.
  >
  > Major contribution to the experimental evaluation.
  >
  > Major contribution to writing and revising the manuscript.

- B. Möller, O. Greß, N. Stöhr, S. Hüttelmaier and S. Posch,
  Adaptive Segmentation of Particles and Cells for Fluorescent Microscope Imaging,
  VISIGRAPP 2010, Revised Selected Papers of Int. Joint Conf. on Computer Vision, Imaging and Computer Graphics. Theory and Applications, pp. 154-169, 2011

  > Major contribution to the concept of the algorithm for cell segmentation.
  >
  > Major contribution to the concept of the particle detection.
  >
  > Predominant contribution to the implementation of the cell segmentation.
  >
  > Predominant contribution to the experimental evaluation of the cell segmentation.
  >
  > Major contribution to writing and revising the manuscript.

- B. Möller, N. Stöhr, S. Hüttelmaier and S. Posch,
  Cascaded Segmentation of Grained Cell Tissue with Active Contour Models,
  Proc. of Intern. Conf. on Pattern Recognition (ICPR), pp. 1481-1484, 2010

  > Major contribution to the concept of the algorithm.
  >
  > Predominant contribution to the implementation of the algorithm.
  >
  > Major contribution to the experimental evaluation.
  >
  > Major contribution to writing and revising the manuscript.

- D. Misiak, S. Posch, M. Lederer, C. Reinke, S. Hüttelmaier and B. Möller,
  Extraction of protein profiles from primary neurons using active contour models and wavelets,
  Journal of Neuroscience Methods, pp. 225:1-12, 2014

  > Major contribution to the concept of the neuron segmentation workflow subsuming the primary supervision of D. Misiak during his Diploma thesis and in the initial phase of his PhD thesis in developing and implementing the neuron segmentation approach.
  >
  > Minor contribution to writing and revising the manuscript.

- B. Möller, B. Schreck and S. Posch,
  Analysis of Arabidopsis Root Images - Studies on CNNs and Skeleton-Based Root Topology,
  IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 7th Workshop on Computer Vision in Plant Phenotyping and Agriculture (CVPPA), pp. 1294-1302, 2021

  > Minor contribution to the study on convolutional neural networks.
  >
  > Predominant contribution to concept and implementation of the main root extraction.
  >
  > Major contribution to the analysis of the challenge data.
  >
  > Major contribution to writing and revising the manuscript.

Chapter 4: Object Shape and Texture

- D. Mitra, S. Klemm, P. Kumari, J. Quegwer, B. Möller, Y. Poeschl, P. Pflug,
  G. Stamm, S. Abel, and K. Bürstenbinder,
  **Microtubule-associated protein IQ67 DOMAIN5 regulates
  morphogenesis of leaf pavement cells in Arabidopsis thaliana**,
  Journal of Experimental Botany, 70(2):529-543, 2019

  > Predominant contribution to implementation of algorithms and tools
  >   for cell shape analysis.

  > Predominant contribution to experimental analysis of
  >   cell wall components (Fig. 6).

- B. Möller, E. Piltz, and N. Bley,
  **Quantification of Actin Structures using Unsupervised
  Pattern Analysis Techniques**,
  Proc. of 22nd Intern. Conf. on Pattern Recognition (ICPR), pp. 3251-3256, 2014

  > Predominant contribution to the concept of the workflow.

  > Predominant contribution to the implementation.

  > Predominant contribution to the experimental evaluation.

  > Predominant contribution to writing and revising the manuscript.

- K. Bürstenbinder, B. Möller, R. Plötner, G. Stamm, G. Hause,
  D. Mitra, and S. Abel,
  **The IQD Family of Calmodulin-Binding Proteins Links Calcium
  Signaling to Microtubules, Membrane Subdomains, and the Nucleus**,
  Plant Physiology, pp. 173(3):1692-1708, 2017

  > Predominant contribution to implementing the image analysis algorithms.

  > Predominant contribution to experimental analysis of microtubule patterns (Fig. 5).

  > Predominant contribution to experimental analysis of
  >   membrane localization patterns (Fig. 3D).

  > Minor contribution to writing the manuscript.

- B. Möller, L. Zergiebel, and K. Bürstenbinder,
  **Quantitative and Comparative Analysis of Global Patterns of (Microtubule)
  Cytoskeleton Organization with CytoskeletonAnalyzer2D**,
  Plant Cell Morphogenesis, Methods in Molecular Biology, Vol. 1992,
  Chap. 10, pp. 151-171, 2019 (invited book chapter)

  > Predominant contribution to the implementation of the software.

  > Major contribution to writing and revising the manuscript.

Chapter 5: Bioimage Analysis Software & Tools

- B. Möller, O. Greß and S. Posch,
  **Knowing what happened - Automatic Documentation of
  Image Analysis Processes**,
  Proc. of Intern. Conf. on Computer Vision Systems (ICVS), LNCS 6962, pp. 1-10, 2011

  Major contribution to the concept and design of Alida.

  Minor contribution to the implementation of Alida's operator concept
  and the mechanisms of automatic process documentation.

  Major contribution to the implementation of MiToBo's core and operators.

  Major contribution to writing and revising the manuscript.

- B. Möller and S. Posch,
  **A Framework Unifying the Development of Image Analysis Algorithms
  and Associated User Interfaces**,
  Proc. of Intern. Conf. on Machine Vision Applications (MVA), pp. 447-450, 2013

  Major contribution to Alida's concept of automatic user interface generation.

  Major contribution to the implementation of Alida's core components
  for automatic user interface generation.

  Predominant contribution to the implementation of the framework
  for graphical user interface generation within Alida.

  Major contribution to concept and implementation of Grappa subsuming
  supervision of S. Kirchner in his Bachelor thesis on the initial
  implementation of a graphical editor.

  Major contribution to writing and revising the manuscript.

- S. Posch and B. Möller,
  **Alida - Advanced Library for Integrated Development of
  Data Analysis Applications**,
  Journal of Open Research Software, pp. 5(1):7, 2017

  Major contribution to the concept, design and implementation of Alida,
  also ongoing after publication.

  Major contribution to software maintainance and infrastructure for Alida,
  also ongoing after publication.

  Major contribution to writing and revising the manuscript.

- B. Möller, M. Glaß, D. Misiak and S. Posch,
  MiToBo - A Toolbox for Image Processing and Analysis,
  Journal of Open Research Software, pp. 4(1):e17, 2016

  Major contribution to the concept, design and implementation of MiToBo,
  also ongoing after publication.

  Major contribution to software maintainance and infrastructure for MiToBo,
  also ongoing after publication.

  Major contribution to writing and revising the manuscript.

- B. Möller and S. Posch,
  MiCA - Easy Cell Image Analysis with Normalized Snakes,
  Proc. of Workshop on Microscopic Image Analysis with Appl. in Biology (MIAAB), 2011

  > Major contribution to the concept of MiCA.

  > Predominant contribution to the implementation of MiCA.

  > Predominant contribution to the experimental evaluation.

  > Major contribution to writing and revising the manuscript.

- Y. Poeschl, B. Möller, L. Müller, and K. Bürstenbinder,
  User-friendly assessment of pavement cell shape features with PaCeQuant: Novel functions and tools,
  Plant Cell Biology, Vol. 160 of Methods in Cell Biology, pp. 349-363, 2020
  (invited book chapter)

  > Predominant contribution to the implementation of image processing functions and tools.

  > Minor contribution to writing and revising the manuscript.

- B. Möller, H. Chen, T. Schmidt, A. Zieschank, R. Patzak, M. Türke,
  A. Weigelt and S. Posch,
  rhizoTrak: a flexible open source Fiji plugin for user-friendly manual annotation of time-series images from minirhizotrons,
  Plant and Soil, pp. 444:519-534, 2019

  > Major contribution to the concept of rhizoTrak.

  > Minor contribution to the software implementation.

  > Major contribution to writing and revising the manuscript.

Appendix B

Supplemental Material

This appendix contains supplemental material for some of the publications presented in this thesis. Only those parts of the material are included here which are suitable for printing and where reprinting is permitted by the copyright owner. The original files and additional material (if existing) can be found on the websites of the journals in which the articles were published.

In the following you will find supplemental material for these publications:

- N. Bley, M. Lederer, B. Pfalz, C. Reinke, T. Fuchs, M. Glaß, B. Möller, and S. Hüttelmaier. **Stress granules are dispensable for mRNA stabilization during cellular stress**, Nucleic Acids Research, 2015, 43(4), pp. e26, doi: 10.1093/nar/gku1275

- B. Möller, Y. Poeschl, R. Plötner, and K. Bürstenbinder. **PaCeQuant: A Tool for High-Throughput Quantification of Pavement Cell Shape Characteristics**, Plant Physiology, 2017, 175(3), pp. 998-1017, doi: 10.1104/pp.17.00961

- D. Mitra, S. Klemm, P. Kumari, J. Quegwer, B. Möller, Y. Poeschl, P. Pflug, G. Stamm, S. Abel, and K. Bürstenbinder. **Microtubule-associated protein IQ67 DOMAIN5 regulates morphogenesis of leaf pavement cells in Arabidopsis thaliana**, Journal of Experimental Botany, 2019, 70(2), pp. 529-543, doi: 10.1093/jxb/ery395

- K. Bürstenbinder, B. Möller, R. Plötner, G. Stamm, G. Hause, D. Mitra, and S. Abel. **The IQD Family of Calmodulin-Binding Proteins Links Calcium Signaling to Microtubules, Membrane Subdomains, and the Nucleus**, In: Plant Physiology, 2017, 173(3), pp. 1692-1708, doi: 10.1104/pp.16.01743

All of the above-mentioned articles and their supplemental material are open access and distributed under the terms of the Creative Commons CC BY license.

For the following papers links are provided to supplemental material on the journal websites:

- D. Misiak, S. Posch, M. Lederer, C. Reinke, S. Hüttelmaier and B. Möller. **Extraction of protein profiles from primary neurons using active contour models and wavelets**, Journal of Neuroscience Methods, pp. 225:1-12, 2014, doi: 10.1016/j.jneumeth.2013.12.009

- B. Möller, H. Chen, T. Schmidt, A. Zieschank, R. Patzak, M. Türke, A. Weigelt, and S. Posch. **rhizoTrak: a flexible open source Fiji plugin for user-friendly manual annotation of time-series images from minirhizotrons**, In: Plant and Soil, 2019, 444, pp. 519-534, doi: 10.1007/s11104-019-04199-3

B.1 Stress granules are dispensable for mRNA stabilization during cellular stress (Bley et al., Nucleic Acids Research, 2015)
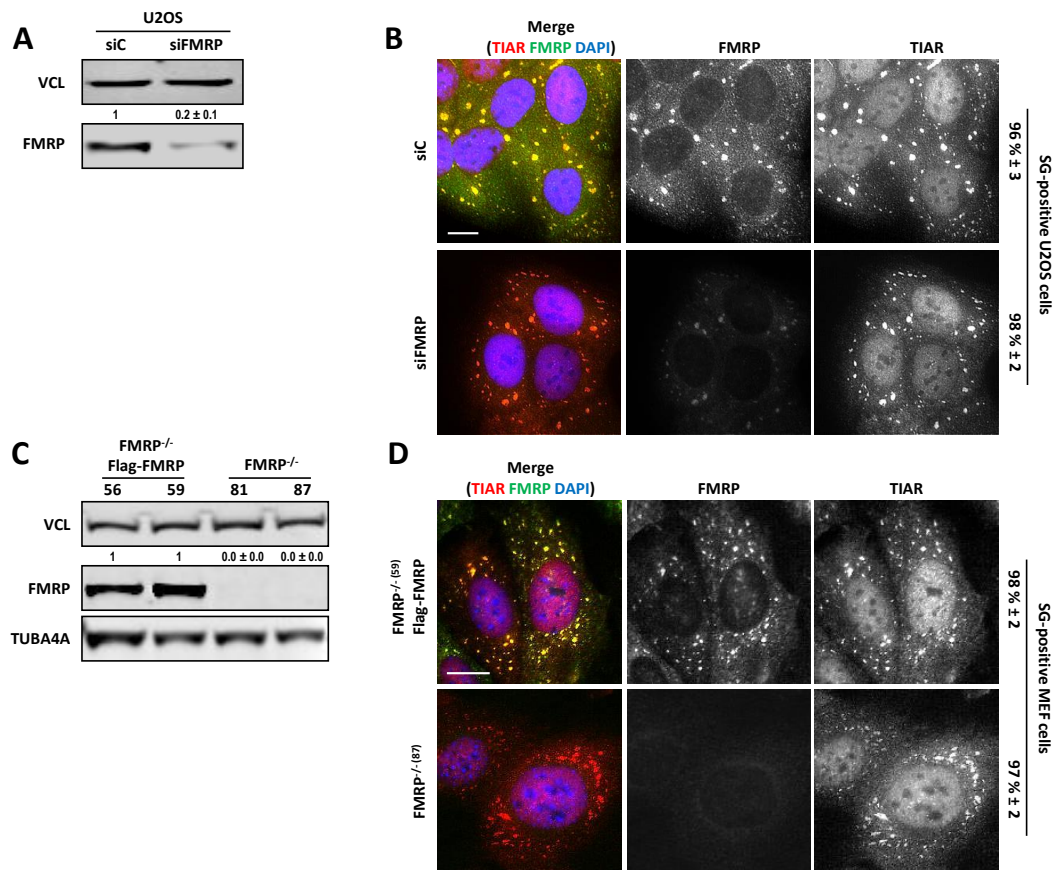
Bley et al., Figure S1
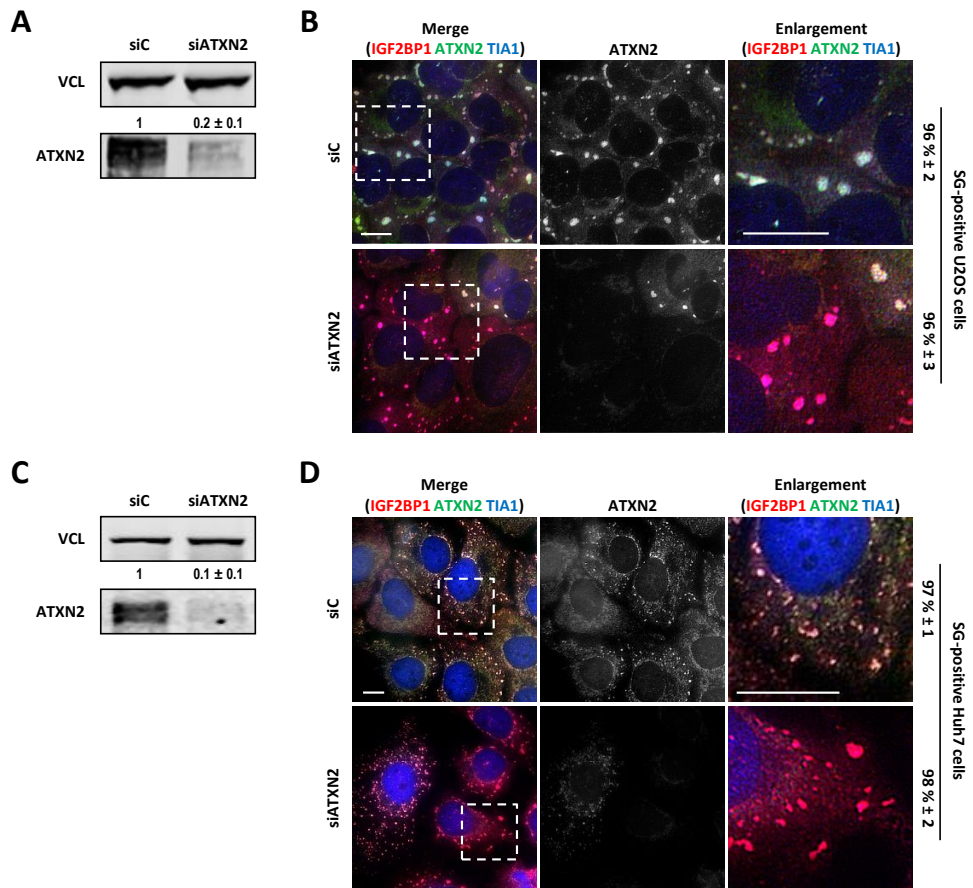
Bley et al., Figure S2
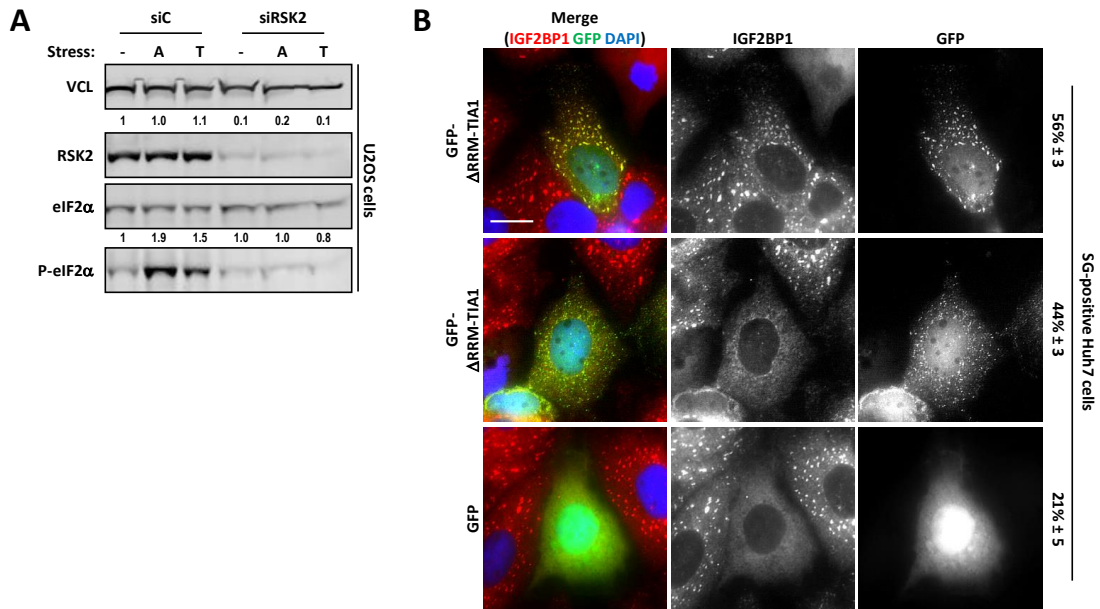
Bley et al., Figure S3
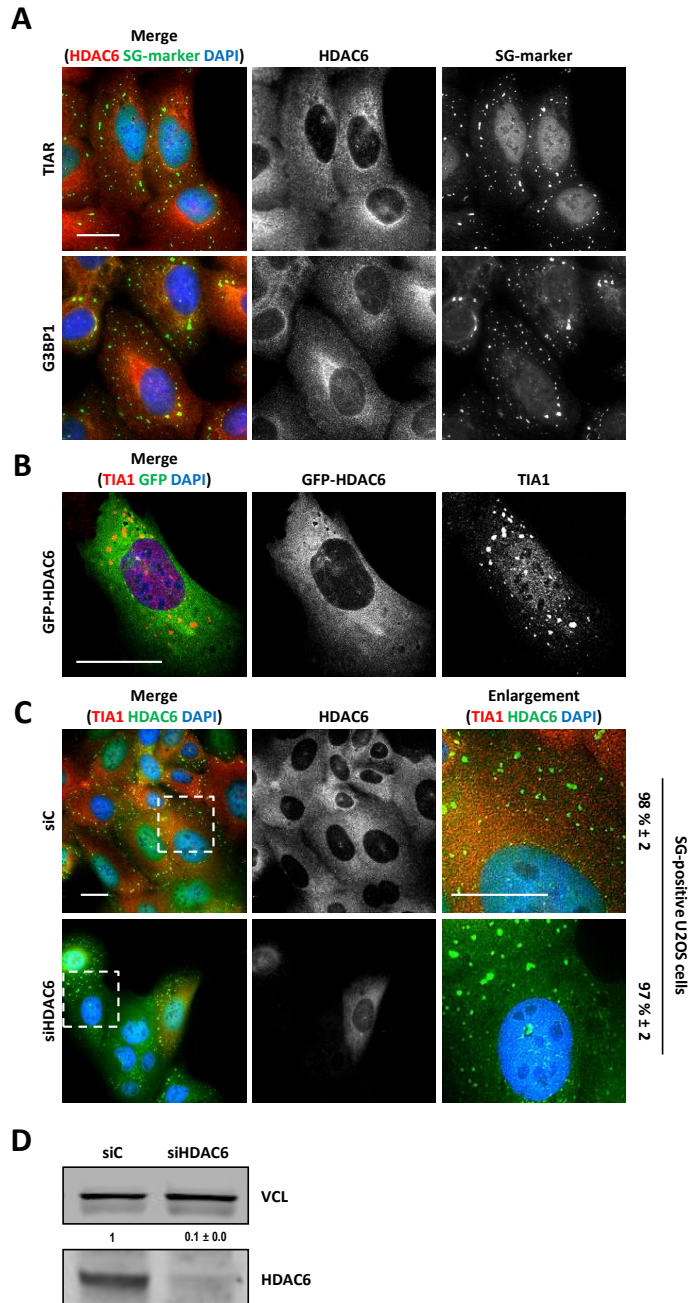
Bley et al., Figure S4
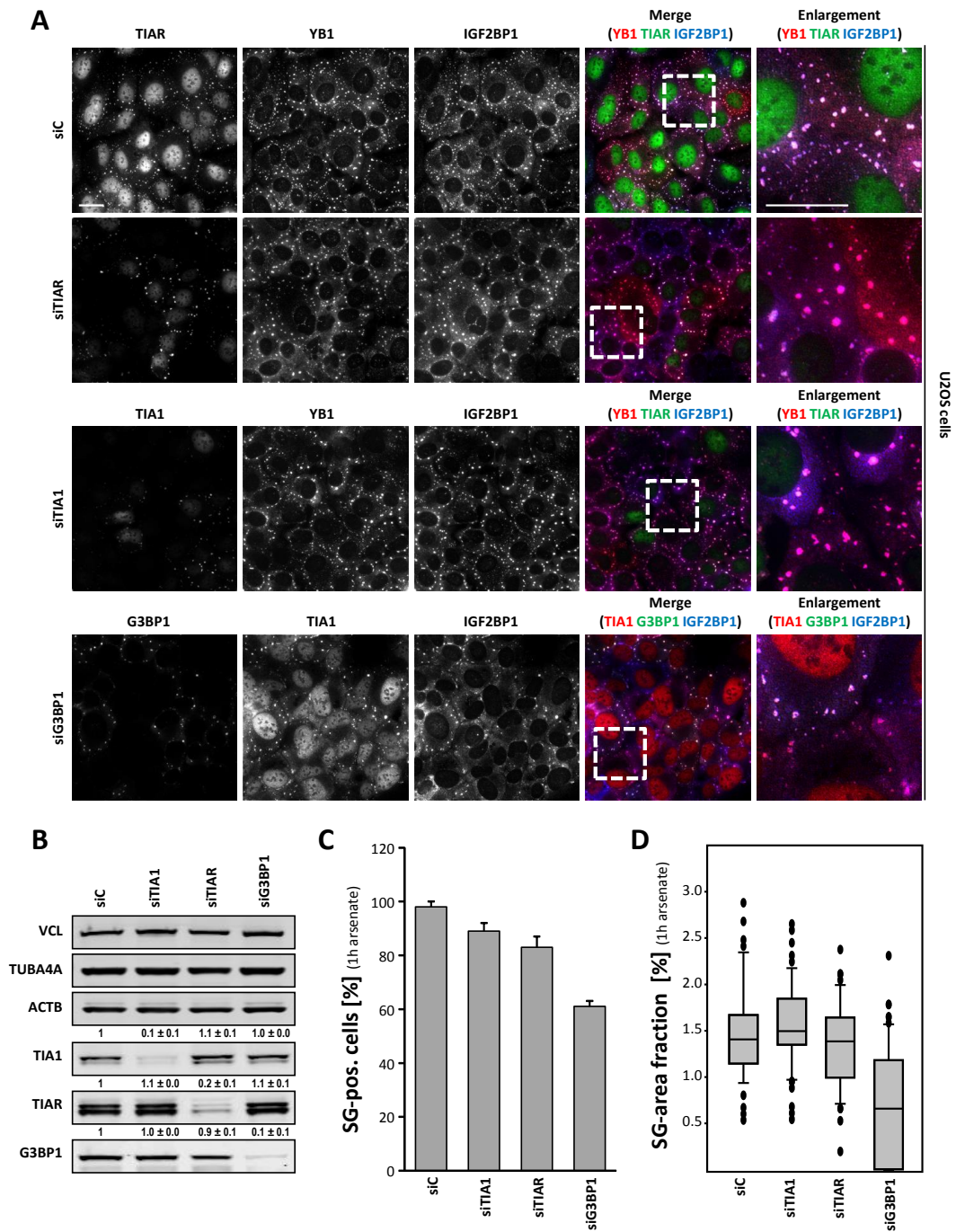
Bley et al., Figure S5
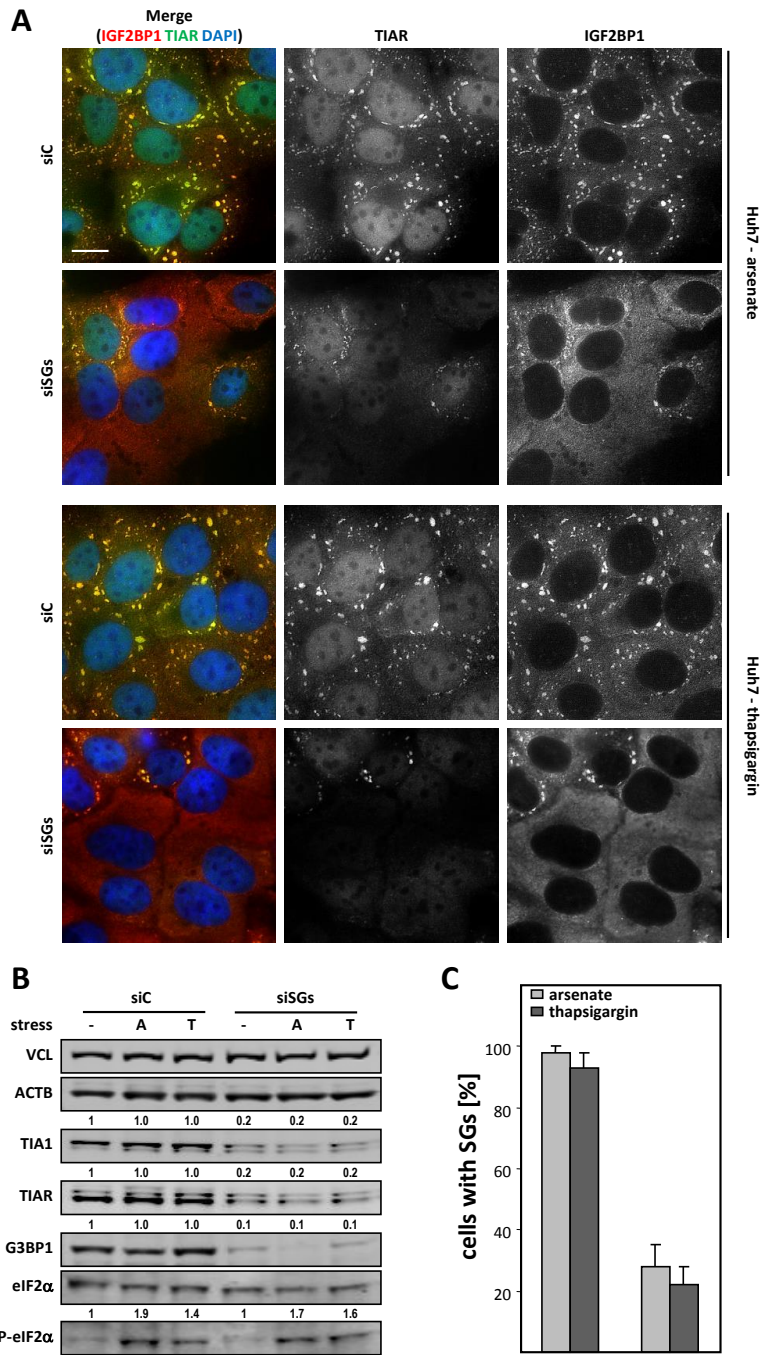
Bley et al., Figure S6
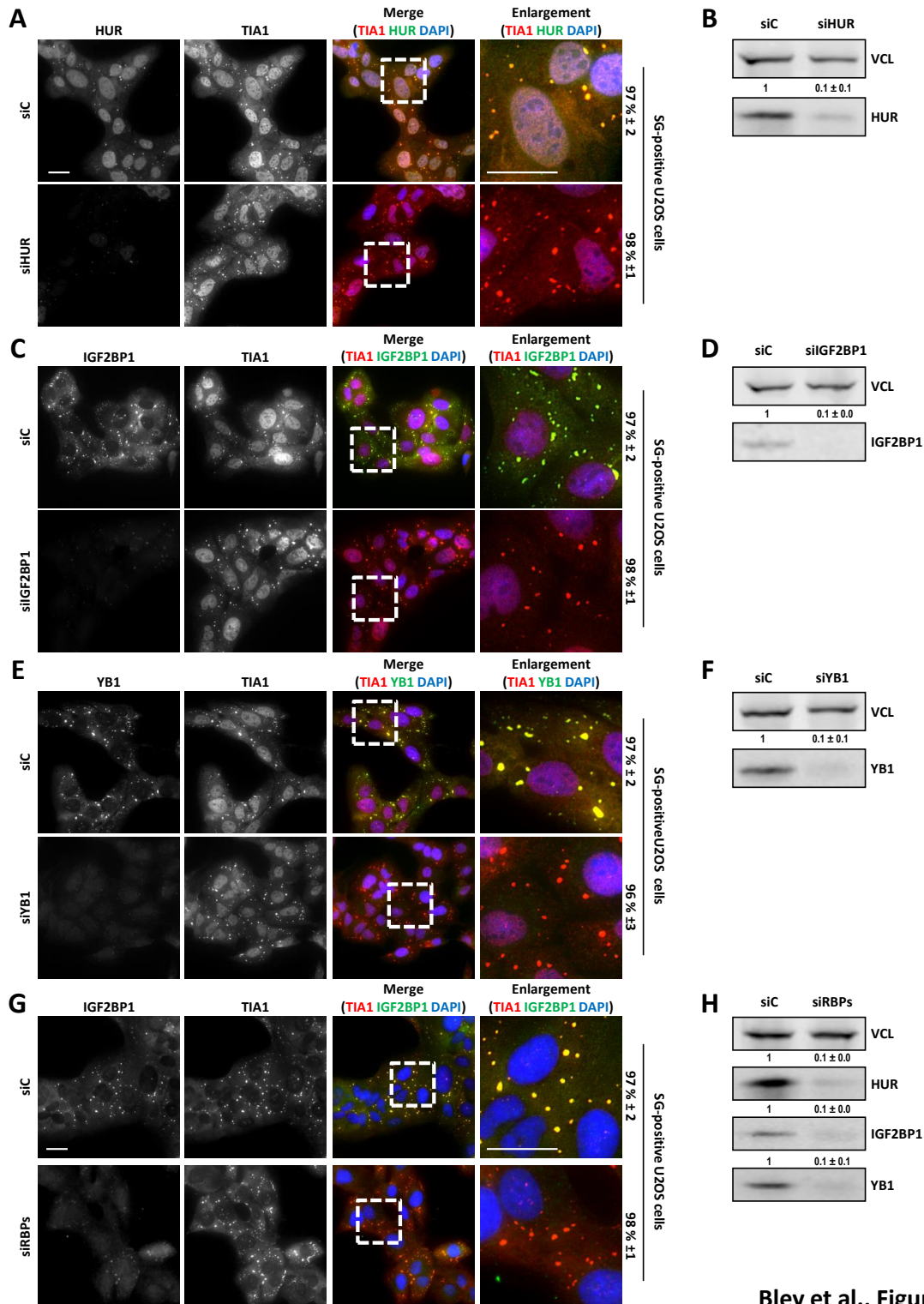
Bley et al., Figure S7

A

protein	conditions	t_{1/2} [sec] SGs	cytoplasm	immobile fraction [%] SGs	cytoplasm	experiments SGs	cytoplasm
G3BP1	OE-induced	6,6	1,1	36,2	15,2	10	10
	transient/arsenate	3,4		30,6		10	
	stable/arsenate	3,5	1,8	21,7	15,8	30	20
	all	3,5	1,6	23,7	15,7	50	30
TIA1	transient/arsenate	2,5	2,8	24,4	26,2	10	5
TIAR	transient/arsenate	2,2	3,0	24,9	29,4	10	5
ZBP1	transient/arsenate	13,6	1,9	71,4	25,8	10	5
	stable/arsenate	15,8	2,0	72,5	24,2	25	10
	all	14,8	2,0	72,3	24,6	35	15
IGF2BP1	transient/arsenate	19,5	5,8	71,7	20,9	10	5
HUR	transient/arsenate	17,7	9,3	57,0	31,9	10	5
YB1	stable/arsenate	20,7	2,7	70,9	29,7	10	5

Note: the $t_{1/2}$ header spans SGs and cytoplasm columns; immobile fraction [%] spans SGs and cytoplasm columns; experiments spans SGs and cytoplasm columns.

B
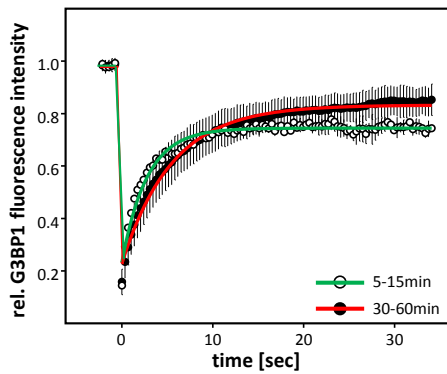
C
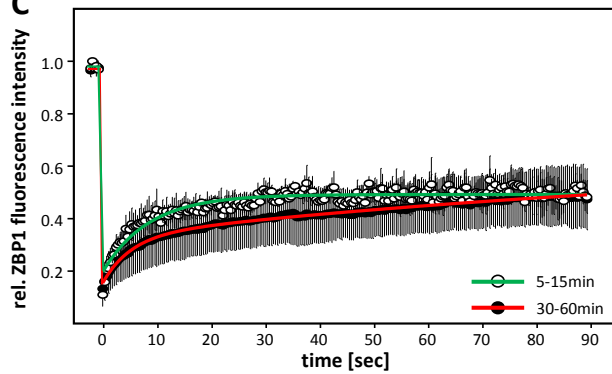
D

protein	arsenate	t_{1/2 in} [sec]	IF [%]
G3BP1	5-15min	1,3	33,3
	30-60min	1,8	21,7
ZBP1	5-15min	7,6	63,2
	30-60min	13,8	66,9

Bley et al., Figure S8

A

B

Bley et al., Figure S9

A

B

Bley et al., Figure S10
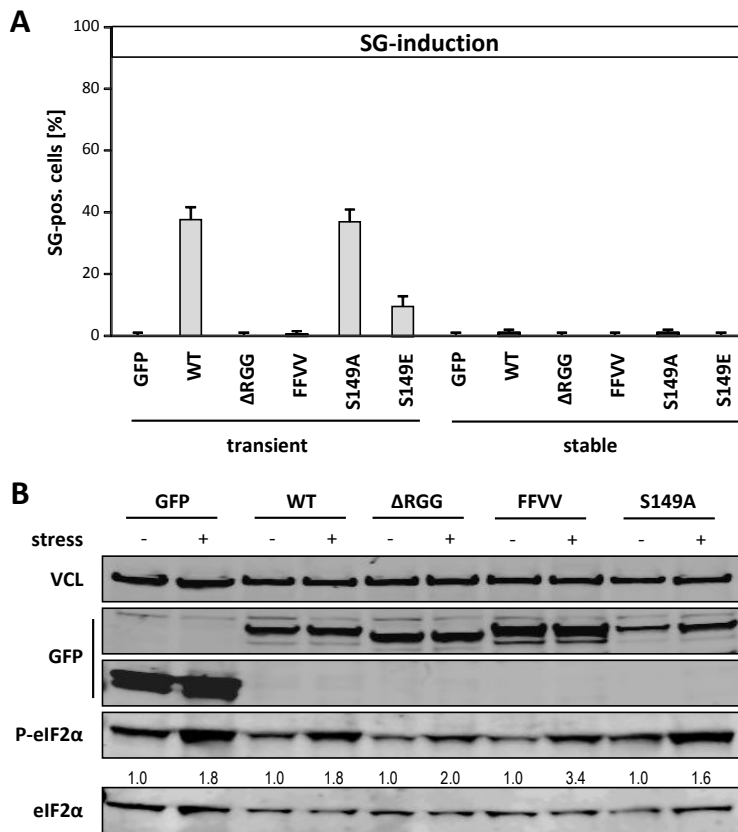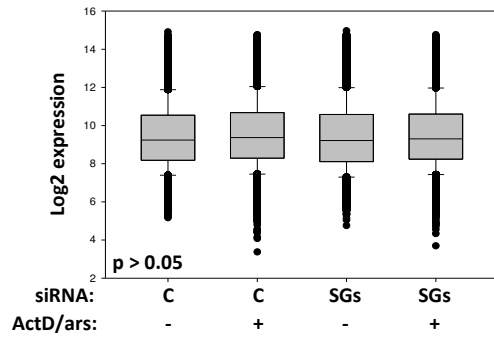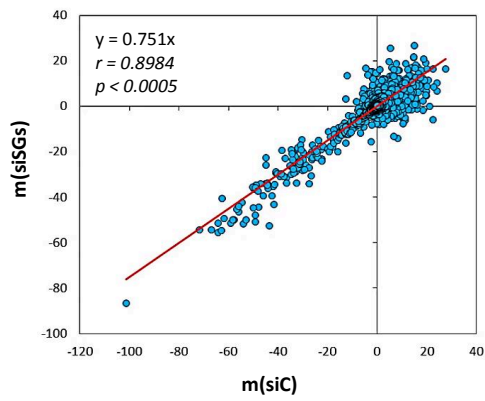
B. Supplemental Material

Table S1: siRNAs, plasmids and oligonucleotides

siRNAs	sequence
siC (C.el.mir239b)	uuguacuacacaaaaguacug
siIGF2BP1	ugaauggccaccaguugga
siYB1	gcgaagguuccaccuuacua
siHUR	ggacaaaaucuuacagguu
siTIA1	auucgagucuuuccagaua
siTIA1 (2)	caggaaagucuaagggaua
siTIAR	ugacagaaguccuuauacu
siTIAR (2)	gaaaggaggucaaaguaaa
siG3BP1	aaagccugagccaguauua
siG3BP1 (2)	cauuaacaguggugggaaa
siRSK2	ccaugaagguauugaaga
siHDAC6	gguaaagaagaaaggcaaa
siATXN2	uaugaggaugguucauaua

plasmids	vector	sense	antisense	cloning
GFP-DRRM-TIA1	pEGFP C2	aagaattcatgcgtcagacttttccacc	ctcgagtcactgggtttcataccctgc	EcoRI/XhoI in SalI
GFP-HDAC6	pEGFP C2	aaaga tct atgacctcaaccggccaggattccacc	aacccgggttttcattttctctgtggcccgg	BglII/Xma
		aacccgggagctgcaccgtgagagttccaac	aatctagattagtgtgggtggggcatatcctcccc	Xma/XbaI
GFP-G3BP1	described in Stohr et al. 2006	gggaattcatggtgatggagaagcctagtc	gggtcgacttactgccgtggcgcaagcccccttc	EcoRI/SalI
GFP-G3BP1	pLVX-puro-GFP	gcacgaattcatggtgatggagaagcctagtcccctgctgg	tggcgtcgactcactgccgtggcgcaagcccccttcccac	EcoRI/SalI
GFP-G3BP1-DRGG	pLVX-puro-GFP	gcacgaattcatggtgatggagaagcctagtcccctgctgg	ggcgtcgactcaatctcgtcggtcgcctttccctggcagctcg	EcoRI/SalI
GFP-G3BP1-FFVV	pLVX-puro-GFP	ggcgtcgactcaatctcgtcggtcgcctcccctggcagctcg	tcagaatcatcaaacacaacaacaacaccaacattgggtaatttcccac	mutagenesis
GFP-G3BP1-S149A	pLVX-puro-GFP	gagcctcaggaggaggctgaagaagaagtagagg	cctctacttcttcttcagcctcctcctgaggctc	mutagenesis
GFP-G3BP1-S149E	pLVX-puro-GFP	gagcctcaggaggaggaagaagaagaagtagagg	cctctacttcttcttcctcctcctgaggctc	mutagenesis
shATXN2	pSuper			
GFP-YB1	pEGFP C2, Clontech			subcloned from pcDNA3.1 via EcoRI/XhoI in SalI
GFP-HUR	pEGFP C2, Clontech			subcloned from pcDNA3.1 via EcoRI/XhoI in SalI
GFP-ZBP1	described in Stohr et al. 2006			
YFP-TIA1	kind gift from R.H. Singer, Albert-Einstein College of Medicine, NYC			
GFP-TIAR	pEGFP C2, Clontech	gggaattcatgatggaagacgacgggcagc	ggctcgagtcactgtgtttggtaacttgccatac	EcoRI/XhoI in SalI
Dendra-ZBP1	pDendra2-C, Clontech, modified MCS			subcloned from pcDNA3.1 via EcoRI/XhoI
Dendra-G3BP1	pDendra2-C, Clontech, modified MCS			subcloned from pcDNA3.1 via EcoRI/XhoI in SalI

qRT-PCR	sense	Antisense
HSP90AA	ggtcctgtgcggtcacttag	aaaggcgaacgtctcaacc
HSPAA/B1	caagatcaccatcaccaacg	tcgtcctccgctttgtactt
HSPA2	caagatcaccatcaccaacg	cgtcctccgctttgtacttc
MAPK4	tgctcaagattggggatttc	gatgagttgcatctgctcca
MYC	agcgactctgaggaggaac	cgtagttgtgctgatgtgtg
ACTB	agaaaatctggcaccacacc	agaggcgtacagggatagca
PPIA	gtcaaccccaccgtgttctt	ctgctgtctttgggaccttgt
RPLP0	ggcgacctggaagtccaact	ccatcagcaccacagccttc
VCL	ttacagtggcagaggtggtg	tcacggtgttcatcgagttc

Table S2: antibodies

antibodies	Company	Clone number
anti-IGF2BP1	Stohr et al., 2012	mouse monoclonal
anti-IGF2BP1	Huttelmaier et al., 2005	rabbit polyclonal
anti-YB1	Santa Cruz	59-Q
anti-HUR	Santa Cruz	sc-5261
anti-TIAR	BD Transductions	610352
anti-TIA1	Santa Cruz	N-17
anti-G3BP1	Santa Cruz	H-10
anti-eIF2a	Cell Signaling	9722
anti-phospho-eIF2a (S51)	Cell Signaling	L57A5
anti-RSK2	Cell Signaling	9340
anti-FMRP	Abcam	ab27455
anti-HDAC6	Santa Cruz	H-300
anti-ATXN2	BD Transductions	611378
anti-GFP	Santa Cruz	B-2

SUPPLEMENTARY FIGURE LEGENDS

Figure S1. FMRP is not essential for SG-formation. **(A, B)** U2OS cells were transiently transfected with indicated siRNAs for 72h. FMRP knockdown efficiencies were determined by Western blotting relative to the control population (siC) and VCL serving as loading control (A). The formation of SGs upon arsenate treatment (2.5 mM for 1h) was analyzed by immunofluorescence staining with indicated antibodies (B). The number of SG-positive cells was determined by counting of at least 50 cells from three independent experiments. **(C)** The expression of FMRP in immortalized FMRP (-/-) MEFs stably transfected with Flag-FMRP (56 & 59) or control (81 & 87) vector, as previously described (1), was analyzed by Western blotting as in (A). **(D)** The formation of stress granules by arsenate (2.5 mM for 1h) was analyzed by immunostaining with indicated antibodies and quantified as in (B). Note that the loss of FMRP does neither prevent the formation of SGs in U2OS nor FMRP (-/-) MEFs (data for clones 56 and 81 are not shown). Standard deviation was determined from three independent experiments. Bars 25 µm.

Figure S2. ATXN2 knockdown is insufficient to prevent SG-formation in U2OS and Huh7 cell. **(A-D)** U2OS (A, B) or Huh7 (C, D) cells transfected with indicated siRNAs for 72h were analyzed by Western blotting with indicated antibodies (A, C). Knockdown efficiencies for ATXN2 depletions depicted as numbers above panels were quantified relative to the control (siC). VCL served as loading control. The formation of stress granules was analyzed as in (Figure S1) using indicated antibodies for immunostaining (B, D). Note that the depletion of ATXN2 does not interfere with SG-assembly in both cell lines. Standard deviation was assessed from three independent analyses. Bars 25 µm.

Figure S3. Depletion of RSK2 or overexpression of ΔRRM-TIA1 are inappropriate to inhibit SG formation. (**A**) U2OS cells transfected with indicated siRNAs for 72h and treated with arsenate (A: 2.5mM for 1h) or thapsigargin (T: 1µM for 1h) were analyzed by Western blotting with indicated antibodies. Note that the knockdown of RSK2 impairs the stress-induced phosphorylation of eIF2α and consequently interferes with the assembly of stress granules (data not shown). (**B**) Huh7 cells transfected (48h) with the indicated GFP-tagged TIA1 mutant (ΔRRM) lacking two RRM domains or GFP alone were stressed by arsenate as in (A) before immunostaining with indicated antibodies. In contrast to previous reports (2), the formation of SGs was only modestly impaired by the overexpression of the TIA1-mutant. In approximately 44% of transfected cells, substantially fewer SGs were formed and granules appeared significantly smaller with barely any IGF2BP1. Most notably, however, SG-formation was substantially impaired in 20% of GFP-only transfected cells. Errors indicate s.d. of three independent analyses. Bars 25 µm.

Figure S4. HDAC6 is dispensable for SG-formation. (**A, B**) The localization of endogenous (A) or GFP-tagged (B) HDAC6 in U2OS cells stressed by arsenate (2.5mM, 1h) was analyzed by immunostaining with indicated antibodies. Note that endogenous as well as GFP-tagged HDAC6 are not recruited to arsenate-induced SGs. (**C, D**) U2OS cells transfected with HDAC6-directed (siHDAC6) or control (siC) siRNAs for 72h were treated as in (A) before immunostaining (C; number of SG-positive cells indicated) or Western blotting (D). Knockdown efficiencies were determined by Western blotting relative to the control population (siC) and VCL serving as loading control. Note that in contrast to previous reports (3), HDAC6 is neither recruited to stress granules nor does its knockdown prevent SG-formation in U2OS cells. Errors indicate s.d. of three independent analyses. Bars 25 µm.

Figure S5. The depletion of TIA proteins or G3BP1 alone barely impairs the formation of SGs. (**A-D**) U2OS cells were transfected with indicated siRNAs for 72h and stressed by arsenate (2.5µM) for 1h. SG-assembly was monitored by indirect immunostaining of indicated proteins (A). Enlargements of the boxed regions in the merged images are shown in the right panel. The knockdown of indicated proteins was monitored by Western blotting (B). VCL and TUBA4A served as loading controls to assess the knockdown efficiencies of all proteins relative to the control population (siC) as indicated by numbers above panels. The number of SG-positive cells (C) as well as the SG-area fraction (D) was determined via the Mica2D particle detector for ImageJ, essentially as previously described (4). Error bars indicate s.d. determined by analyzing at least 100 cells. Bars 25µm.

Figure S6. The concomitant depletion of TIA proteins and G3BP1 inhibits SG-assembly. (**A-C**) Huh7 cells transfected with indicated siRNAs (siC: control; siSG: siTIA1, siTIAR and siG3BP1) for 72h were stressed by arsenate (A; 2.5µM) or thapsigargin (T; 1µM) for 2h. SG-formation was monitored by immunostaining of indicated proteins (A). The knockdown efficiencies as well as the phosphorylation status of eIF2α were analyzed by Western blotting with indicated antibodies and quantified relative to the control population as indicated by numbers above panels using VCL and ACTB as loading controls. (B). The number of SG-containing cells was determined by counting. Error bars indicate s.d. of three independent experiments. Bars 25µm. Note that the concomitant depletion of TIA proteins and G3BP1 impairs SG-formation in a cell line and stressor independent manner (compare to Figure 1).

Figure S7. The formation of SGs remains unaffected by the single or concomitant depletion of the RBPs IGF2BP1, YB1 and HUR. (**A-H**) U2OS cells were transfected with control (siC), indicated RBP-directed siRNAs or a mixture of siRNAs targeting all three RBPs concomitantly. 72h post-transfection, cells were stressed by arsenate (2.5 mM for 1h) and SG-formation was monitored by immunostaining

with indicated antibodies (A, C, E, G). The average number of SG-positive cells was determined by counting, as indicated on the right. Errors indicate s.d. of three independent experiments. Bars 25 μm. The knockdown of indicated proteins was monitored by Western blotting and quantified relative to the control population (siC) as indicated by numbers above panels with VCL serving as loading control (B, D, F, H). Note that the formation of SGs remains unaffected by the single or concomitant depletion of the three RBPs (HUR, IGF2BP1 and YB1) analyzed.

Figure S8. FRAP (fluorescence recovery after photo-bleaching) parameters. (A) The table summarizes the exchange rates ($t_{1/2}$) and immobile fractions (in percent) averaged over the indicated number of analyzed SGs (experiments). The parameters were obtained by fitting a first order kinetic to the data. The conditions analyzed are indicated: OE-induced, SG-formation induced by transient overexpression without chemical stressor; transient/arsenate, SG-formation monitored in transiently transfected cells stressed by arsenate; stable/arsenate, SG-formation analyzed in arsenate-stressed cells stably expressing the indicated transgene. The data obtained for G3BP1, TIA1, TAIR, ZBP1, HUR or YB1 are shown in Figure 2A and B. **(B-D)** The average recovery of GFP-fluorescence in U2OS cells stably transfected with GFP-G3BP1 (B) or GFP-ZBP1 (C) was monitored shortly after SG-initiation (5-15 min arsenate treatment; green graph) or at steady state levels when SGs were completely assembled (30-60 min arsenate treatment; red graph). For ZBP1 the recovery time was increased to 90 sec to determine putative effects on the immobile fraction upon prolonged recovery time. A first order kinetic model was fitted to the data to determine the half time of fluorescence recovery ($t_{1/2\ IN}$) and immobile fractions (IF) (D). Note that the calculation of exchange rates does only apply at steady state levels, when $k_{IN} = k_{OUT}$. At the stage of SG-initiation however, we expect $k_{IN} > k_{OUT}$ since SGs are growing. Therefore only the half time of fluorescence recovery ($t_{1/2\ IN}$) can be determined.

 Figure S9. The RNA-binding domains of G3BP1 are essential to promote SG-assembly. (**A**) U2OS cells were transiently (48h) or stably transfected with GFP-tagged G3BP1 wild type or mutant constructs. GFP served as negative control. Stable transfections do not represent single clones but mixed populations of approximately 100% GFP-positive cells. The number of SG-containing cells in the absence of stress was determined by immunostaining as described in Figure 4A (data not shown). Error bars indicate s.d. determined by analyzing at least a 100 transfected cells for each mutant. Note that SG-formation is only induced by the transient expression of wild type and S149-modified G3BP1 mutants. (**B**) The phosphorylation of eIF2α in stressed (+, 2.5 mM arsenate for 1h) or non-stressed cells stably expressing the indicated G3BP1 mutant proteins or GFP was analyzed by Western blotting. Note that the overexpression of G3BP1 mutant proteins does not induce or impair the phosphorylation of eIF2α.

Figure S10. SGs are dispensable for the stabilization of bulk mRNA during cellular stress. (**A**) U2OS cells were concomitantly transfected (72h) with TIA1-, TIAR- and G3BP1-directed (siSGs) or control (C) siRNAs. Where indicated cells were stressed by arsenate (2.5mM) and treated with actinomycin D (ActD, 5µM) for 2h. For each experimental condition two independent total RNA populations were analyzed by microarrays (also compare to Figure 5C-E; Affimetrix HG133plus2.0). The expression of transcripts was analyzed without normalization and background correction. The averaged transcript abundance per condition is depicted as log2-expression in box plots. ANOVA testing revealed no statistical significant differences in-between the four experimental conditions (p > 0.01). This indicates that bulk mRNA stabilization during cell stress is neither impaired by the depletion of TIA1/TIAR/G3BP1 proteins nor by preventing the formation of SGs. (**B**) U2OS cells were transfected with two alternative sets of TIA1-, TIAR- and G3BP1-directed siRNAs or a control siRNA for 72h. Post-transfection, cells were treated with arsenate and ActD for 1h, 2h or 4h. The change in mRNA abundance was monitored by comparative microarray analyses (Solexa HumanHT12 chips; Illumina) using untreated cells as the input control. The quantile-normalized and background corrected signals reliably detected in all three untreated conditions were averaged for both alternative sets of siRNAs. The decay-rate for each detected signal was defined as the slope of signal intensities over time (m) determined by linear regression. The m-values determined for each signal in the siSG-transfected populations are plotted over the respective m-values determined in the control (siC) populations. Note that only ~3% of all transcripts decayed more than 2-fold during cellular stress. Moreover, Pearson correlation parameters indicated in the plot, confirmed a strong coherence (p < 0.0005) of the determined m-values. This indicates that the impairment of SG-formation does not interfere or promote mRNA degradation during cellular stress.

Table S1: siRNAs, plasmids and oligonucleotides. The table summarizes all siRNAs with the respective sequences used in this study. If not otherwise stated all plasmids used in this study were generated by PCR on cDNA obtained from HEK293 cells using the depicted oligonucleotide sequences. PCR products were cloned into ZeroBlunt- (Life Technologies) or pGemT-Vectors (Promega), sequenced and subcloned into the described expression vectors using the restrictions enzymes shown in the table. Sequences of primer pairs used for SYBR Green I based qRT-PCR are summarized.

Table S2: Antibodies. The table summarizes all primary antibodies as well as clone numbers and providers used in this study. Secondary antibodies used for Western blotting or immunostaining were previously described (4).

1. Castets, M., Schaeffer, C., Bechara, E., Schenck, A., Khandjian, E.W., Luche, S., Moine, H., Rabilloud, T., Mandel, J.L. and Bardoni, B. (2005) FMRP interferes with the Rac1 pathway and controls actin cytoskeleton dynamics in murine fibroblasts. *Hum Mol Genet*, **14**, 835-844.
2. Kedersha, N.L., Gupta, M., Li, W., Miller, I. and Anderson, P. (1999) RNA-binding proteins TIA-1 and TIAR link the phosphorylation of eIF-2 alpha to the assembly of mammalian stress granules. *J Cell Biol*, **147**, 1431-1442.
3. Kwon, S., Zhang, Y. and Matthias, P. (2007) The deacetylase HDAC6 is a novel critical component of stress granules involved in the stress response. *Genes Dev*, **21**, 3381-3394.
4. Stohr, N., Kohn, M., Lederer, M., Glass, M., Reinke, C., Singer, R.H. and Huttelmaier, S. (2012) IGF2BP1 promotes cell migration by regulating MK5 and PTEN signaling. *Genes Dev*, **26**, 176-189.

B.2 PaCeQuant: A Tool for High-Throughput Quantification of Pavement Cell Shape Characteristics (Möller et al., Plant Physiology, 2017)
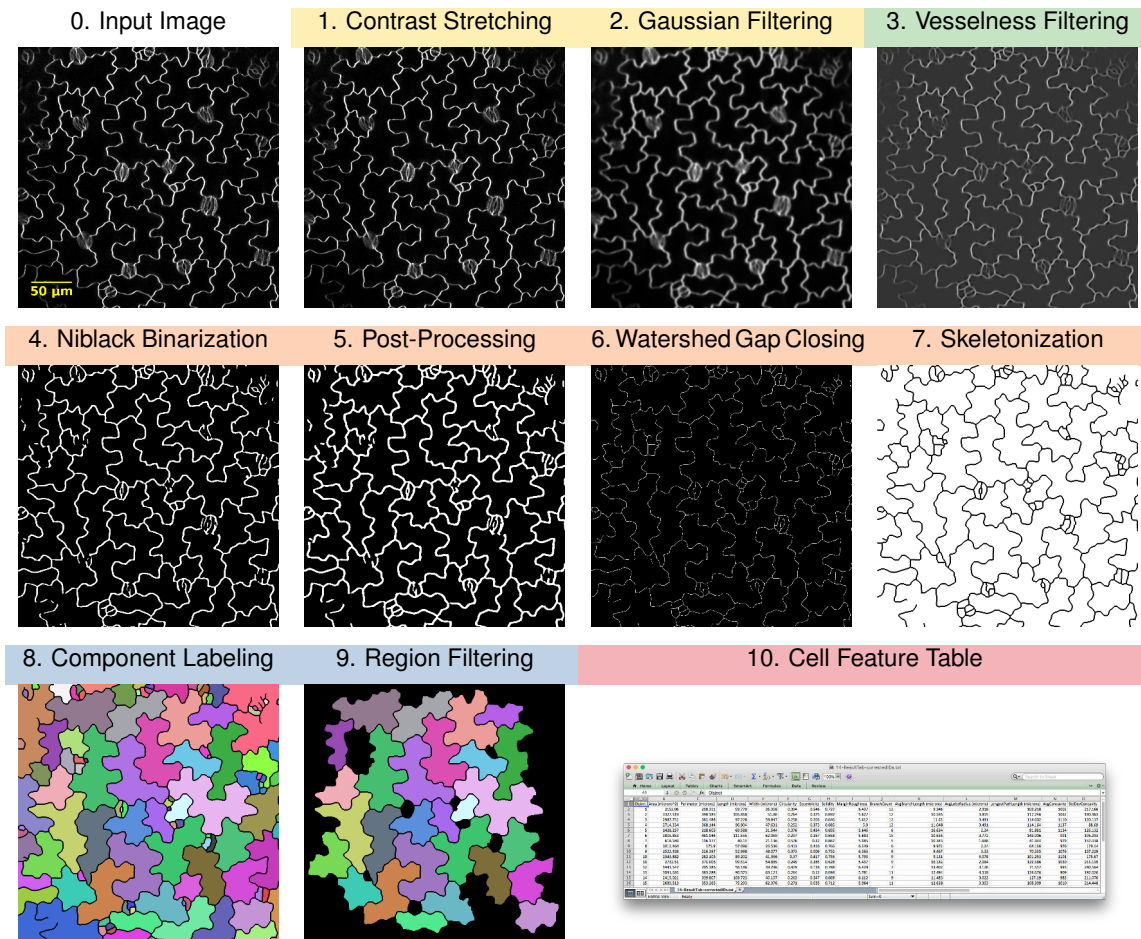
Figure S1

Figure S1. Effects of the Individual Processing Steps Implemented in PaCeQuant on the Input Image.

Shown are the input image (0.), the resulting intermediate images (1. to 8.), the output image (9.), and the final cell feature table for one exemplary sample during (I) image quality improvement, (II) boundary enhancement, (III) binarization and morphological post-processing, (IV) region filtering, and (V) feature extraction.

Figure S2
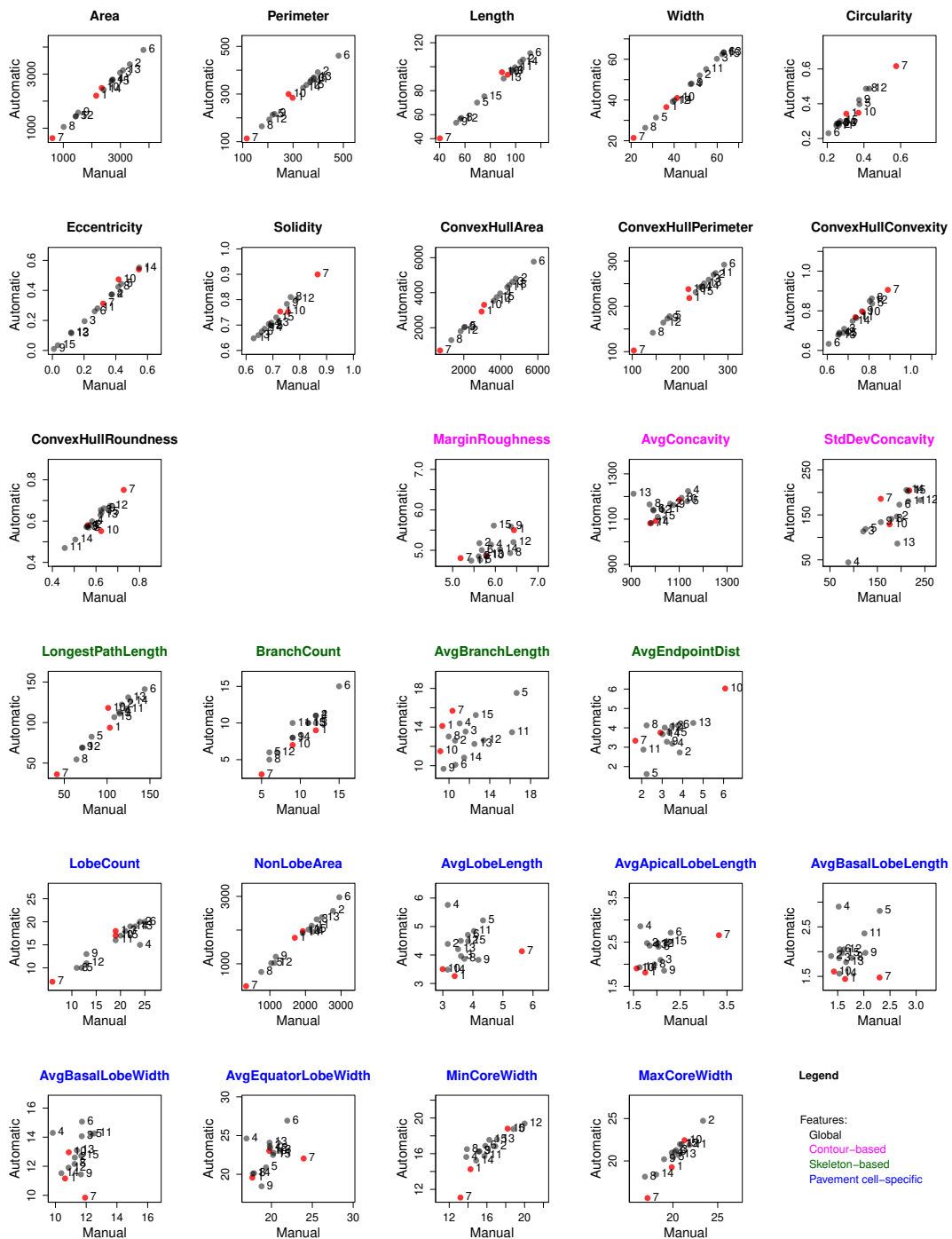
Figure S2. Comparison of PaCeQuant Features Between Manually and Automatically Segmented Cells.

Scatterplots of all 27 features (see Fig. 2 and Supplemental Tab. S1) from a pairwise comparison between the automatic and manual segmentation of the test set of 15 individual cells. Cells shown in Fig. 3B (ID 1, ID 7 and ID 10) are highlighted in red. Different heading colors represent the four feature groups (global, black; contour-based, pink; skeleton-based, green; and pavement cell-specific, blue).

Figure S3
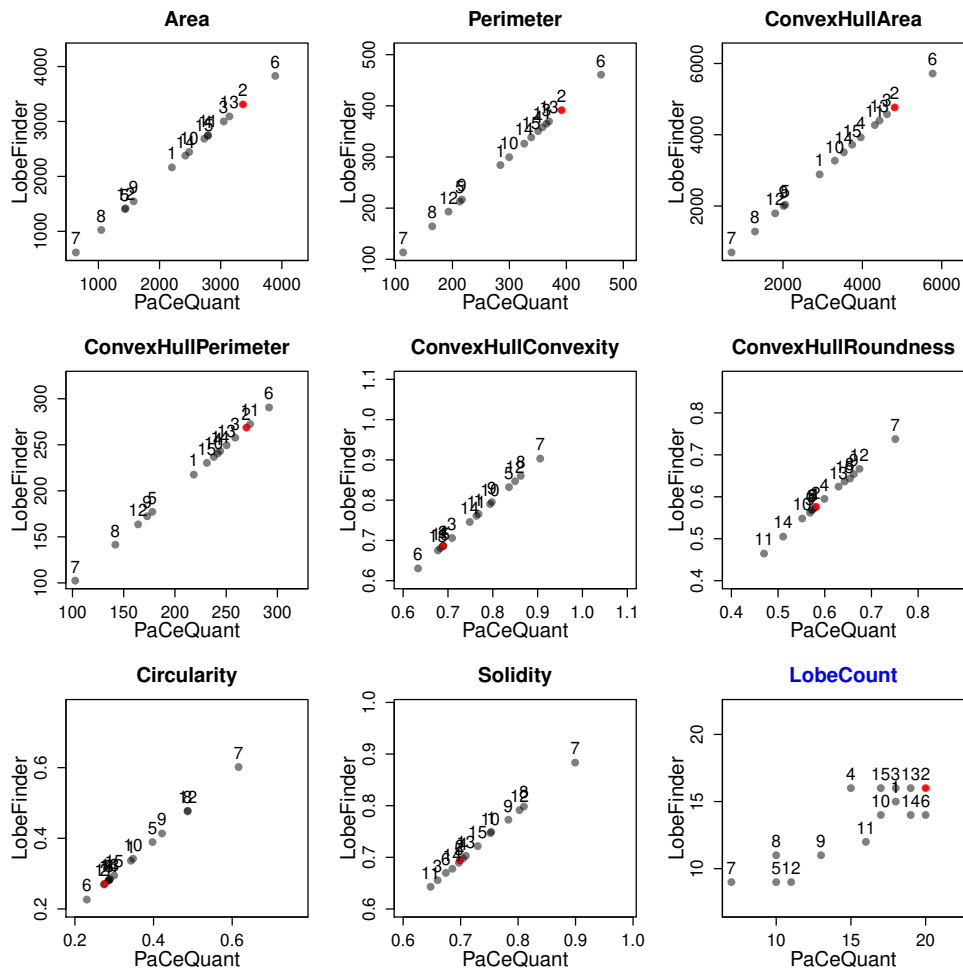
Figure S3. Pairwise Comparison of Features Computed by PaCeQuant and LobeFinder.

Scatterplots are shown for pairwise comparisons of features computed in the sample set of the 15 automatically segmented cells by PaCeQuant and LobeFinder. Black headings refer to global features, the blue heading indicates a pavement cell-specific feature. The cell shown in Fig. 4B (ID 2) is highlighted in red.

Figure S4

A
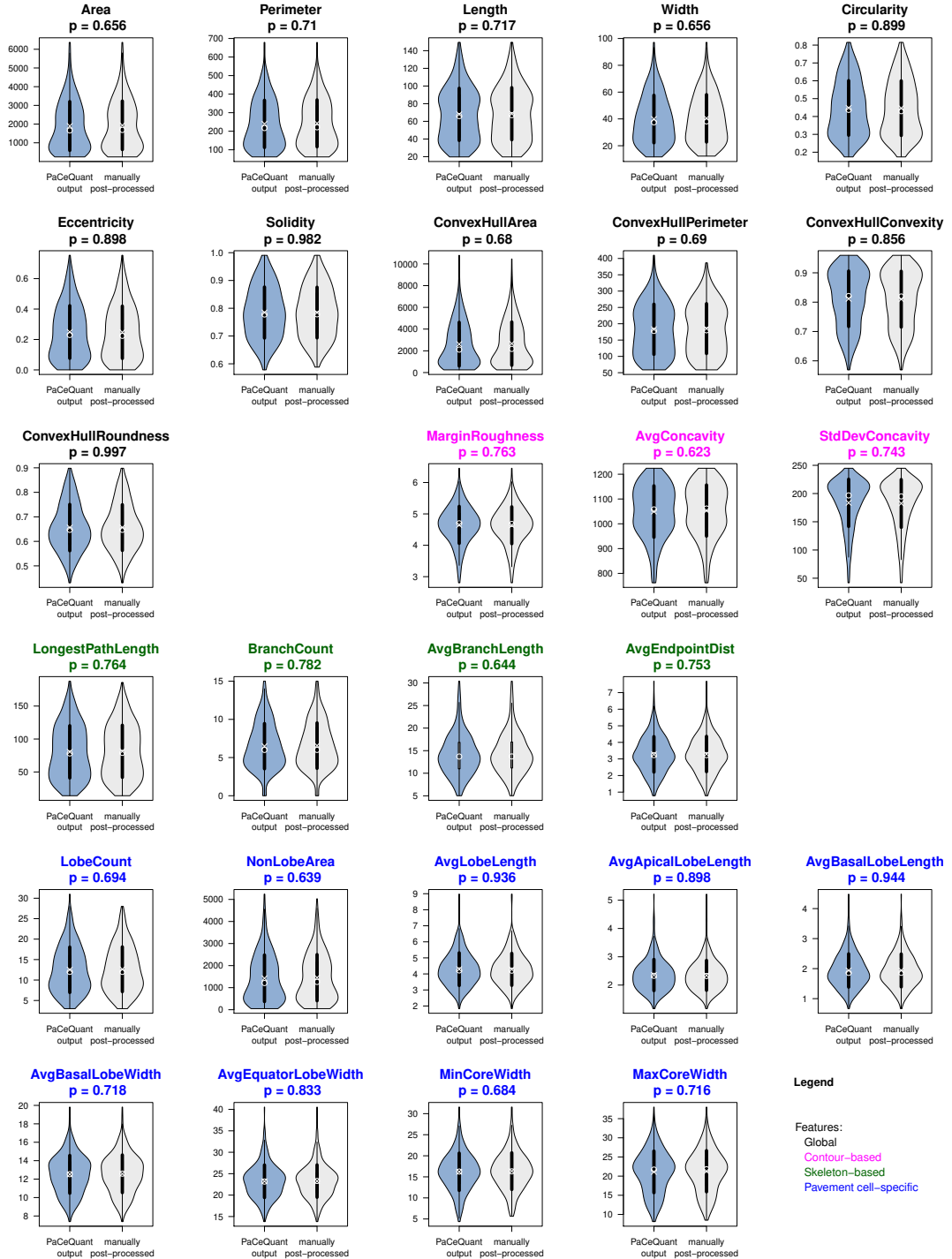
B
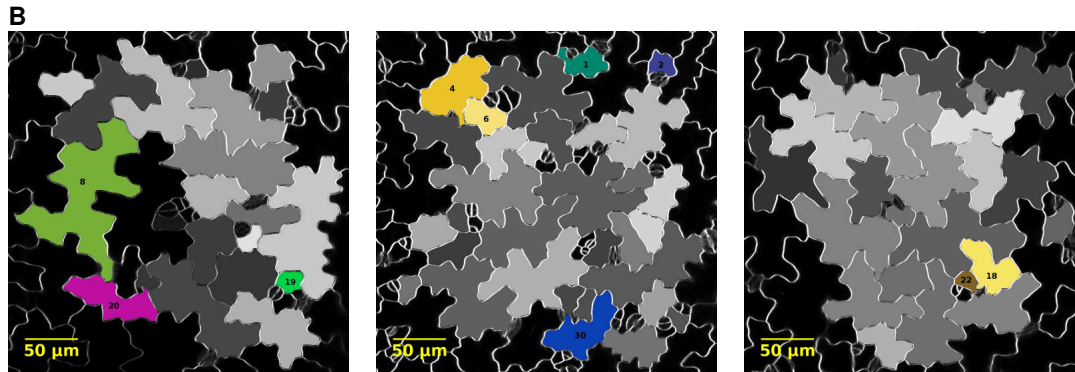
Figure S4. Comparison of Feature Values Between Raw PaCeQuant Output Data and After Cell Removal by Manual Filtering.

A sample set of 14 individual images from cotyledons of 5-day-old wild-type (Col-0) seedlings was subjected to PaCeQuant, which identified a total of 373 cells. Segmented regions with local deviations from cell contours were removed from the sample set by manual post-processing, which reduced the data set to 342 cells. **(A)** Violin plots for all 27 features from the 373 cells detected by PaCeQuant fully automatically (PaCeQuant output) and the 342 remaining cells after manual post-processing. Circles and crosses refer to medians and means, the vertical black lines in each category represent the standard deviation (thick lines) and the 95% confidence intervals (thin lines). The width of each violin box represents the local distribution of feature values along the y axes. P-values are derived from a two-sample Kolmogorov-Smirnov test, and are given for each feature. The different heading colors refer to the four feature groups. **(B)** Results from manual post-processing are shown for three exemplary images. Cells removed by manual filtering are highlighted in color; cells with accurately detected contours are shown in grey.

Figure S5
A
Group A - Global Features
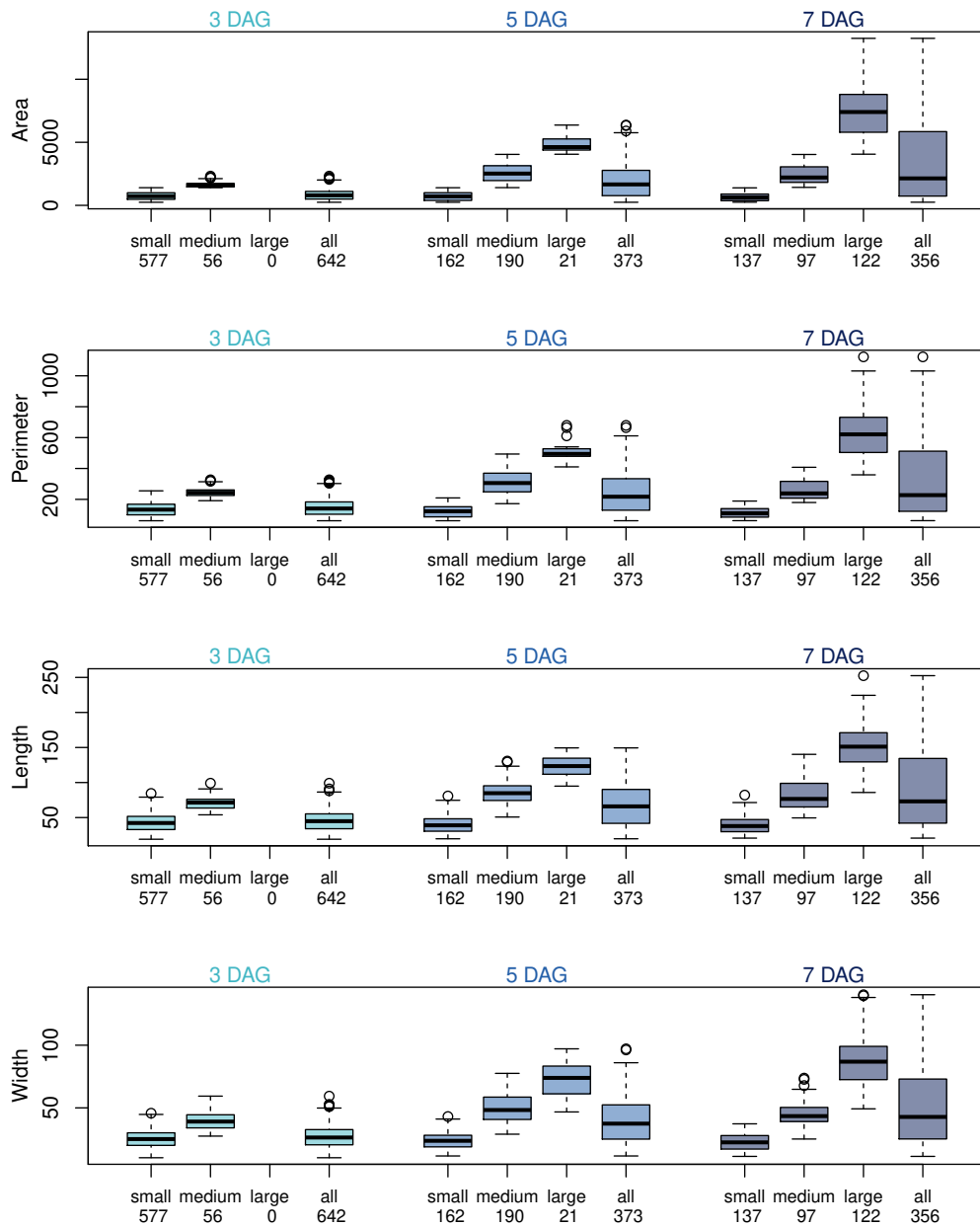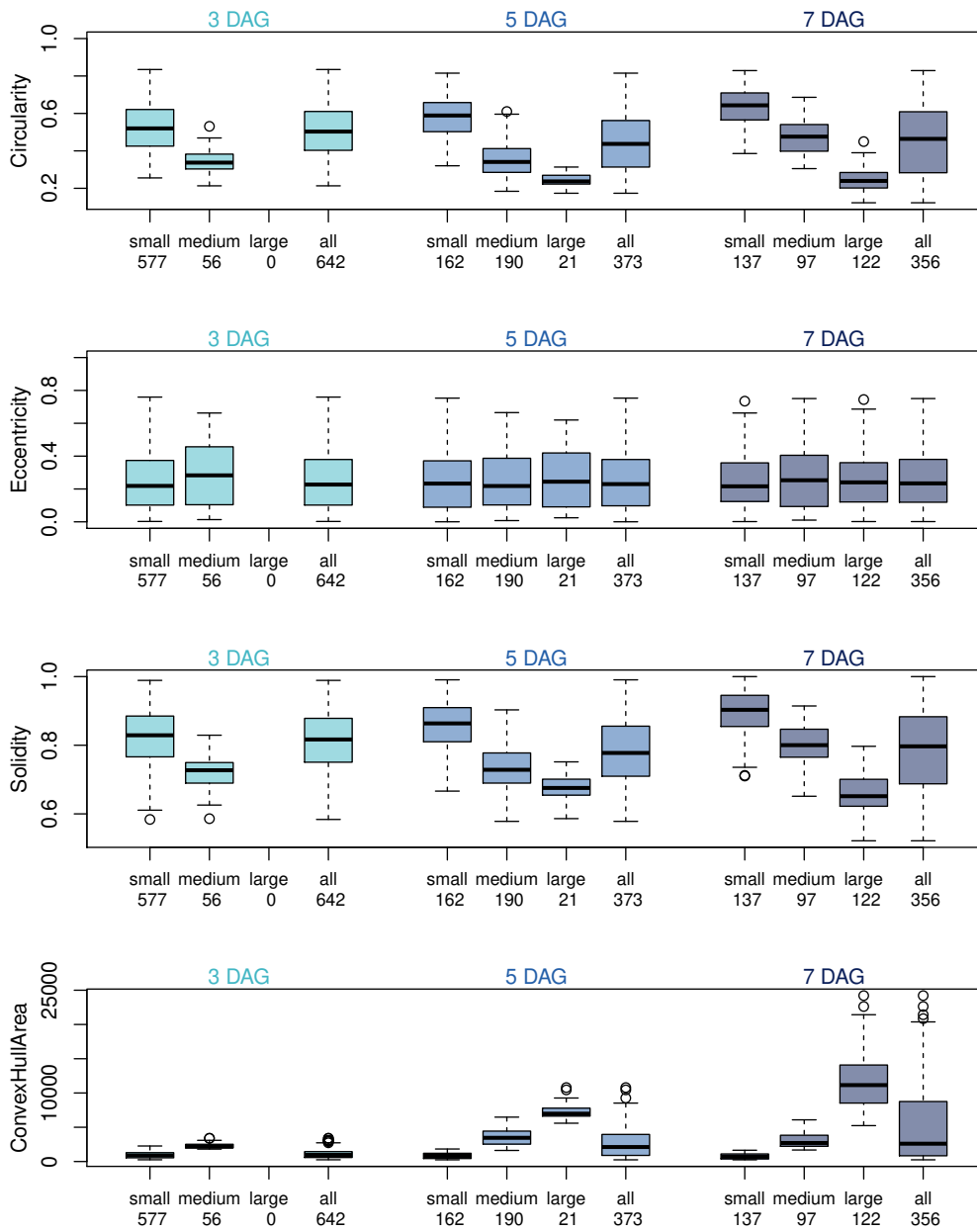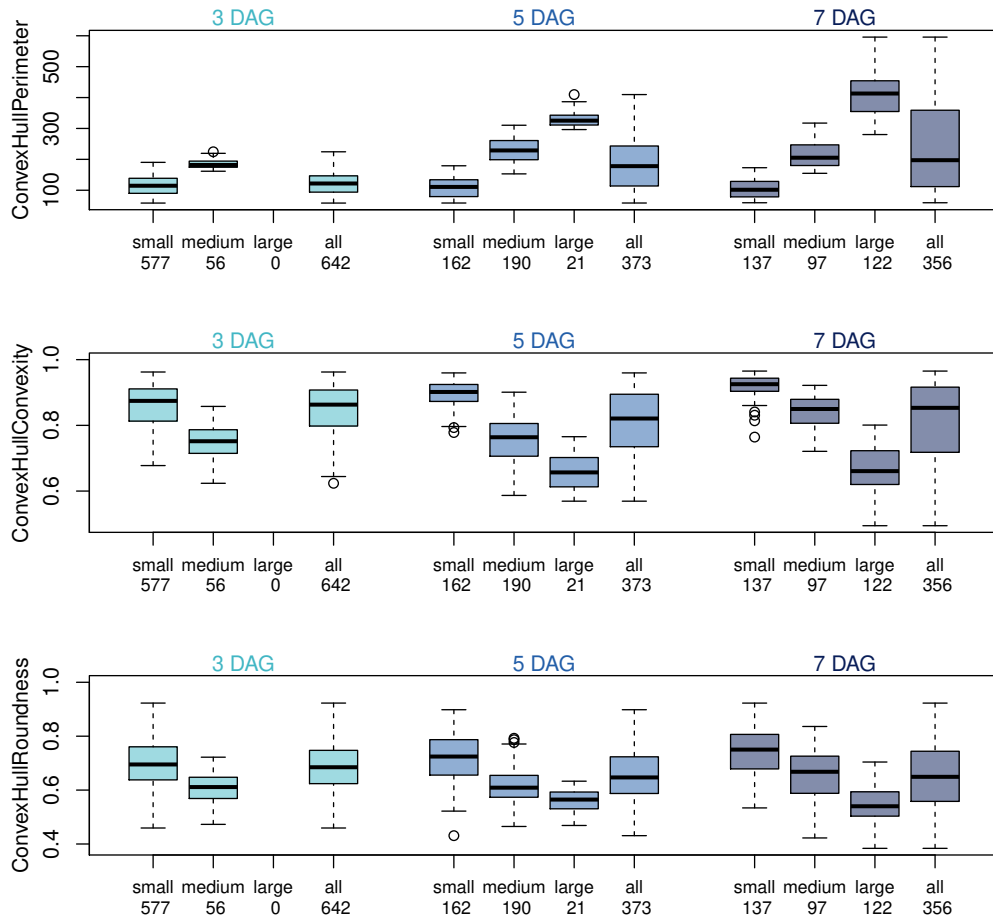
Group B - Contour-based Features

Group C - Skeleton-based Features
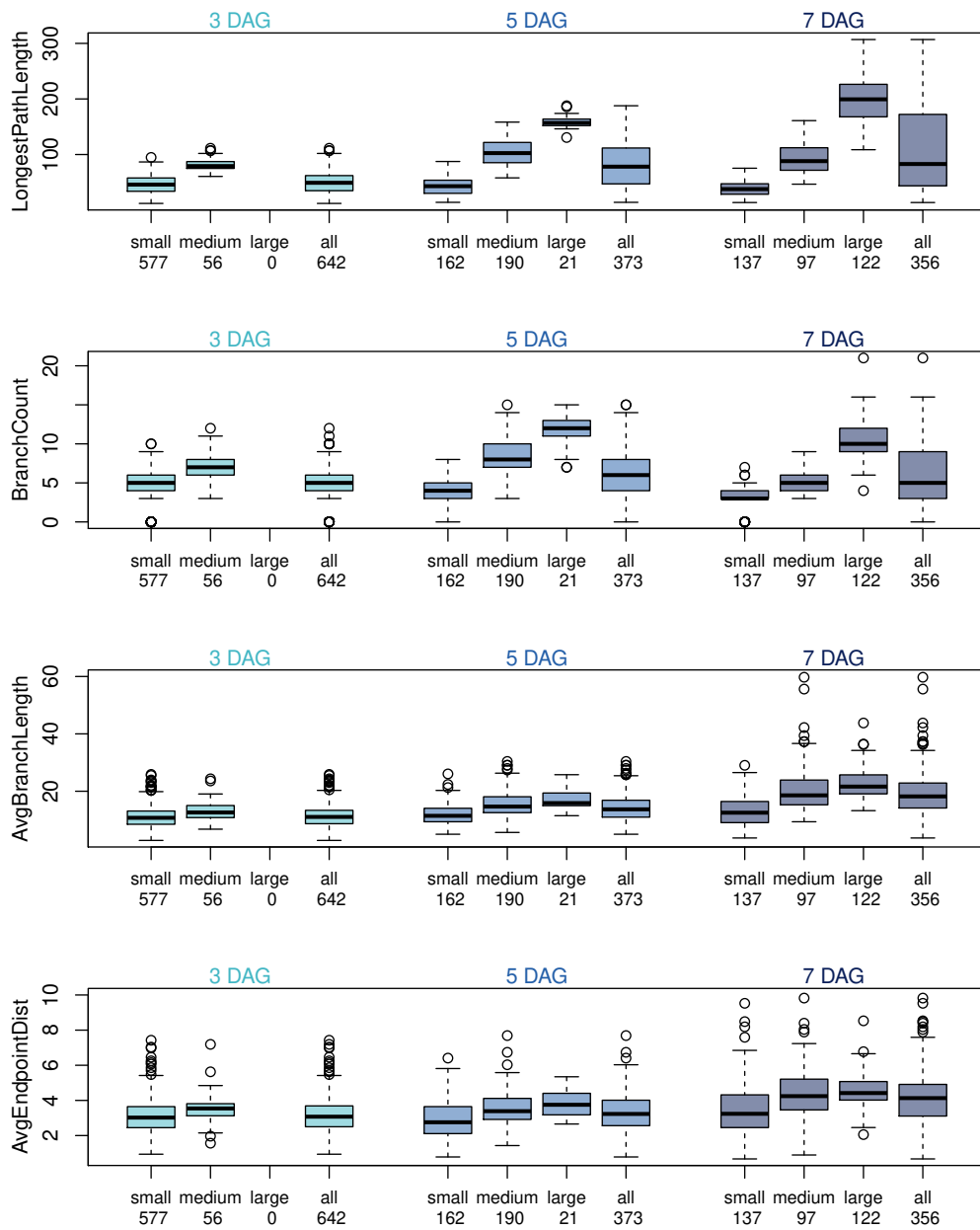
Group D - Pavement Cell-Specific Features

6
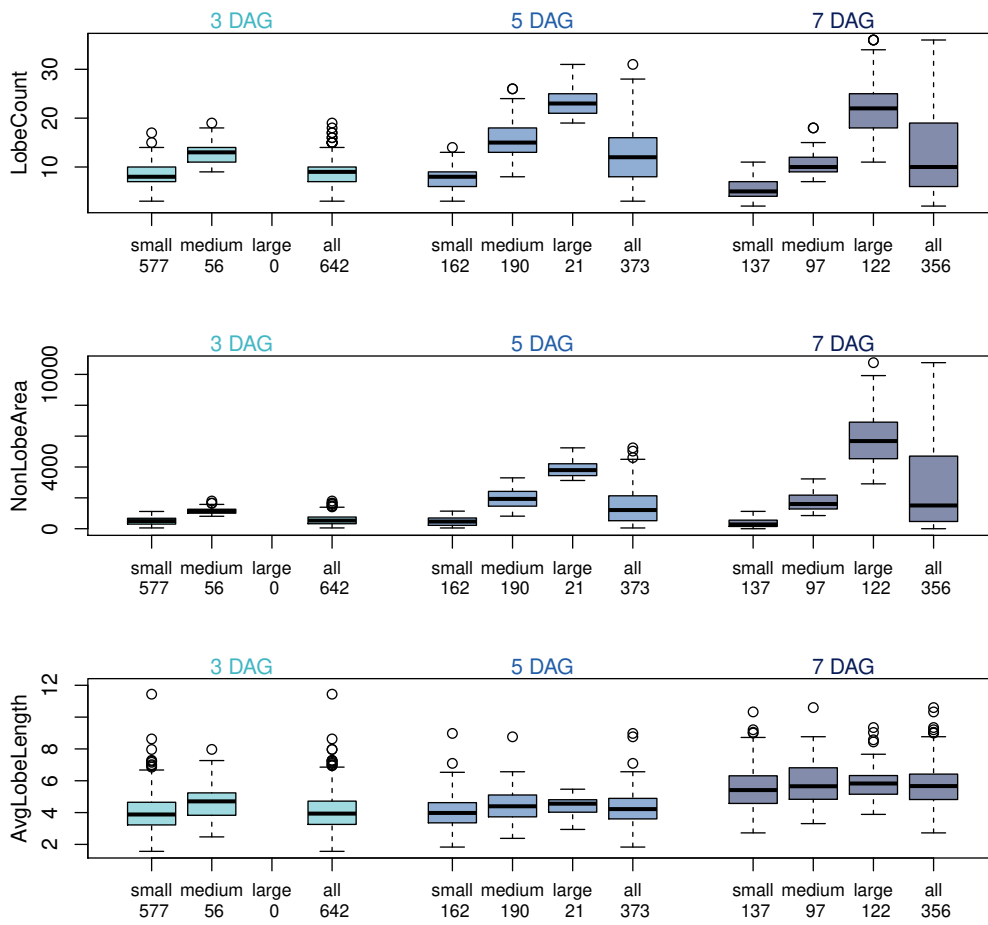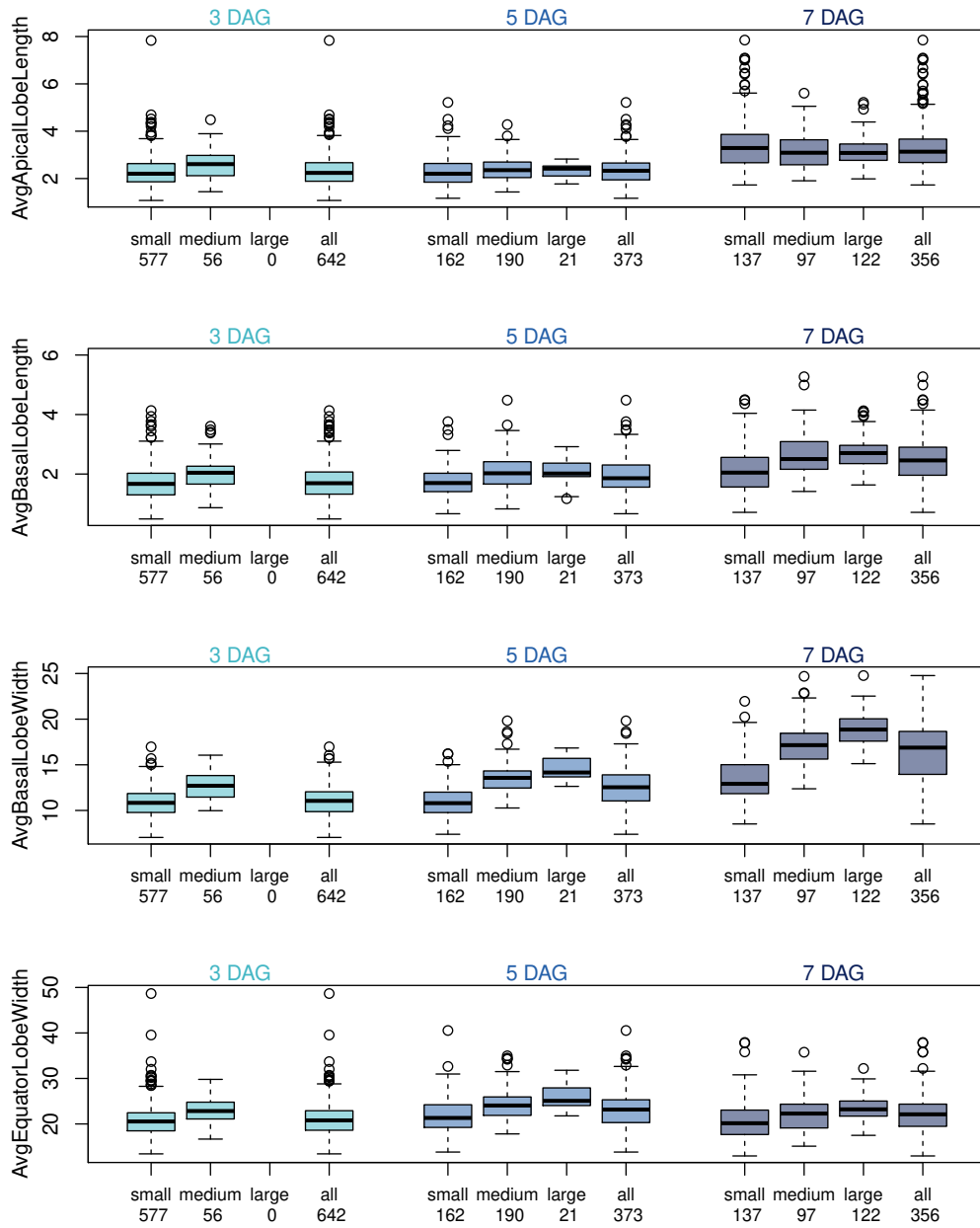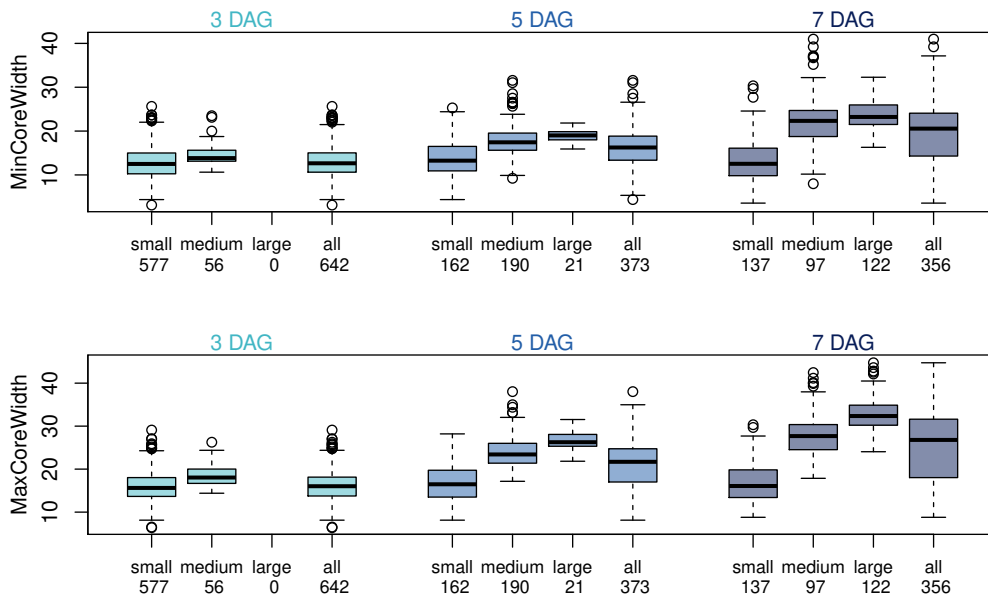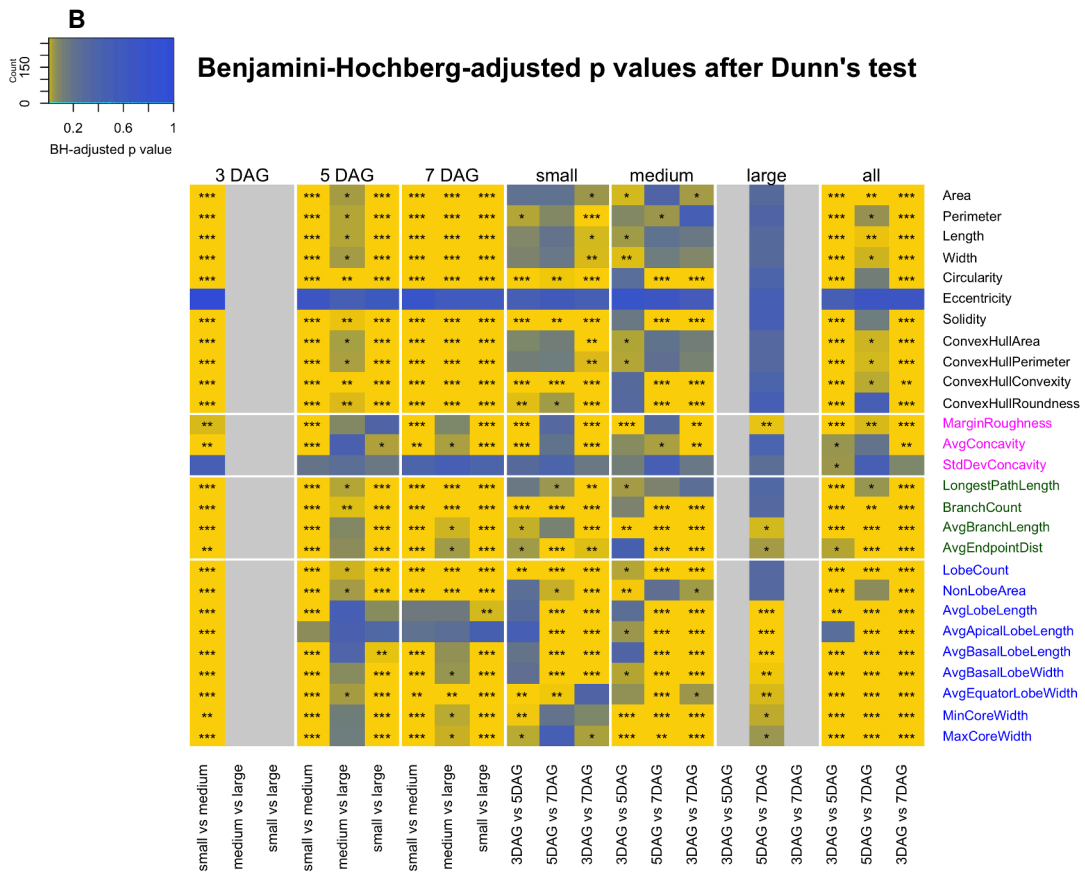
Figure S5. Quantification and Statistical Analysis of Pavement Cell Shape Features During Development.
(A) Cells were grouped according to their area (see Fig. 5), and feature values were calculated for small, medium and large-sized cells, and for the complete input dataset (all). Numbers on the x-axis refer to the number of cells analyzed per sample set. Feature values are shown in box plots. Box plots show medians (horizontal bar), interquartile ranges (IQRs; boxes), and data ranges (whiskers) excluding outliers (defiened as >1.5×IQR). **(B)** Statistical analysis from pairwise comparisons within the three data sets (3 DAG, 5 DAG, 7 DAG) and the four size categories (small, medium, large, all). Shown are Benjamini-Hochberg-adjusted p-values after Dunn's pairwise test. Blue colors represent p-values close to 1, yellow colors represent values close to 0. Stars indicate statistically significant differences (* padj < 0.05, ** padj < 0.01, *** padj < 0.001).

9

Figure S6

A

B

Name of Feature	Short Description
Object	numerical lobe identifier
Type	lobe type, i.e. UNDEFINED, TYPE_1 or TYPE_2
EquatorLength	length of lobe equator
BaselineLength	length of lobe baseline
ApicalLength	distance from equator to lobe contour
BasalLength	distance from baseline to equator
TotalLength	total distance from baseline to lobe contour, i.e. sum of apical and basal distances
ApicalContourLong (only TYPE_2 lobes)	length of longer contour segment between contact point and inflection point
ApicalContourShort (only TYPE_2 lobes)	length of shorter contour segment between contact point and inflection point

C

D

E

F

Count 150 0

0.2 0.6 1
BH-adjusted p value

Benjamini-Hochberg-adjusted p values after Dunn's test

	3 DAG	5 DAG	7 DAG	small	medium	large	all

LobeLen/EquatorWidth (TYPE_1)
LobeLen/EquatorWidth (TYPE_2)
ACShort/ACLong (TYPE_2)

Figure S6. Classification and Quantification of Type I and Type II Lobe Characteristics.

(A) Total number of lobes at two-cell (type I), three-cell (type II) and undefined contact points detected in the three size categories (small, medium and large) and at the three developmental time points (3 DAG, 5 DAG, 7 DAG). For information on the total number of cells per category and time point, and on filtering thresholds, see Fig. 5. **(B)** Overview of lobe features extracted for each individual lobe of the input data set. **(C-E)** Quantification of lobe characteristics within the three data sets (3 DAG, 5 DAG, 7 DAG) and the four size categories (small, medium, large, all). Feature values are shown in box plots. Box plots show medians (horizontal bar), interquartile ranges (IQRs; boxes), and data ranges (whiskers) excluding outliers (defiened as >1.5xIQR). Ratio of lobe length to lobe equator width for type I lobes in **(C)** and type II lobes in **(D)**. **(E)** Ratio of short versus long contour segments in type II lobes. **(F)** Statistical analysis from pairwise comparisons within the three data sets (3 DAG, 5 DAG, 7 DAG) and the four size categories (small, medium, large, all). Shown are Benjamini-Hochberg-adjusted p-values after Dunn's pairwise test. Blue colors represent p-values close to 1, yellow colors represent values close to 0. Stars indicate significant differences (* padj < 0.05, ** padj < 0.01, *** padj < 0.001).

Figure S7

A

B

Count
50
20
0

0.2 0.8

BH-adjusted p value

**Benjamini-Hochberg-adjusted
p values after Dunn's test**

Col-0 vs *ktn1 – 5*	*ktn1 – 5* vs *oxIQD16*	Col-0 vs *oxIQD16*	
***	***		Area
***		***	Perimeter
	***	***	Length
*	***	***	Width
***	**	***	Circularity
*	***	***	Eccentricity
***		***	Solidity
	**	**	ConvexHullArea
			ConvexHullPerimeter
***	***	***	ConvexHullConvexity
***	***		ConvexHullRoundness
***	***	***	MarginRoughness
			AvgConcavity
	*	**	StdDevConcavity
*			LongestPathLength
***	***	***	BranchCount
***		***	AvgBranchLength
***	**	***	AvgEndpointDist
		*	LobeCount
***	***	*	NonLobeArea
***	***	***	AvgLobeLength
***	***	***	AvgApicalLobeLength
***	***	***	AvgBasalLobeLength
	***	***	AvgBasalLobeWidth
	***	**	AvgEquatorLobeWidth
***	***	***	MinCoreWidth
***	***	***	MaxCoreWidth

Figure S7. Quantification of Pavement Cell Shape Features in Wild-Type and Mutant Lines.
(A) Violin plots showing the distribution of values for all 27 features calculated from medium and large sized cells of wild-type (Col-0) and *ktn1-5* and *oxIQD16* mutant seedlings. Circles and crosses refer to medians and means, the vertical black lines in each category represent the standard deviation (thick lines) and the 95% confidence intervals (thin lines). The width of each violin box represents the local distribution of feature values along the y axes. P-values are derived from Kruskal-Wallis-Tests, and are given for each feature. **(B)** Statistical analysis of all 27 features from pairwise comparisons between the three analyzed genotypes (Col-0, *ktn1-5* and *oxIQD16*). Shown are Benjamini-Hochberg-adjusted p-values after Dunn's pairwise test. Blue colors represent p-values close to 1, yellow colors represent values close to 0. Stars indicate statistically significant differences (* padj < 0.05, ** padj < 0.01, *** padj < 0.001).

Figure S8
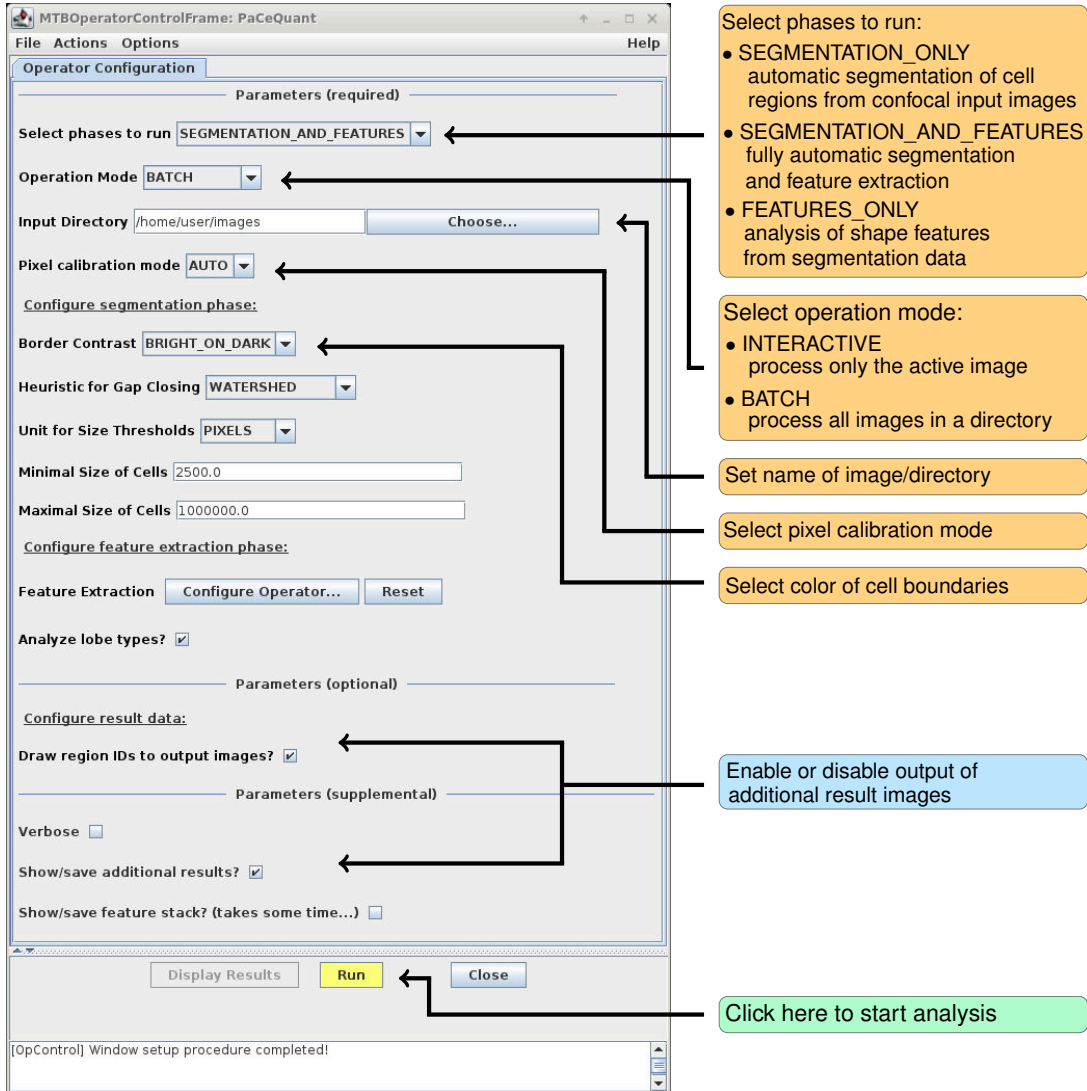
Figure S8. Graphical User Interface of PaCeQuant.
PaCeQuant is available as part of the Microscope Image Analysis Toolbox MiToBo (http://www.informatik.uni-halle.de/mitobo, version ≥ 1.8.6). For integration in ImageJ/Fiji activate MiToBo's update site in Fiji. For more information on Fiji update sites see http://imagej.net/How_to_follow_a_3rd_party_update_site. After installation, PaCeQuant is included in the MiToBo plugins menu in Fiji. Selecting PaCeQuant opens the graphical user interface, which allows to configure and run PaCeQuant. The user just needs to specify the phases to run (segmentation and/or feature extraction), set the input data format, and choose between processing single images interactively or all images in a directory in batch mode as shown above. All other parameter settings are optional and we recommend to leave them unchanged for easing comparative evaluations. An extended instruction manual explaining all parameters in detail can be found on the PaCeQuant website http://mitobo.informatik.uni-halle.de/index.php/Applications/PaCeQuant.

Table S1: Detailed Definitions of Features Analyzed by PaCeQuant. PaCeQuant extracts 27 individual features per cell that are classified into four general feature groups: (A) global features (11), (B) contour-based features (3), (C) skeleton-based features (4), and (D) pavement cell-specific features (9). Shown are the features and the definitions for their calculation.

Group A – Global Features
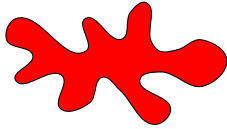
1. Area $A(R)$	The area $A(R)$ of a cell region R is defined as the total number of pixels belonging to the cell region, measured in μm^2 according to the physical resolution of the image.
2. Perimeter $P(R)$	The perimeter $P(R)$ of a cell region R is calculated by tracing the contour pixels of a region (assuming 8-neighborhood) and summing up distances between all pairs of subsequent pixels measured in μm.
3. Length $L(R)$ **and** **4. Width** $W(R)$	The length and width of a cell region R are defined as follows: $$L(R) = 2 \cdot \sqrt{\frac{2 \cdot \left(\mu_{2,0} + \mu_{0,2} + \sqrt{(\mu_{2,0} - \mu_{0,2})^2 + 4\mu_{1,1}^2}\right)}{\mu_{0,0}}}$$ $$W(R) = 2 \cdot \sqrt{\frac{2 \cdot \left(\mu_{2,0} + \mu_{0,2} - \sqrt{(\mu_{2,0} - \mu_{0,2})^2 + 4\mu_{1,1}^2}\right)}{\mu_{0,0}}}$$ $\mu_{k,l}$ denotes the central moment of order (k, l) of the set of region pixels: $$\mu_{k,l} = \sum_{(x,y) \in R} (x - \bar{x})^k (y - \bar{y})^l$$ with $(\bar{x}, \bar{y})$ being the center of mass of the region. The length and width refer to the major and minor axes of an ellipse fitted to the region pixels. They are measured in μm.
5. Circularity $C(R)$	The circularity $C(R)$ of a cell region R is defined as $$C(R) = \frac{4 \cdot \pi \cdot A(R)}{P(R) \cdot P(R)}$$ where $A(R)$ is the area and $P(R)$ is the perimeter of the region. A circularity value of 1 indicates that the shape is perfectly circular. Decreasing values towards 0 indicate increasing deviation from a circular shape.
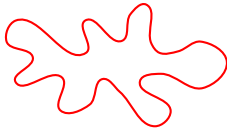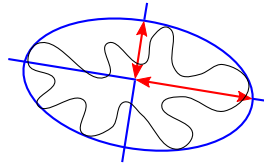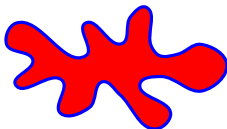
1

Group A – Global Features – Continued

|	
6. Eccentricity $E(R)$	The eccentricity $E(R)$ of a cell region R is defined as follows: $$E(R) \;=\; \frac{(\mu_{2,0} - \mu_{0,2})^2 + 4 \cdot \mu_{1,1}^2}{(\mu_{2,0} + \mu_{0,2})^2} \;.$$ $\mu_{k,l}$ denotes the central moment of order (k,l) of the region (see also features Length and Width). The eccentricity quantifies the elongation of a region. An almost circular region has an eccentricity value of 0.
7. Area of convex hull $A_{CH}(R)$	The convex hull is defined as the convex polygon with the smallest area including the complete region. Hence, the convex hull of a region is always larger than the region itself. The area $A_{CH}(R)$ of the convex hull is defined as the total number of pixels of the convex hull measured in μm^2.
8. Perimeter of the convex hull $P_{CH}(R)$	The perimeter $P_{CH}(R)$ of the convex hull is defined analogously to the perimeter $P(R)$ of the region itself, i.e. the contour of the hull is traced and distances between subsequent pixels are summed up. It is measured in μm.
9. Roundness $round(R)$	The roundness $round(R)$ of a cell region R is defined similar to the region circularity $C(R)$, however, replacing the perimeter $P(R)$ of the region by the perimeter $P_{CH}(R)$ of the convex hull (see also Wu et al., (2016)): $$round(R) \;=\; \frac{4 \cdot \pi \cdot A(R)}{P_{CH}(R) \cdot P_{CH}(R)}$$
10. Convexity $conv(R)$	The convexity $conv(R)$ of a cell region R is defined as the ratio of the perimeter of the convex hull and the perimeter of the region (see also Wu et al., (2016)): $$conv(R) \;=\; \frac{P_{CH}(R)}{P(R)}$$
11. Solidity $S(R)$	The solidity $S(R)$ calculates the ratio of the region area $A(R)$ and the area of the region's convex hull: $$S(R) \;=\; \frac{A(R)}{A_{CH}(R)}$$ The value of solidity decreases with decreasing convexity of the region.
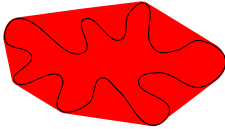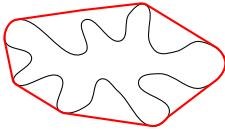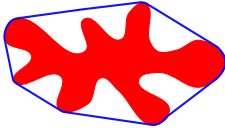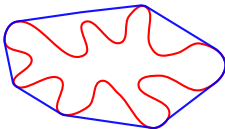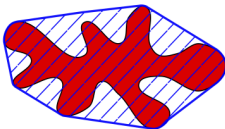
2

Group B – Contour-based Features
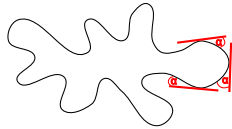
|	
12. Margin roughness $MR(R)$	The margin roughness $MR(R)$ of a cell region R is defined as follows, $$MR(R) \;=\; \left(\frac{1}{N} \sum_{i=1}^{N} |A_i| \right) - \frac{360}{N} \;,$$ where N is the total number of contour points of the region. A_i denotes an estimate of the local curvature at contour pixel position i (for details cf. features 13 and 14 below). To quantify margin roughness first for each contour pixel the local curvature is estimated and all estimates are averaged. Subsequently the margin roughness is defined as the deviation of the resulting average value from the expected average angle of a perfect circle sampled by N discrete points (McLellan et al., 1998).
13. Average local contour concavity and 14. Standard deviation of contour concavity	The local contour concavity is calculated by applying the algorithm of Fernàndez et al., (1995) which measures angles between tangents of neighboring points along the contour. The average concavity and its standard deviation is given by calculating the mean and standard deviation of all local concavities along the contour.

Group C – Skeleton-based Features
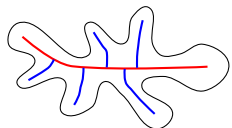
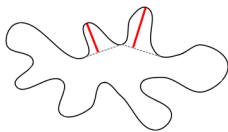|	
15. Longest skeleton path length	A region skeleton consists of end-points, branch points and pixels in between. End-points are defined as pixels having only a single neighbor in the skeleton, while branch points have more than two neighbors. A path in the skeleton connects two end-points. Its length is given by the sum of all distances between neighboring pixels along the path. Here the length of the longest path in a skeleton is measured in µm.
16. Branch count and 17. Average branch length	A branch in a skeleton consists of a set of pixels connecting an end-point of a skeleton with the closest branch-point. The branch count specifies the total number of branches in a skeleton, and the average branch length quantifies the average length of the branches measured in µm.
18. Average (branch) end-point distance	The number of branches yields a rough estimation for the number of lobes in a cell region, but usually underestimates the real count. The radius of each of these lobes can be approximated by the distance of the end-point of the corresponding branch to the closest background pixel. The average lobe radius given in µm is calculated over the radii of all branches of a given skeleton.
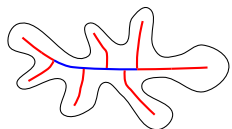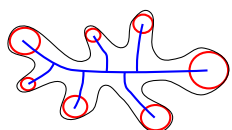
3

Group D – Pavement Cell-Specific Features
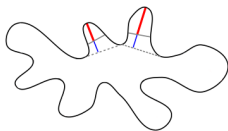
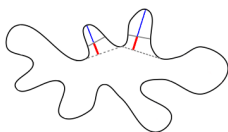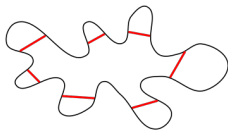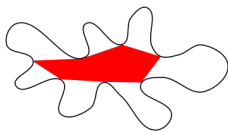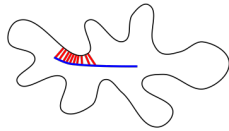|	
19. Lobe count	This features counts the total number of lobes along a cell contour. To detect these lobes we basically rely on estimates of local contour curvature calculated using the algorithm of Freeman and Davis (Freeman et al., 1977). For details of the calculation refer to Section "Pavement Cell Features".
20. Average lobe length	To calculate the length of a lobe first the distances of all pixels of the lobe to the lobe baseline are computed. The maximum of these distances defines the overall length of the lobe. The average lobe length of a cell is the mean value of all these maximal distances of all lobes of the cell measured in μm.
21. Average apical lobe length	The length of a lobe as defined in Feature 20, i.e. the distance between contour and baseline, is separated into two parts by the equator of the lobe. With the average apical length the mean distance between the contour points and the equators for all lobes of a cell is measured in μm.
22. Average basal lobe length	Complementary to Feature 21 the basal length of a lobe measures the distance between the baseline and the equator of the lobe. This feature quantifies the average basal lobe length of all lobes of a cell in μm.
23. Average basal lobe width	The basal lobe width measures the length of the baseline of a lobe, and this feature extracts the average baseline length of all lobes of a cell in μm.
24. Average equator lobe width	Each lobe contour subsumes two inflection points where the sign of the curvature changes. The equator of a lobe, i.e. the line connecting these two points, is usually located at half-height of the lobe and yields an approximation for the vertical extension of a lobe. The average equator lobe width is calculated as the mean of all equator lengths of all lobes and is measured in μm.
25. Non-lobe area	The non-lobe part of a cell region is defined as the part of the region which does not belong to any lobe. The size of this region yields an estimate for the degree of lobing in the cell. Smaller non-lobe areas refer to an increased number and length of lobes in a cell. The non-lobe area is measured in μm^2.
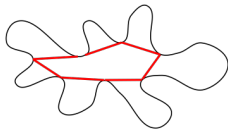
4

Group D – Pavement Cell-Specific Features

26. Minimal core width and 27. Maximal core width 	As an estimate of the neck width of a region the core of a cell region is analyzed for narrow and wide sections. The core part of a cell region is roughly given by the part of the cell region represented by all non-branch parts of the region skeleton. For pixels belonging to these parts their distances to the background are calculated and sorted. The minimal core width is then given by twice the 1st quartile value and the maximal core width is given by twice the 3rd quartile value of the sorted distance list measured in µm.

Supplemental References

Fernàndez G, Kunt M, Zrÿd J-P (1995) A new plant cell image segmentation algorithm. Springer Berlin Heidelberg: 229-234

Freeman H, Davis LS (1977) A corner-finding algorithm for chain-coded curves. IEEE Trans. Comput. **26**: 297-303

McLellan T, Endler J (1998) The relative success of some methods for measuring and describing the shape of complex objects. Syst. Biol. **47**: 264-281

Wu T-C, Belteton S, Pack J, Szymanski DB, Umulis D (2016) LobeFinder: a convex hull-based method for quantitative boundary analyses of lobed plant cells. Plant Physiol. **171**: 2331-2342

5

Table S2: Quantitative Evaluation of Segmentation Quality. For the 15 cells of the test dataset a quantitative comparison of PaCeQuant's automatic segmentation results with manual segmentation by a human expert was performed. The table shows Hausdorff distances for differences between contours, and recalls, precisions and Dice coefficients for deviations in cell areas for each of the 15 cells.

Segmentation Quality Evaluation Measures				
Cell-ID	Hausdorff	Recall	Precision	Dice index
1	3.606	0.980	0.974	0.977
2	11.705	0.980	0.985	0.983
3	4.243	0.984	0.980	0.982
4	4.000	0.985	0.976	0.980
5	5.000	0.969	0.985	0.977
6	6.000	0.983	0.977	0.980
7	4.000	0.959	0.960	0.960
8	6.403	0.965	0.955	0.960
9	5.000	0.983	0.968	0.976
10	44.181	0.984	0.943	0.963
11	3.606	0.983	0.979	0.981
12	6.000	0.963	0.978	0.970
13	4.243	0.980	0.980	0.980
14	4.123	0.971	0.985	0.978
15	12.166	0.970	0.972	0.971
Min	3.606	0.959	0.943	0.960
Max	44.181	0.985	0.985	0.983
Average	8.574	0.976	0.973	0.974
Median	5.000	0.980	0.977	0.977

1

B.3 Microtubule-associated protein IQ67 DOMAIN5 regulates morphogenesis of leaf pavement cells in *Arabidopsis thaliana* (Mitra et al., Experimental Botany, 2019)
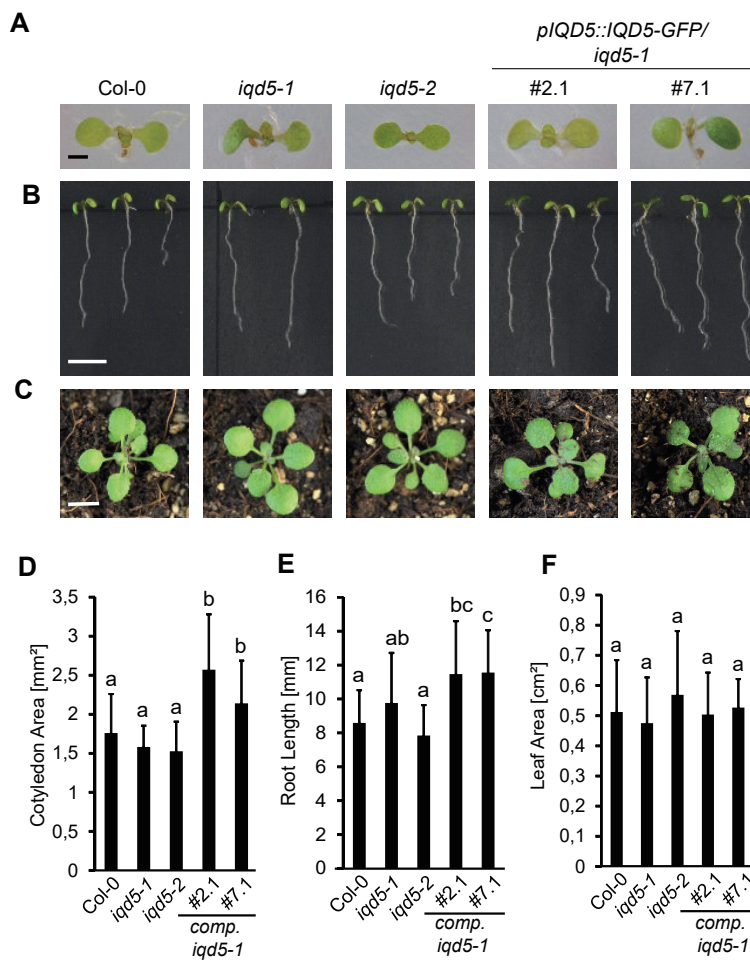
Supplementary Fig. S1.
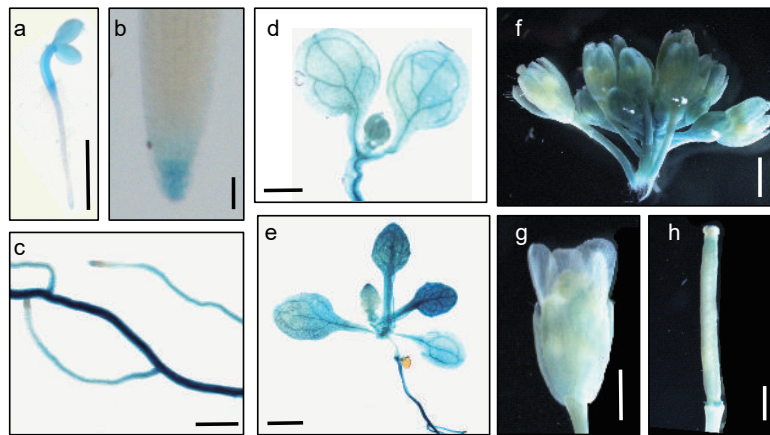
Supplementary Fig. S1.
IQD5 expression analysis in *pIQD5$_{long}$::GFP-GUS* reporter lines. Whole mount histochemical GUS staining of 2-day-old seedlings (a), in the primary root meristem (b), lateral roots (c) and cotyledons (d) of 5-day-old seedlings, in the shoot of 10-day-old seedlings (e), and in flower buds f), flowers (g) and siliques (h) of 5-week-old plants. Scale bars represent 1 mm (a, c-h) and 10 μm (b).
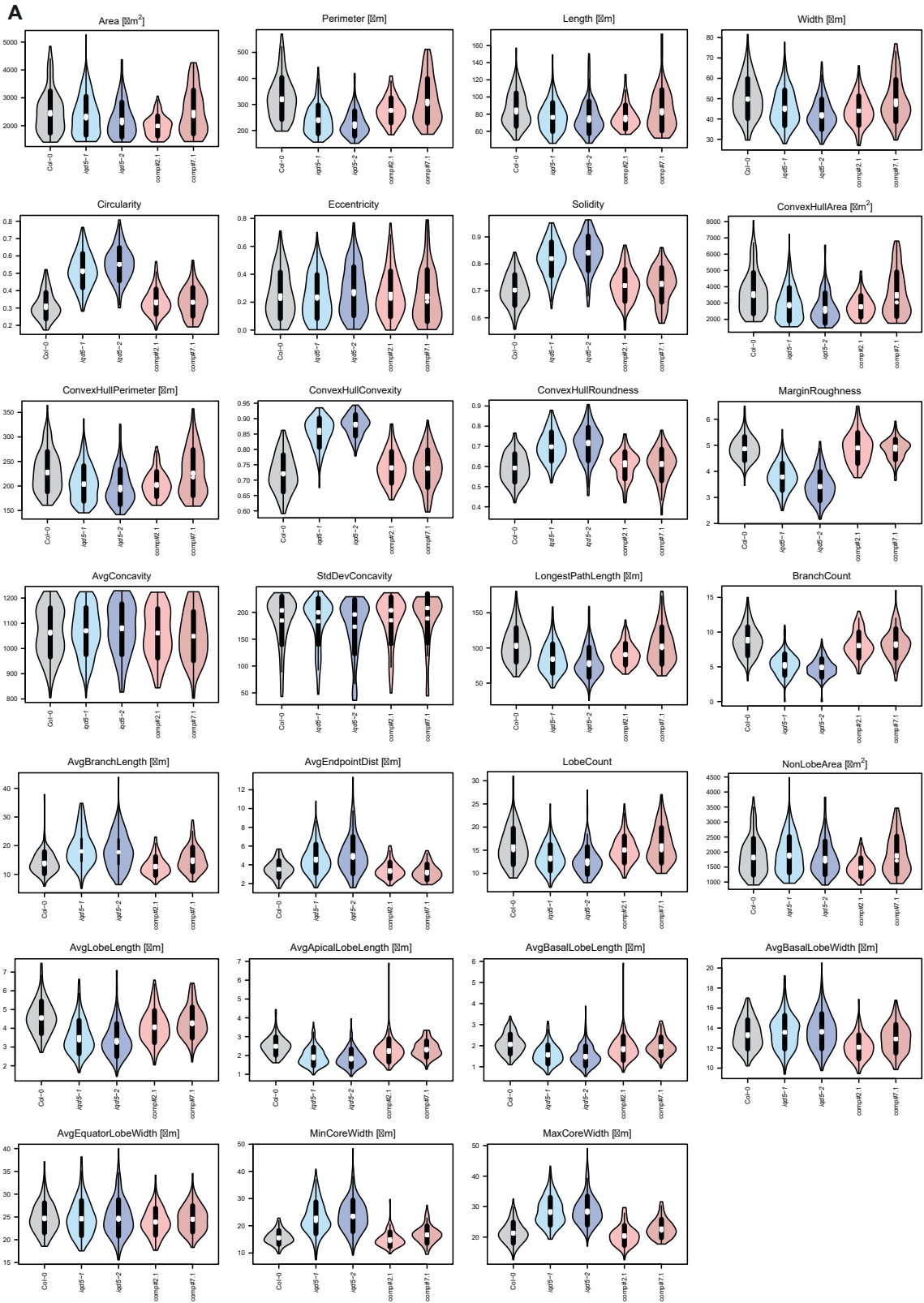
Supplementary Fig. S2.

A

Col-0   *iqd5-1*   *iqd5-2*   *pIQD5::IQD5-GFP/ iqd5-1*   #2.1   #7.1

B

C

D   **E**   **F**
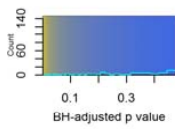
Supplementary Fig. S2.
Macroscopic analysis of growth parameters in wild type and *iqd5* mutants. Seedlings of
wild type, the two *iqd5* mutant alleles *iqd5-1* and *iqd5-2*, and two independent
pIQD5::IQD5-GFP/iqd5-1 complementation lines grown under long-day conditions at
5 days after germination (A, B) and at 3 weeks after germination (C). Surface view of cotyledons
(A) and side view of complete seedlings (B). Surface view of rosettes (C). Bars, 1 mm (A);
1 cm (B, C). Quantification of coteldon area (D) and root length (E) in seedlings shown in
A and B. Quantification of leaf area (F) in seedlings shown in C. Data represent mean values
± standard deviation from n = 25-56 (D), n = 53-72 (E) and n = 30 (F) seedlings or plants.
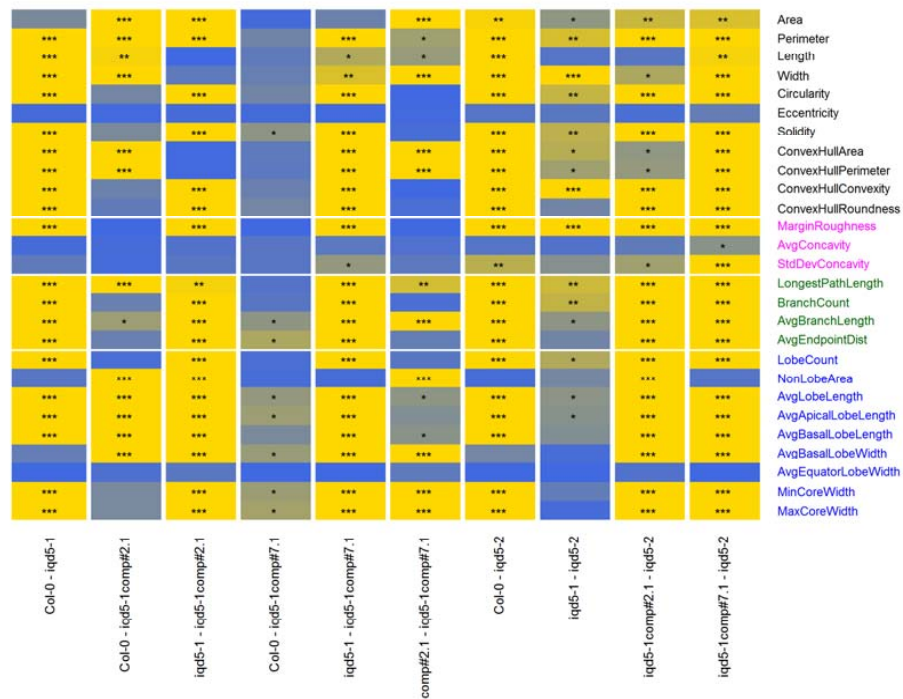Different letters indicate statistically significant differences by one-way ANOVA; $p < 0.01$.

Fig. S3.

A
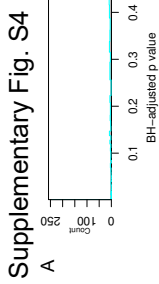
B

Supplementary Fig. S3.
Quantification and statistical analysis of PC shape features in 5-day-old seedlings of the wild type and *iqd5* mutants. Quantitative analysis of pavement cell shape features in 5-day-old seedlings of the wild type, *iqd5-1*, *iqd5-2* and two independent *pIQD5::IQD5-GFP/iqd5-1* complementation lines. Violin plots of all 27 features quantified with PaCeQuant (A). Statistical analysis from pairwise comparison between the analyzed genotypes (B). Shown are Benjamini-Hochberg-adjusted p-values after Dunn's pairwise test. Blue colors represent p-values close to 1, yellow colors represent values close to 0. Stars indicate statistically significant differences (* padj < 0.05, ** padj < 0.01, *** padj < 0.005).
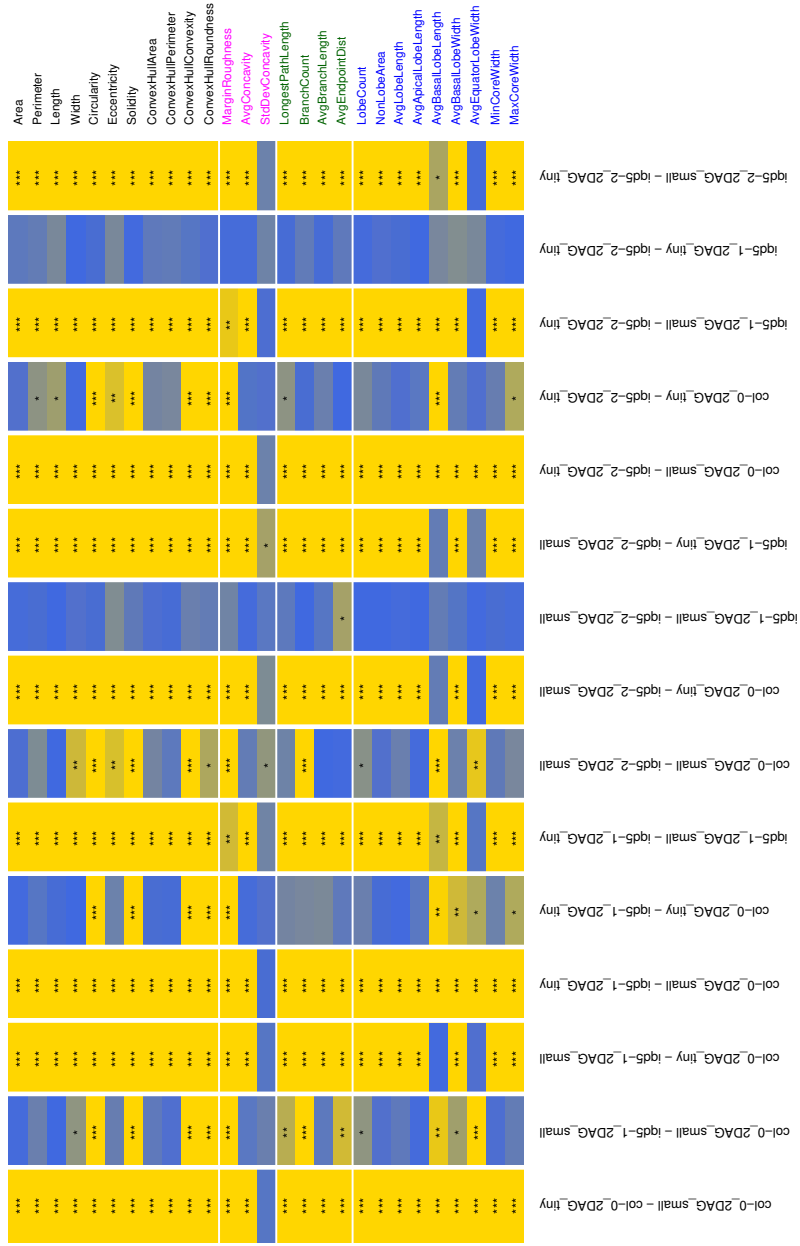
Supplementary Fig. S4
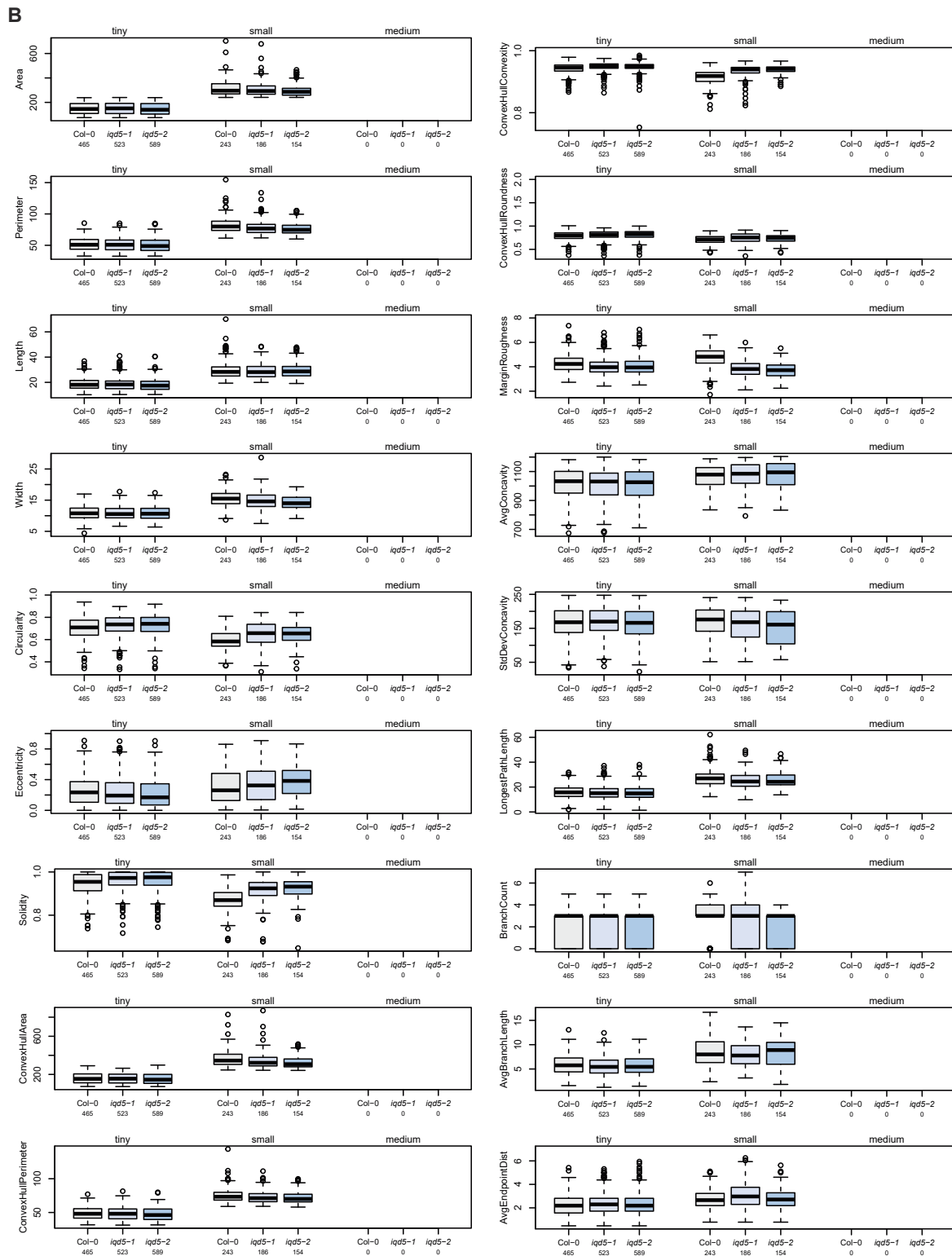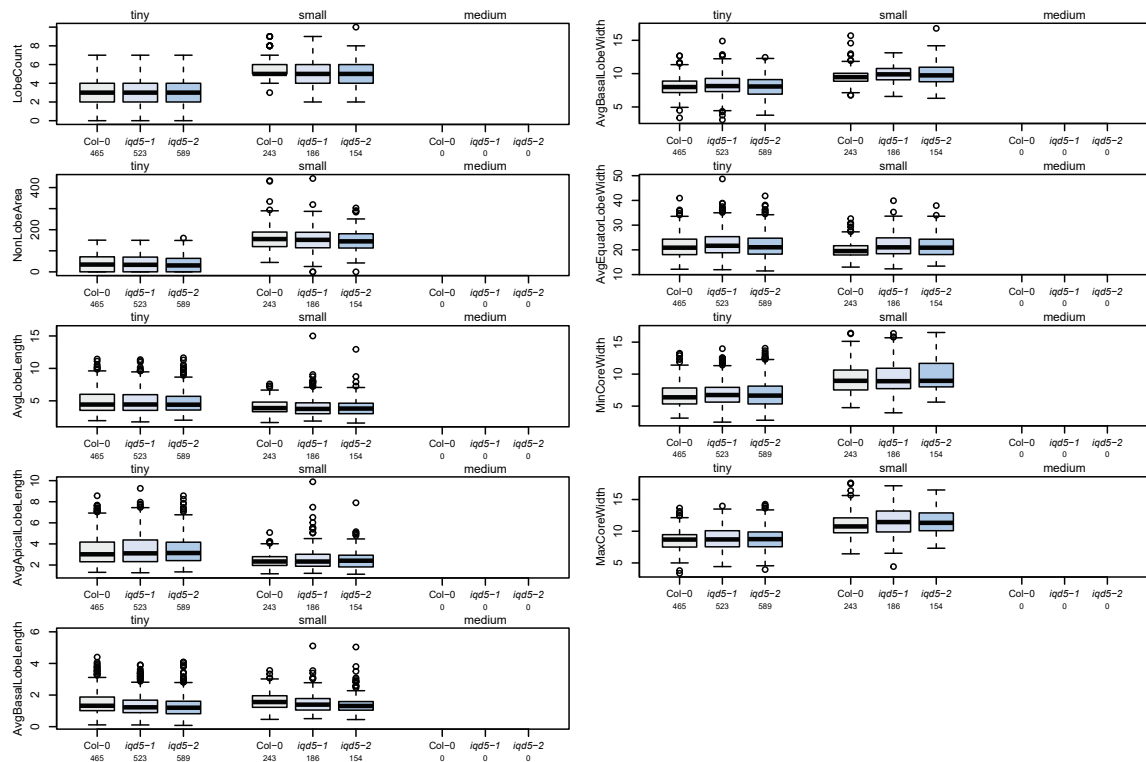
Supplementary Fig. S4.

Quantification and statistical analysis of PC shape in cotyledons at 2 DAG. Cells were grouped into tiny ($t_{tiny} \leq 240$ μm²) and small ($t_s \leq 1,400$ μm²) sized populations. Statistical analysis from pairwise comparisons within the three genotypes (Col-0, *iqd5-1* and *iqd5-2*) and the two size categories (tiny, small) (A). Quantification of shape features (B). Results are medians, boxes range from first to third quartile.
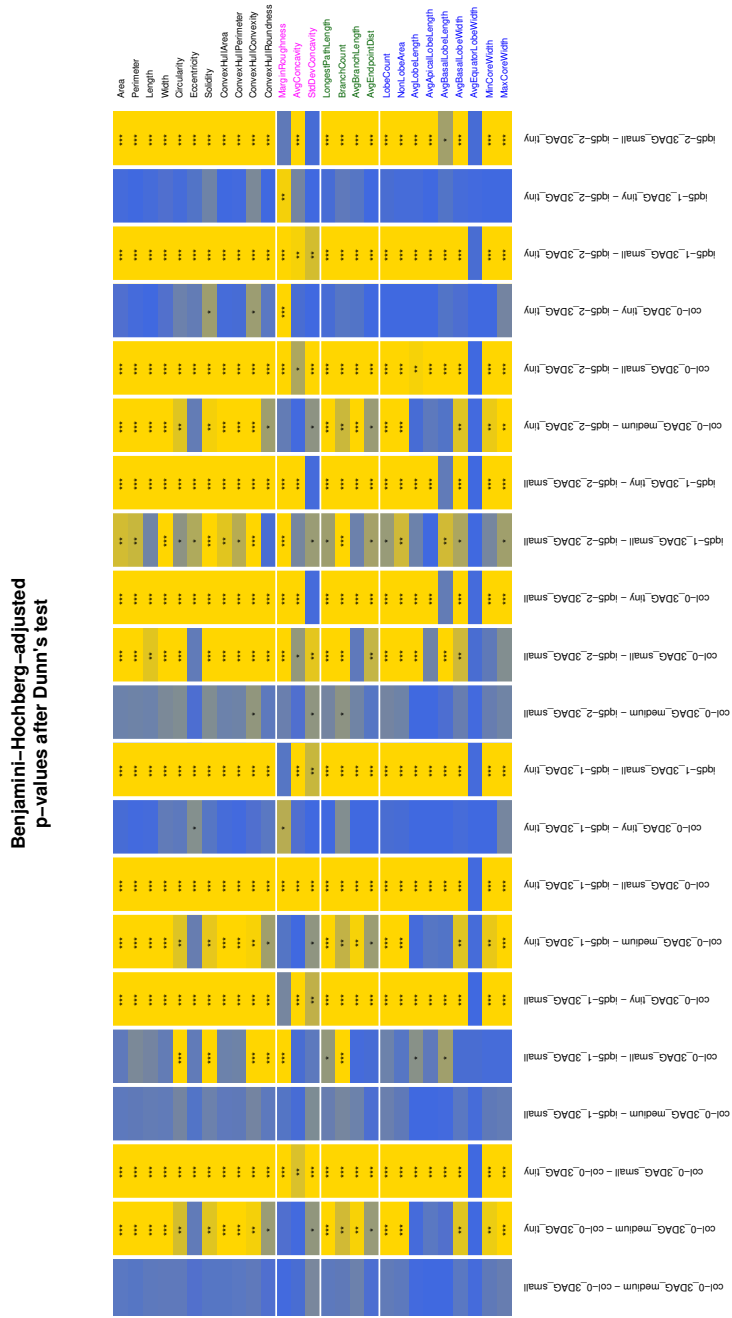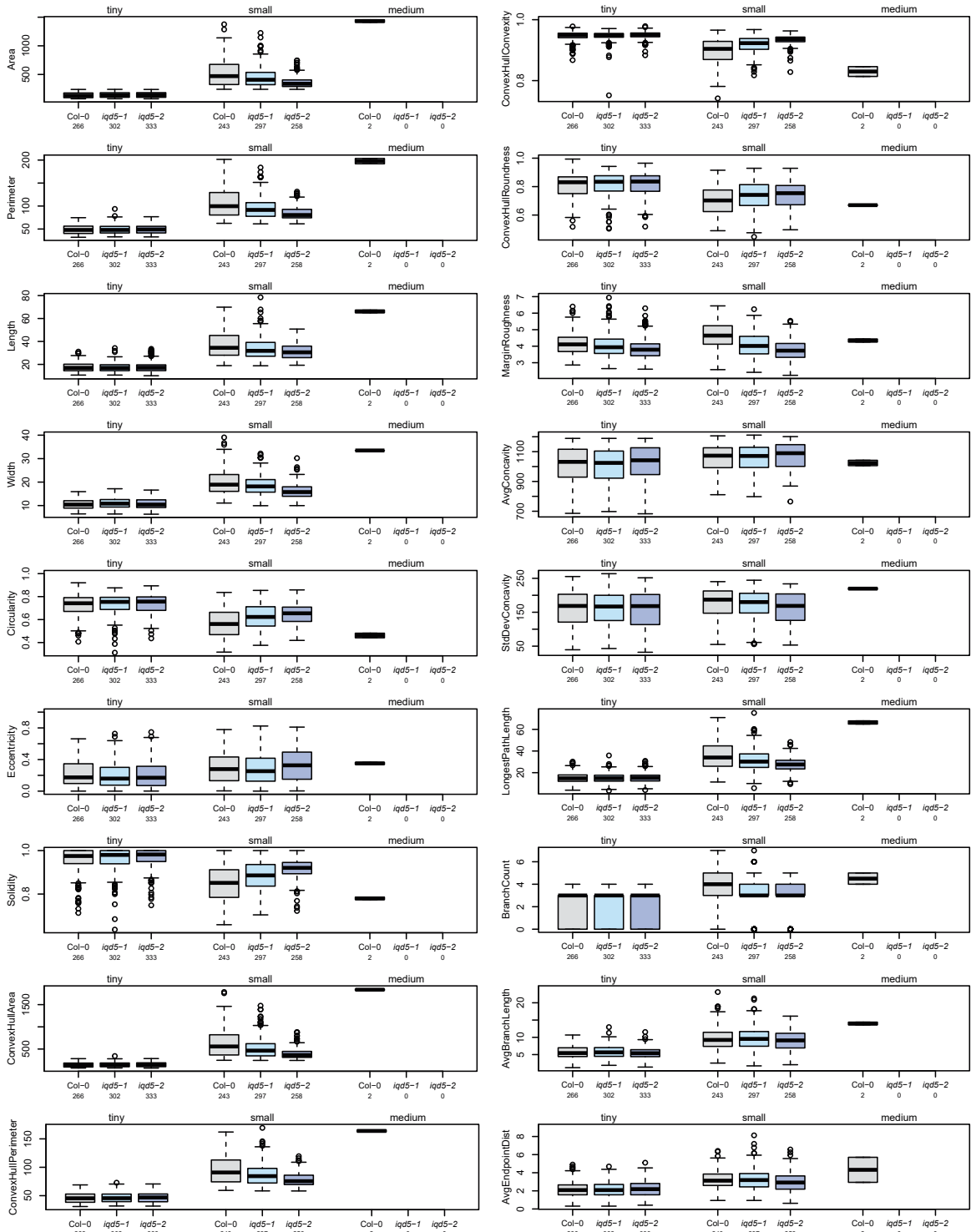
Supplementary Fig. S5

B

Supplementary Fig. S5.
Quantification and statistical analysis of PC shape in cotyledons at 3 DAG. Cells were grouped into tiny
($t_{tiny} \leq 240\ \mu m^2$), small ($t_s \leq 1,400\ \mu m^2$) and medium ($t_m \leq 4,042\ \mu m^2$) sized populations. Statistical analysis
from pairwise comparisons within the three genotypes (Col-0, *iqd5-1* and *iqd5-2*) and the three size categories
(tiny, small, medium) (A). Note that only two medium-sized cells were detected in Col-0, and no medium-sized
cells were present in the two *iqd5* mutants.Quantification of shape features (B). Results are medians, boxes
range from first to third quartile.

Supplementary Fig. S6

B

Supplementary Fig. S6.

Quantification and statistical analysis of PC shape in cotyledons at 5 DAG. Cells were grouped into small ($t_s \le 1{,}400\ \mu m^2$), medium ($t_m \le 4{,}042\ \mu m^2$) and large ($t_l > 4{,}042\ \mu m^2$) sized populations. Statistical analysis from pairwise comparisons within the three genotypes (Col-0, *iqd5-1* and *iqd5-2*) and the three size categories (small, medium, large) (A). Quantification of shape features (B). Results are medians, boxes range from first to third quartile.
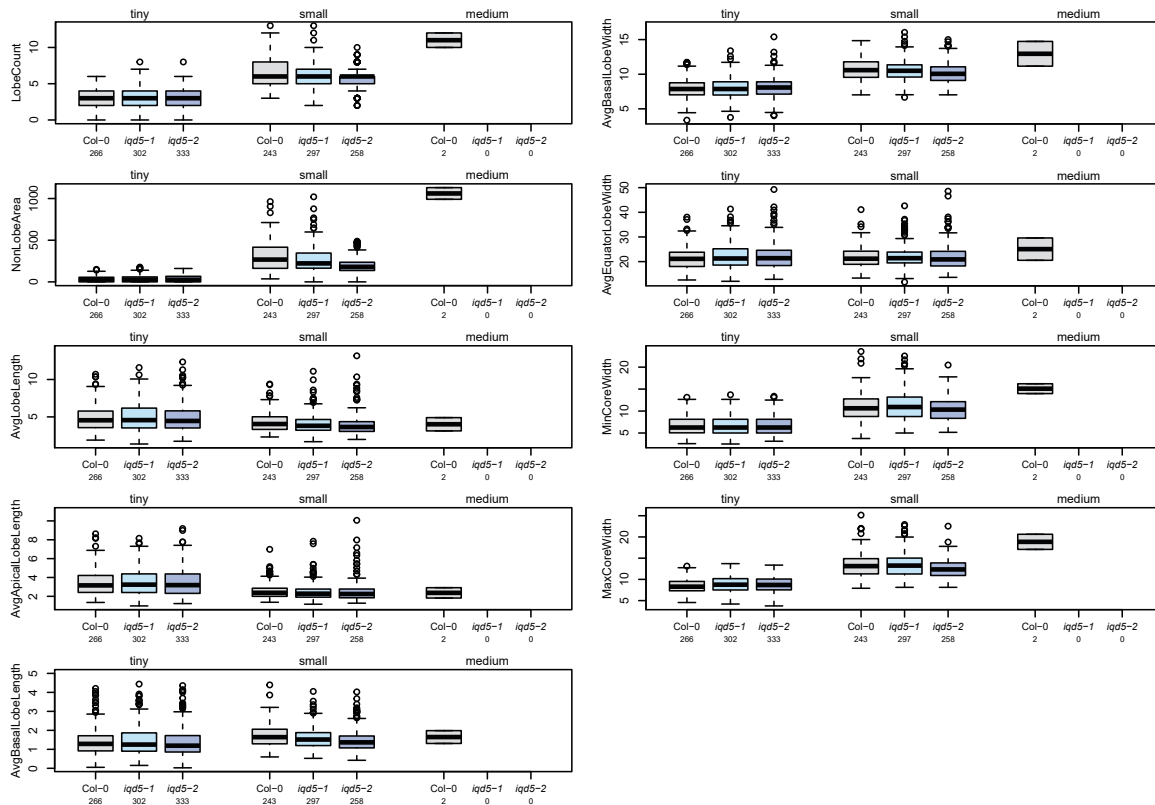
Supplementary Fig. S7

B
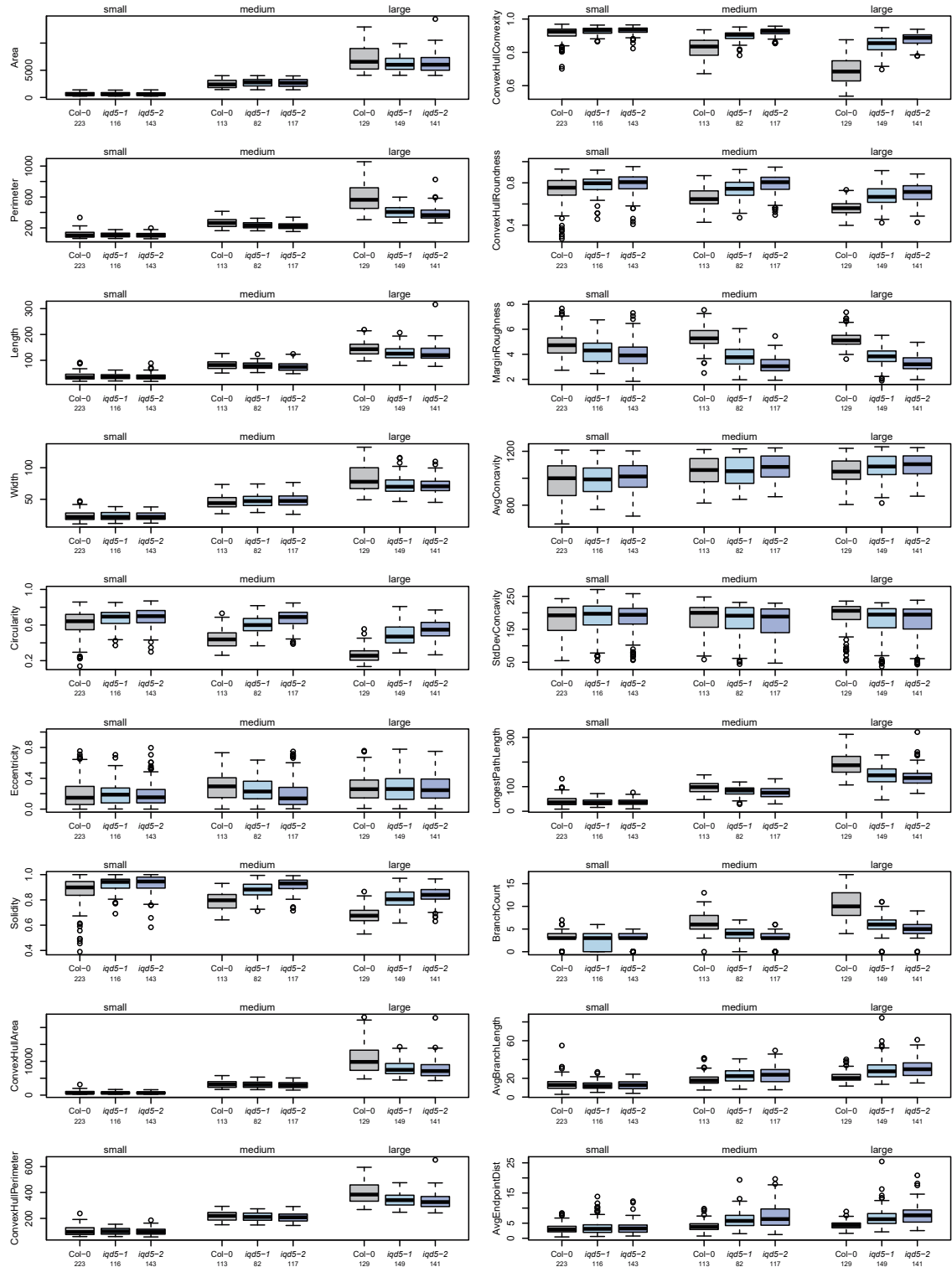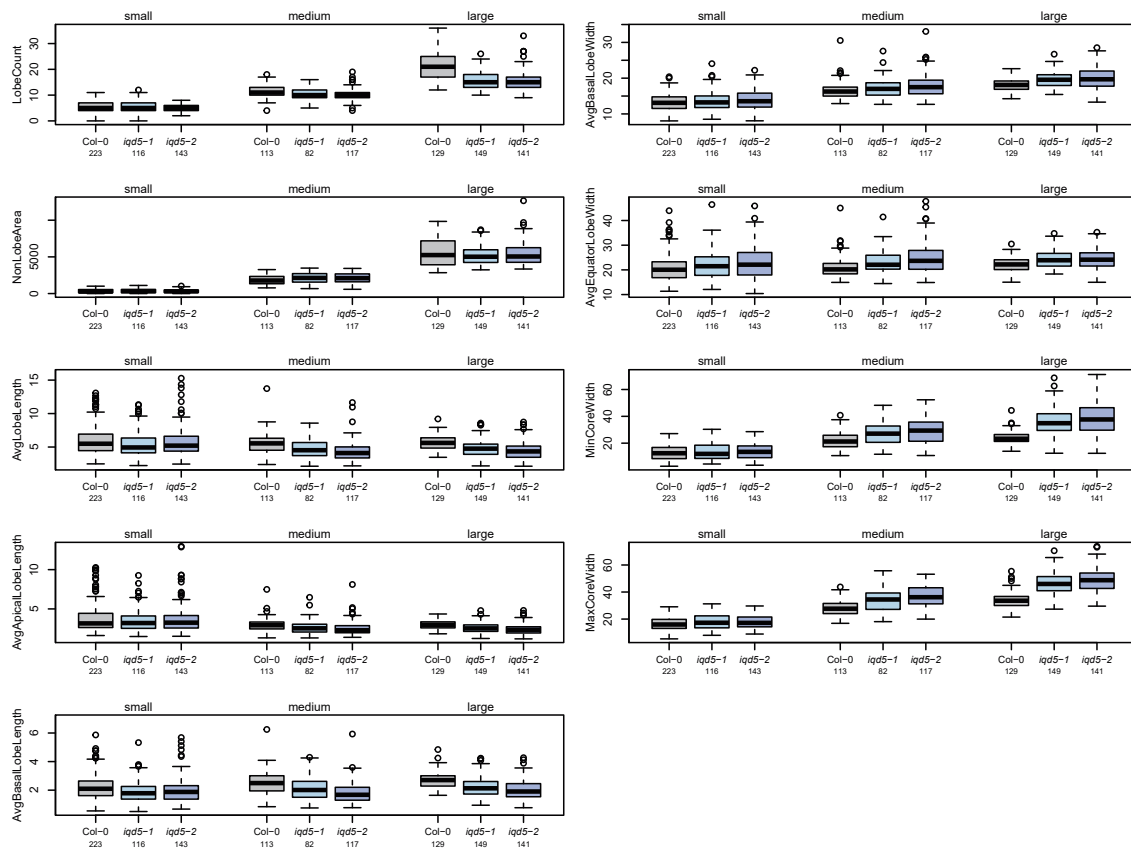
Supplementary Fig. S7.
Quantification and statistical analysis of PC shape in cotyledons at 7 DAG. Cells were grouped into small ($t_s \leq 1,400$ µm²), medium ($t_m \leq 4,042$ µm²) and large ($t_l > 4,042$ µm²) sized populations. Statistical analysis from pairwise comparisons within the three genotypes (Col-0, *iqd5-1* and *iqd5-2*) and the three size categories (small, medium, large) (A). Quantification of shape features (B). Results are medians, boxes range from first to third quartile.

Supplementary Fig. S8

B

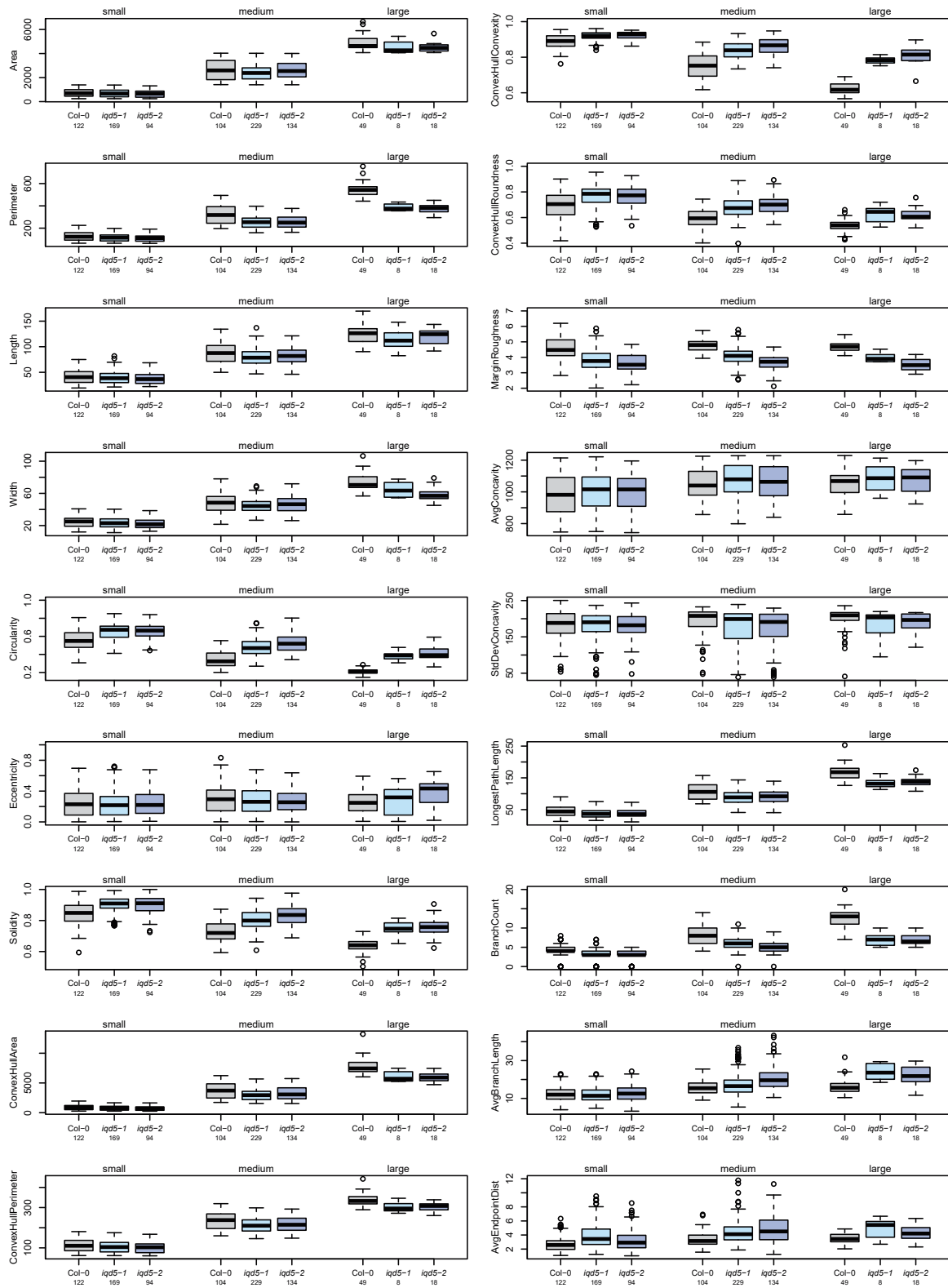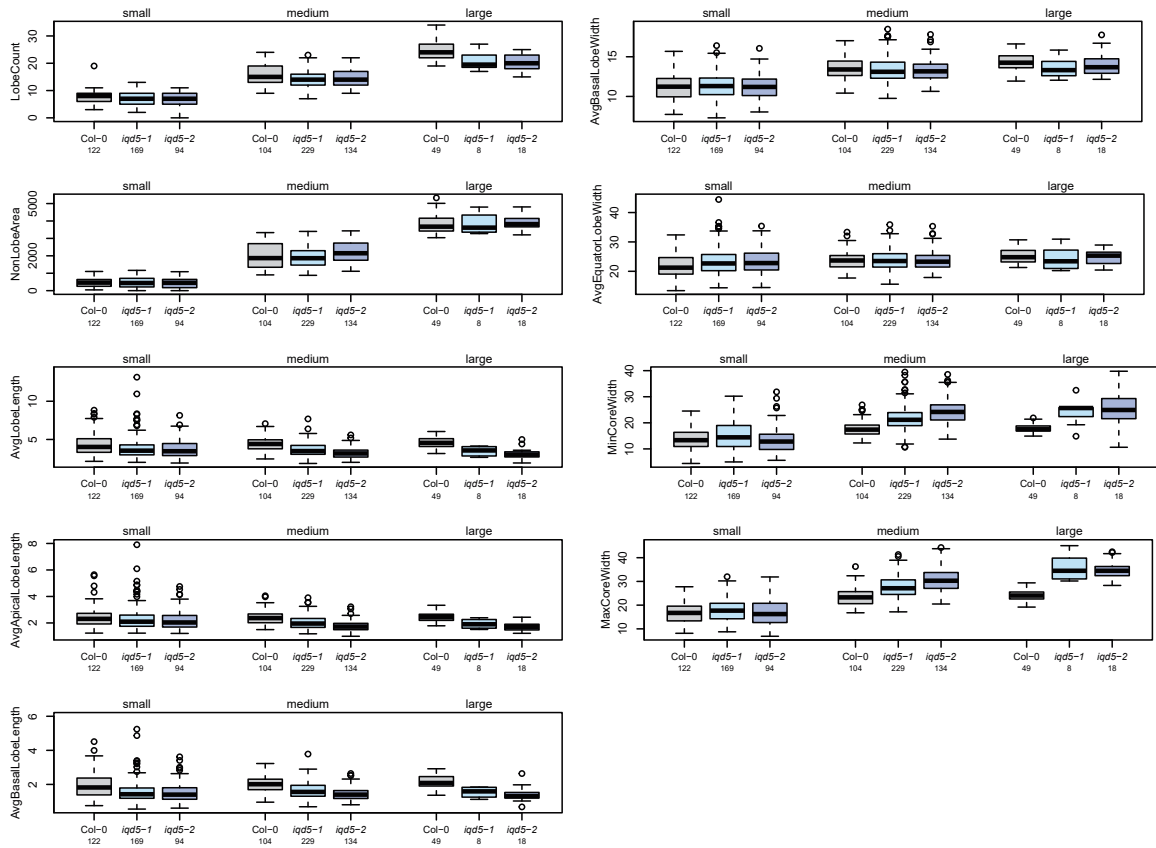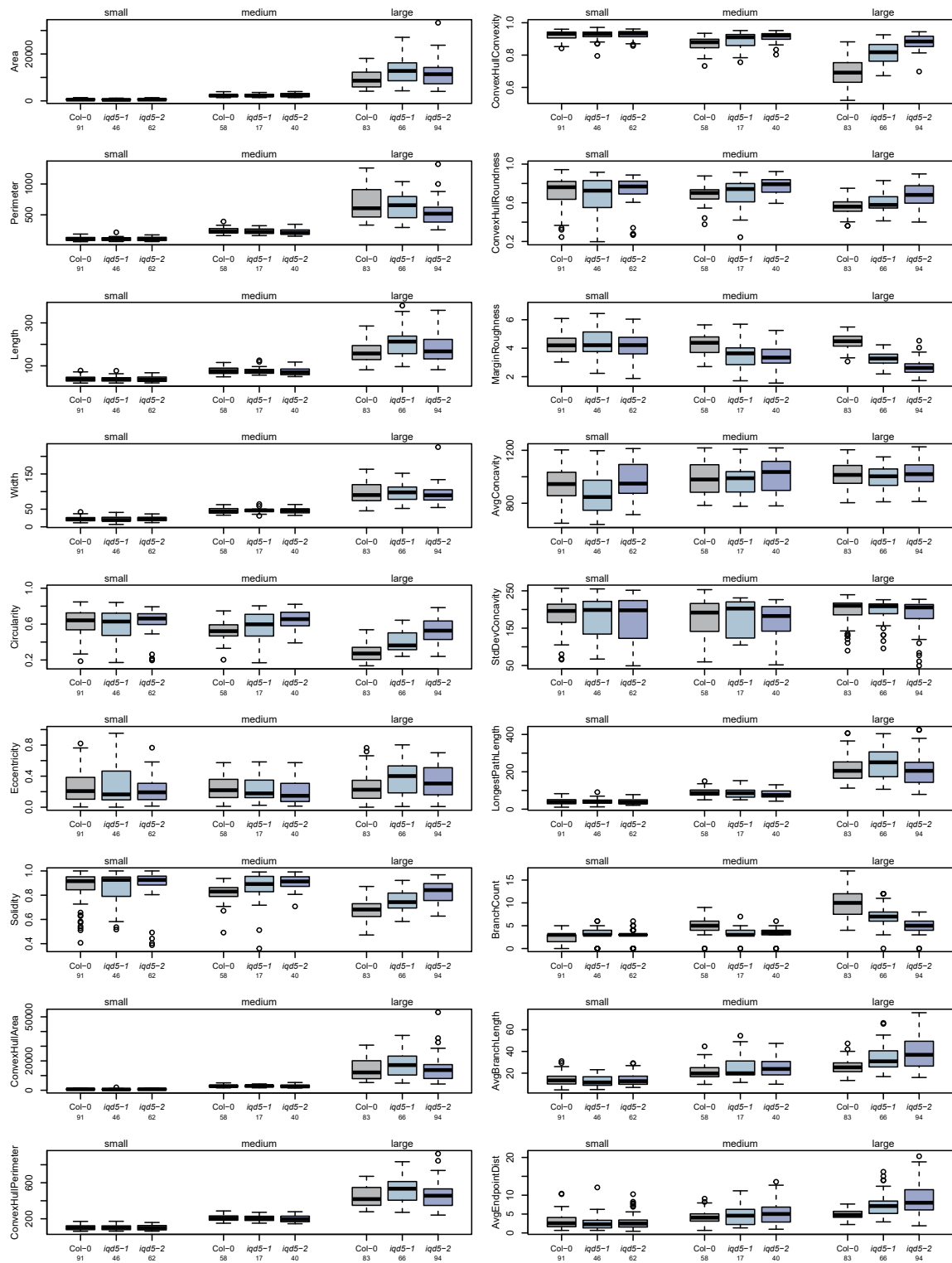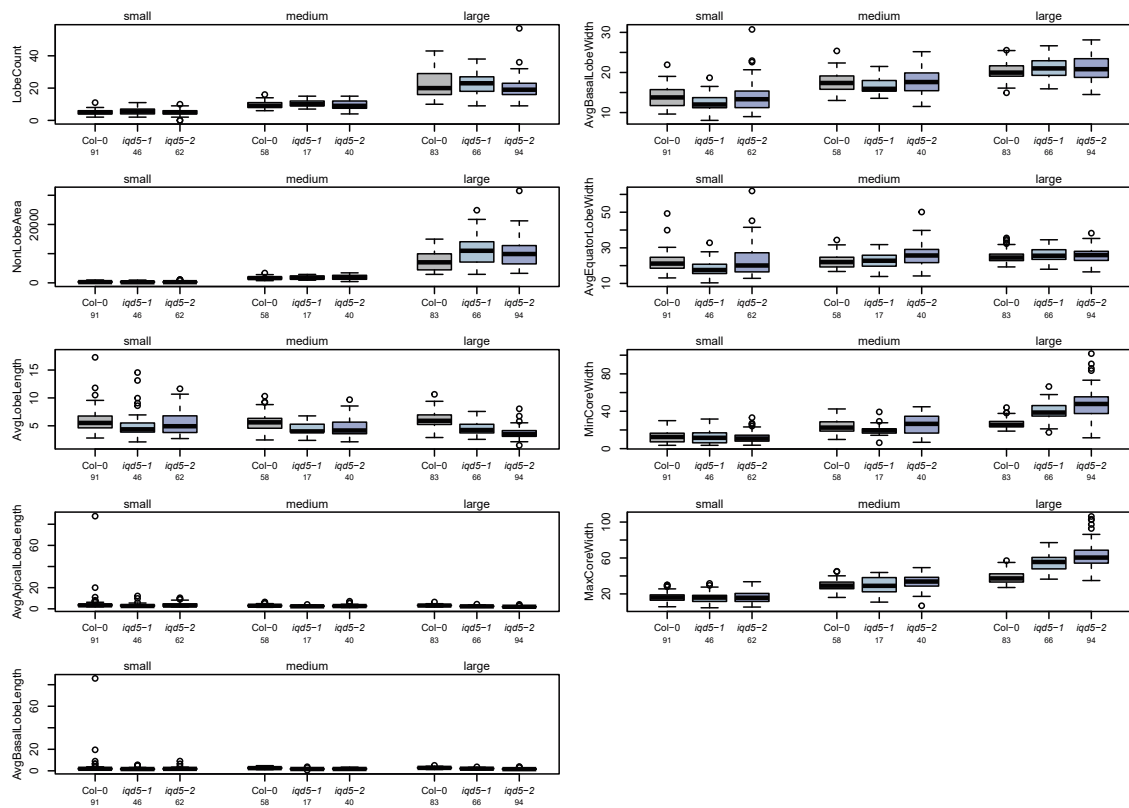**Supplementary Fig. S8.**
Quantification and statistical analysis of PC shape in cotyledons at 10 DAG. Cells were grouped into small
($t_s \leq 1,400\ \mu m^2$), medium ($t_m \leq 4,042\ \mu m^2$) and large ($t_l > 4,042\ \mu m^2$) sized populations. Statistical analysis
from pairwise comparisons within the three genotypes (Col-0, *iqd5-1* and *iqd5-2*) and the three size categories
(small, medium, large) (A). Quantification of shape features (B). Results are medians, boxes range from
first to third quartile.

Fig. S9.

A

B
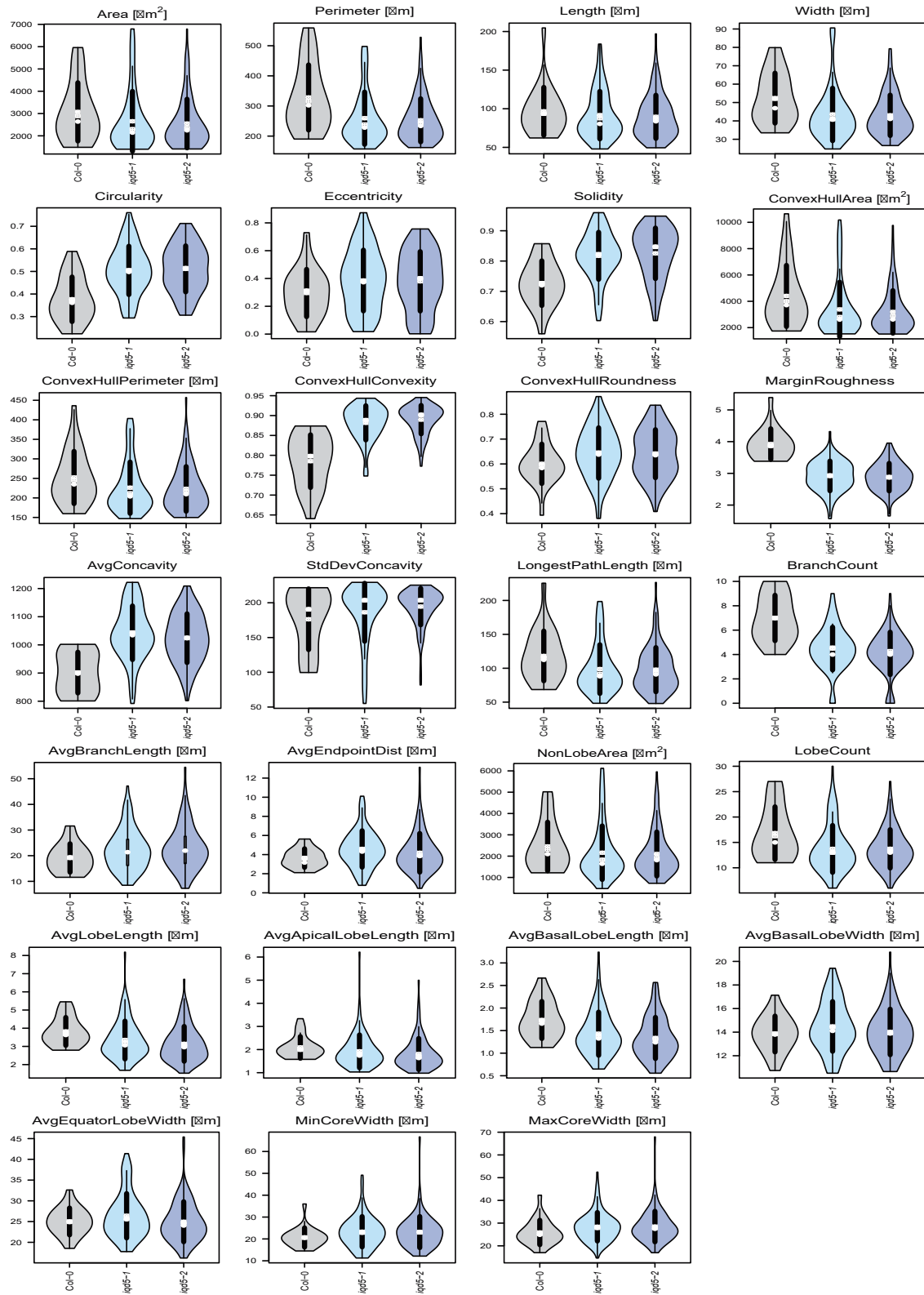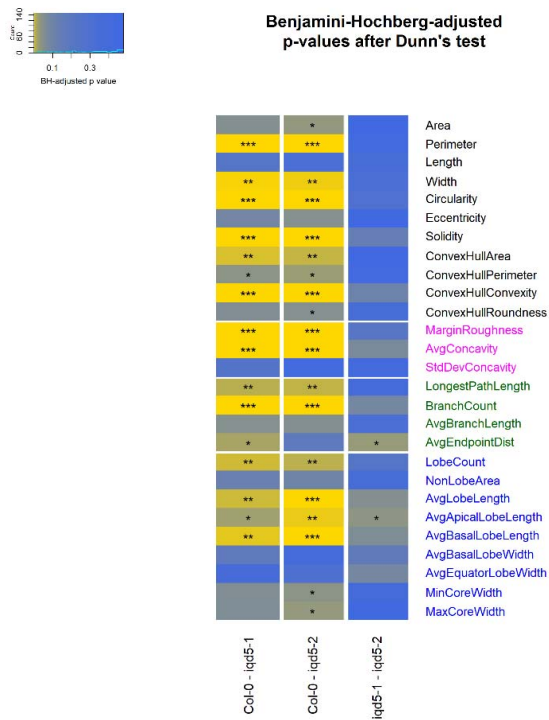
Supplementary Fig. S9.
Quantification and statistical analysis of PC shape in true leaves. PC shape features in the third and fifth true leaves of wild type and *iqd5* mutants. Violin plots of all 27 features quantified with PaCeQuant (A). Statistical analysis from pairwise comparison between the analyzed genotypes (B). Shown are Benjamini-Hochberg-adjusted p-values after Dunn's pairwise test. Blue colors represent p-values close to 1, yellow colors represent values close to 0. Stars indicate statistically significant differences (* padj < 0.05, ** padj < 0.01, *** padj < 0.005).

Supplementary Fig. S10.
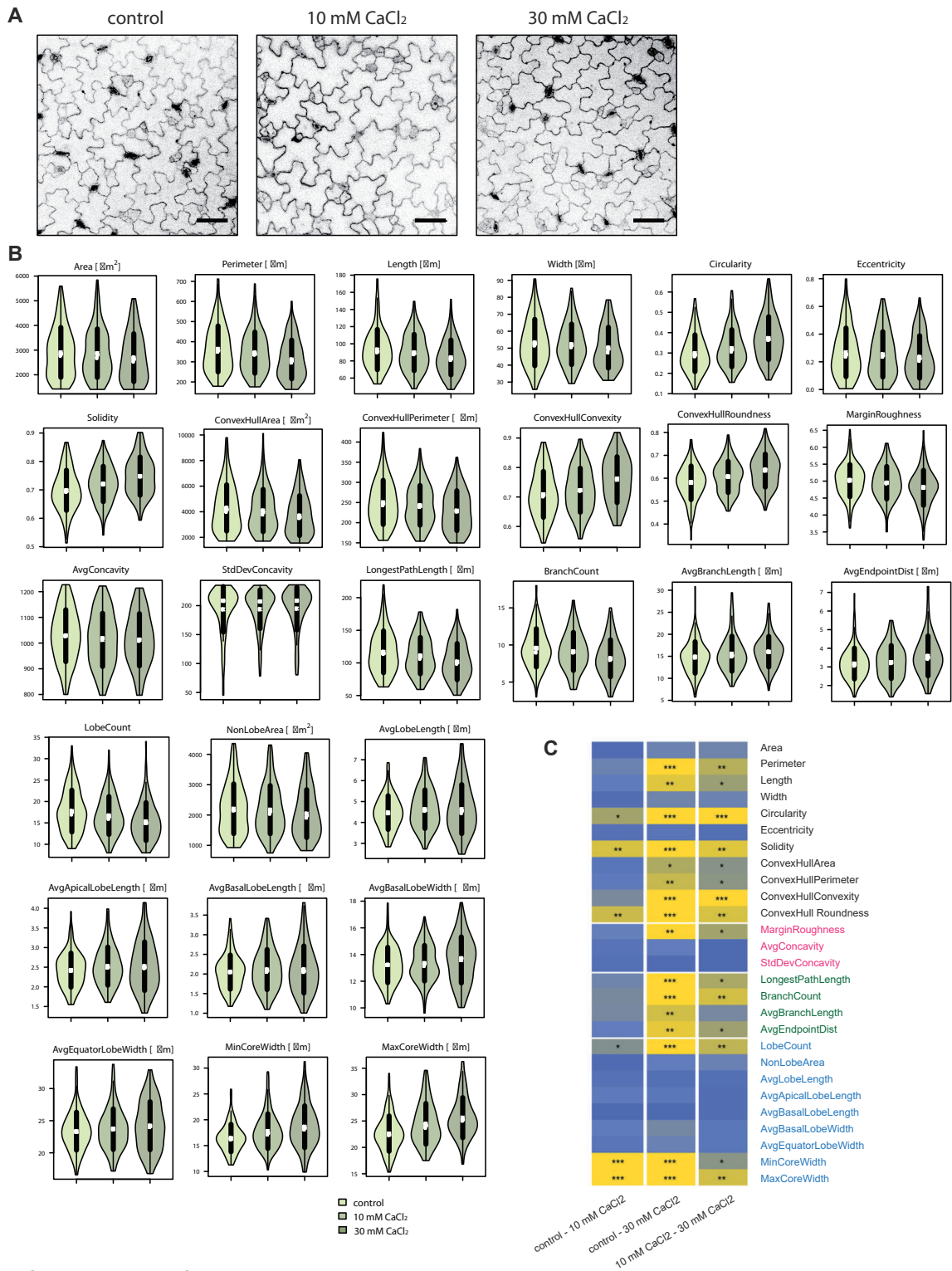
Supplementary Fig. S10.
Calcium-dependent changes in PC shape. A) Representative images of epidermis cells in seedlings 5 DAG grown on control media or on supplemented with 10 or 30 mM CaCl$_2$. B) Violin plots of all 27 shape features quantified with PaCeQuant. C) Statistical analysis from pairwise comparisons. Blue colors represent p-values close to 1, yellow colors represent values close to 0. Stars indicate statistically significant differences (*padj<0.05, **pqdj<0.01, ***padj<0.005).

Supplementary Fig. S11
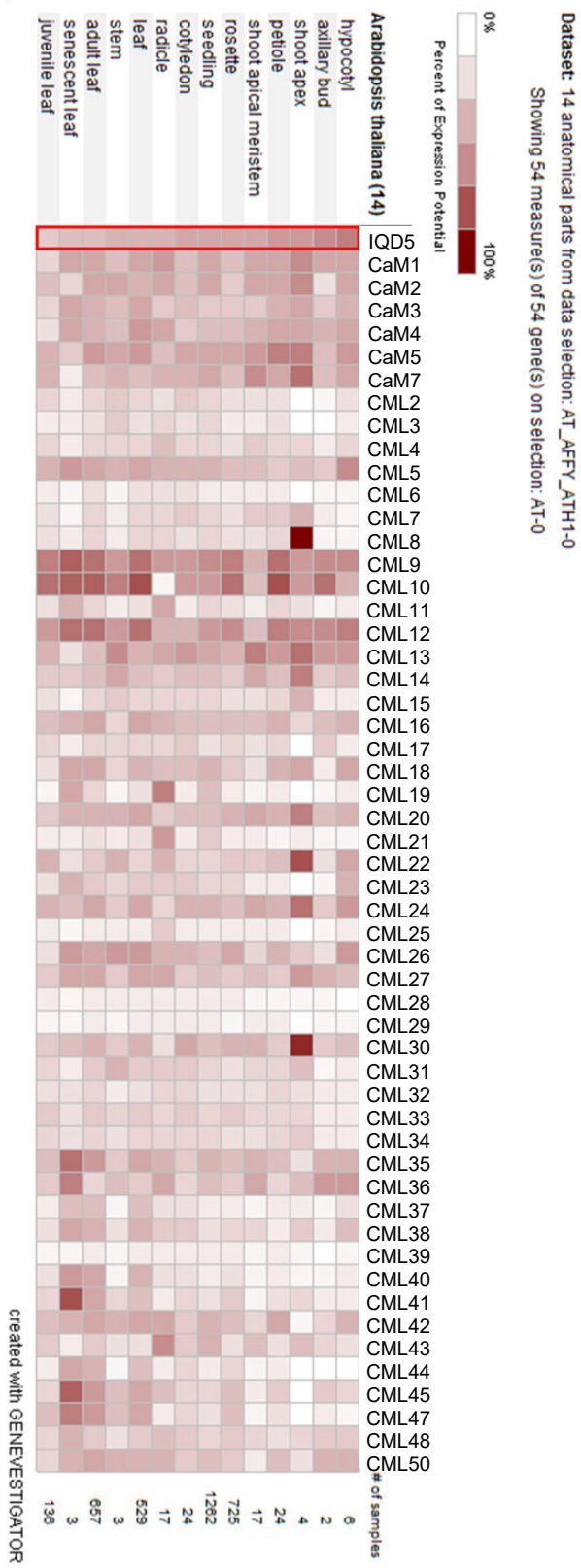
Supplementary Fig. S11.
In silico expression analysis of *IQD5,* and *CaM* and *CML* genes using Genevestigator (Zimmermann et al., 2014, BioData Mining, 7:18). Note that *CaM2/3/5* and *CaM1/4* encode for identical proteins. No expression information is available for *CaM6, CML1, CML46*, and *CML49*.

B.4 The IQD Family of Calmodulin-Binding Proteins Links Calcium Signaling to Microtubules, Membrane Subdomains, and the Nucleus (Bürstenbinder et al., Plant Physiology, 2017)

Figure S1   Bürstenbinder et al.
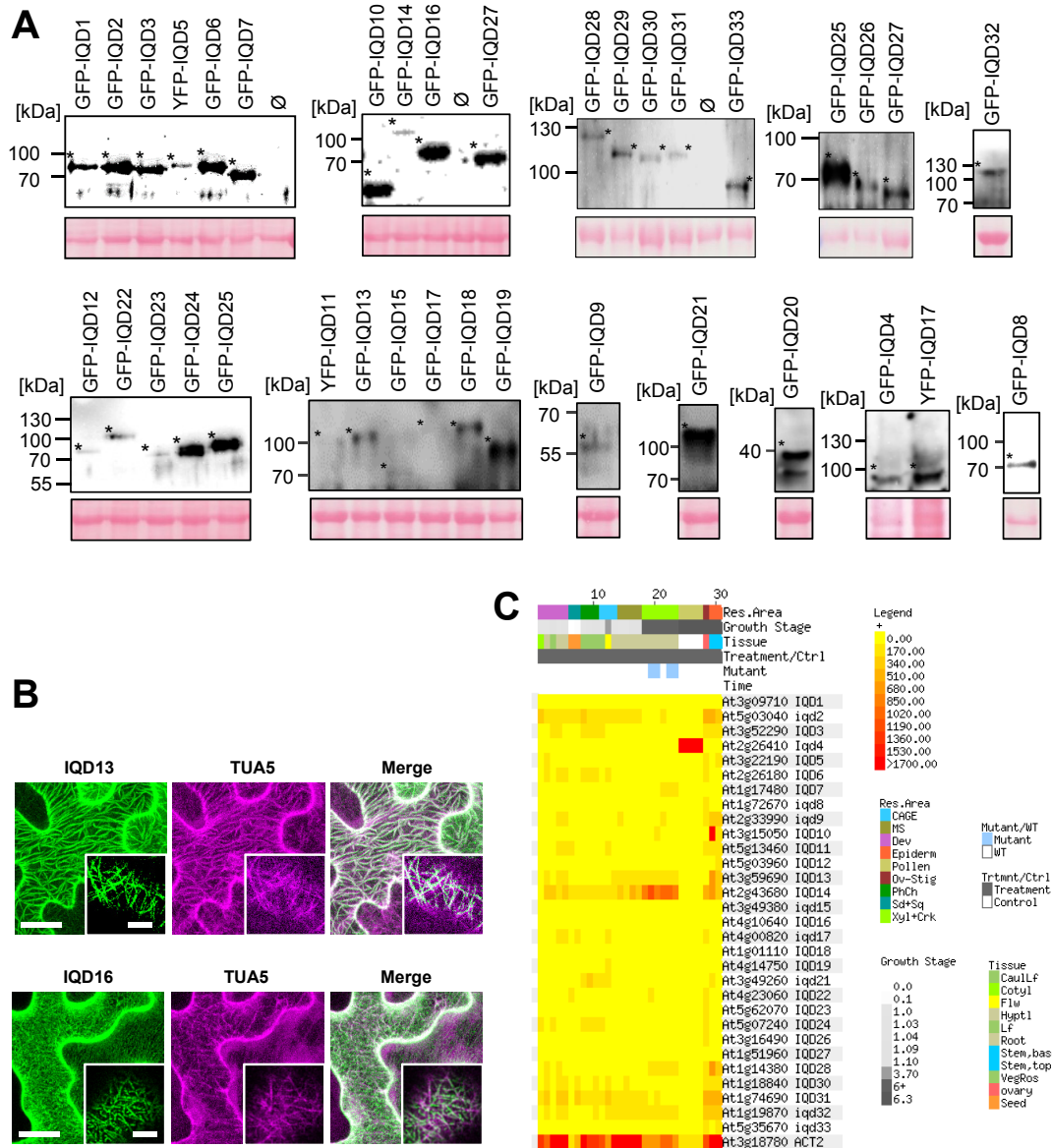
Supplemental Figure 1. Expression Analysis and Subcellular Localization of *A. thaliana IQD* Genes.
A, Expression of full-length GFP-IQD fusion proteins was confirmed by immunoblot analysis using an anti-GFP antibody, as shown for a few examples. For a list of calculated molecular masses of the GFP-IQD fusions see Supplemental Table S1. B, Co-expression of RFP-TUA5 with GFP-IQD13 (left) and GFP-IQD16 (right) in *N. benthamiana* epidermis cells. Note the differences in RFP-TUA5 labeled MT networks. Micrographs of cells are projections of Z-stacks; insets are single-layer images. Bars, 20 µm and 5 µm (insets). C, *IQD* gene expression was analyzed in various tissues and developmental stages using the AtGenExpress_Plus_Extended Tissue Series dataset provided by the BAR expression browser in comparison to *ACTIN2*. Color code in the heat map represents relative expression intensities. Analyzed tissues and growth stages, as listed in the legend are highlighted on the right of the heat map. Expression information is not available for *IQD20, IQD25*, and *IQD29*.

Figure S2   Bürstenbinder et al.
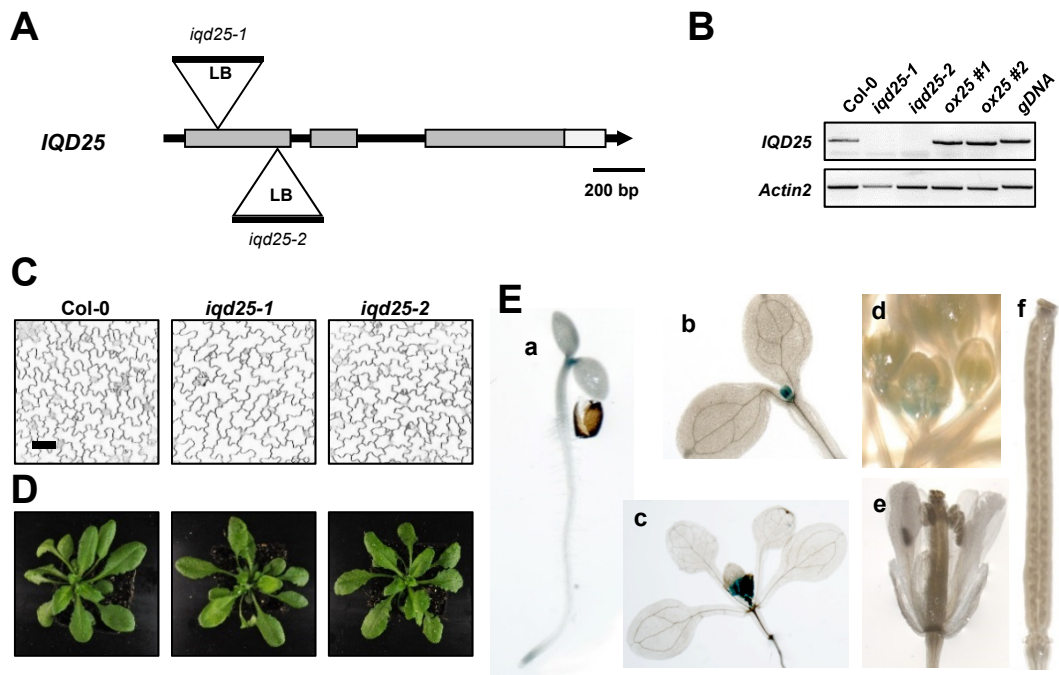
Supplemental Figure 2. Phenotypes of *iqd25* Mutant Lines and Expression Analysis of *ProIQD25:GFP-GUS*.
A, Gene model and position of the T-DNA insertions in *iqd25-1* and *iqd25-2* mutants. B, RT-PCR analysis of *IQD25* transcript levels in Col-0, two T-DNA insertion lines (*iqd25-1* and *iqd25-2*) and in 2 independent *Pro35S:IQD25 A. thaliana* lines (*ox25 #1* and *ox25 #2*). *ACTIN2* was included as control for RT efficiency and for semi-quantitative analysis of *IQD25* transcript levels. C, Epidermis pavement cell shape in cotyledons of Col-0 and *iqd25* loss-of-function lines (top panels). The scale bar represents 50 μm. D, Shoots of 4-week-old Col-0 and *iqd25* mutant plants grown on soil under long-day conditions. E, Histochemical GUS analysis of transgenic *ProIQD25:GFP-GUS A. thaliana* seedlings. a, 4-d-old seedling, b, c shoots of b) 6-d-old and c) 10-d-old seedlings, and in d) flower buds, e) flowers, and f) siliques of 5-week-old plants.

Figure S3   Bürstenbinder et al.
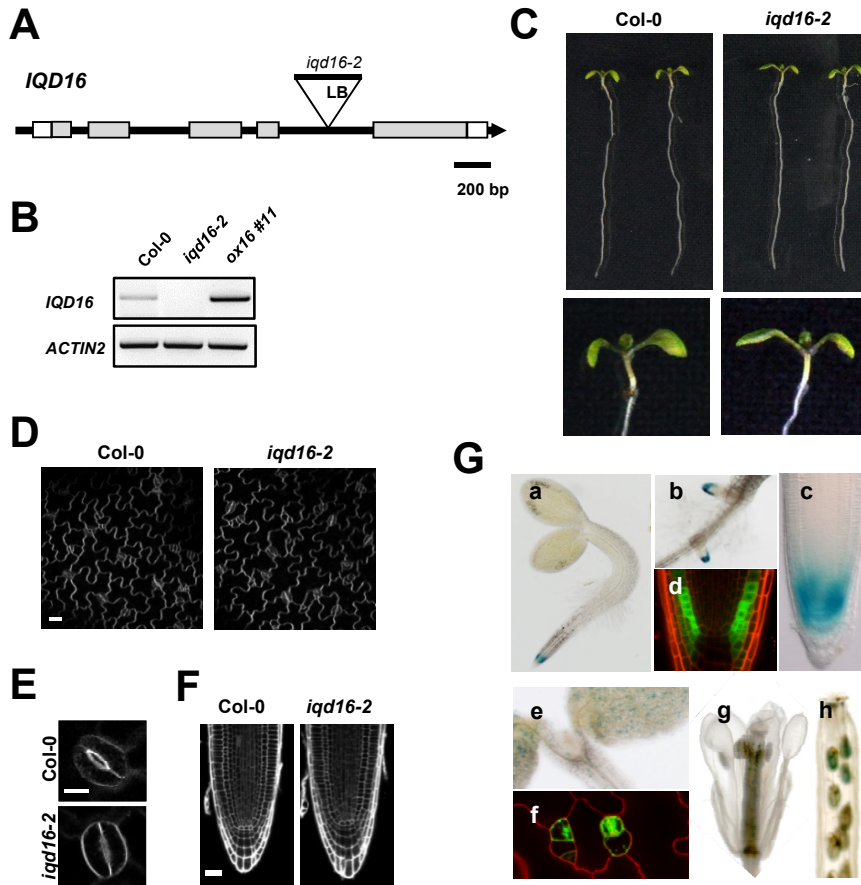
Supplemental Figure 3. Phenotypes of *iqd16* Lines and Expression Analysis of *ProIQD16:GFP-GUS*.
A, Gene structure of *IQD16* with insertion site and orientation of the T-DNA. White boxes represent the 5'UTR and 3'UTR, grey boxes the exons, black line the introns and intergenic regions. LB, left border. B, RT-PCR amplification of *IQD16* transcripts in wild-type (Col-0), *iqd16-2*, and *p35S::IQD16* #11 seedlings relative to *ACTIN2*. C-E) Phenotype of 5-day-old wild-type (Col-0) and *iqd16-2* seedlings grown on ATS media under long-day condition. C, Complete seedlings (upper panels), and close-up of hypocotyls and cotyledons (lower panels), D-F) Single optical sections. Cell walls were visualized by PI staining. D, Epidermis pavement cells of the adaxial site of cotyledons. Scale bars represent 20 µm. E, Individual stomata. Scale bars represent 5 µm. F, Longitudinal sections of primary root tips. Scale bars represent 20 µm. G, a-c, e, g, h) Histochemical localization of *IQD16* promoter driven GUS expression in 2-day-old seedlings (a), in lateral roots (b), in the primary root tip (c), and in cotyledons and the hypocotyl of 5-day-old seedlings (e), in flowers (g), and in siliques of 5-week-old plants (h). d, f Fluorescence microscopic analysis of *IQD16* promoter driven GFP expression (green), cell walls were stained with PI (red). Primary root tip (d), and epidermis and stomata cells in cotyledons of 4-day-old seedling (f).

Figure S4   Bürstenbinder et al.
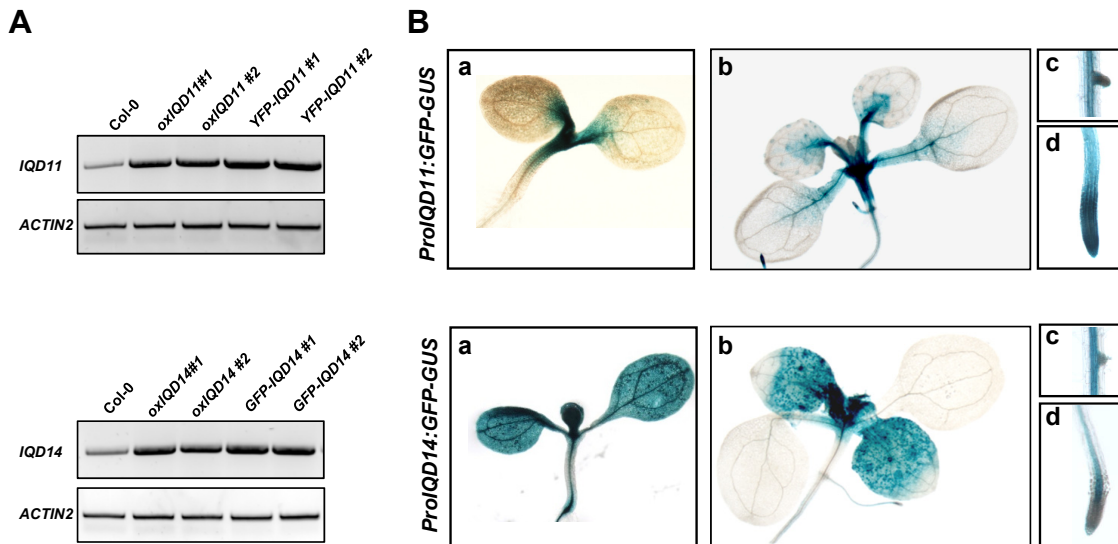
Supplemental Figure 4. Expression Analysis of *IQD11* and *IQD14* Overexpression Lines and of *ProIQD11:GFP-GUS* and *ProIQD14:GFP-GUS* Lines

A, RT-PCR analysis of *IQD11* and *IQD14* expression levels in wild type and in two independent transgenic lines of *Pro35S:IQD11* (oxIQD11), *Pro35S:YFP-IQD11* (YFP-IQD11), *Pro35S:IQD14* (oxIQD14) and *Pro35S:GFP-IQD14* (GFP-IQD14) relative to *ACTIN2*. B, Histochemical GUS analysis transgenic *ProIQD11:GFP-GUS* and *ProIQD14:GFP-GUS* Arabidopsis seedlings. Seedlings were grown on ATS medium under long-day conditions; a) cotyledons of 5-day-old seedlings, and b) shoots, c) lateral roots, d) primary root tip of 10-day-old seedlings.

Figure S5   Bürstenbinder et al.
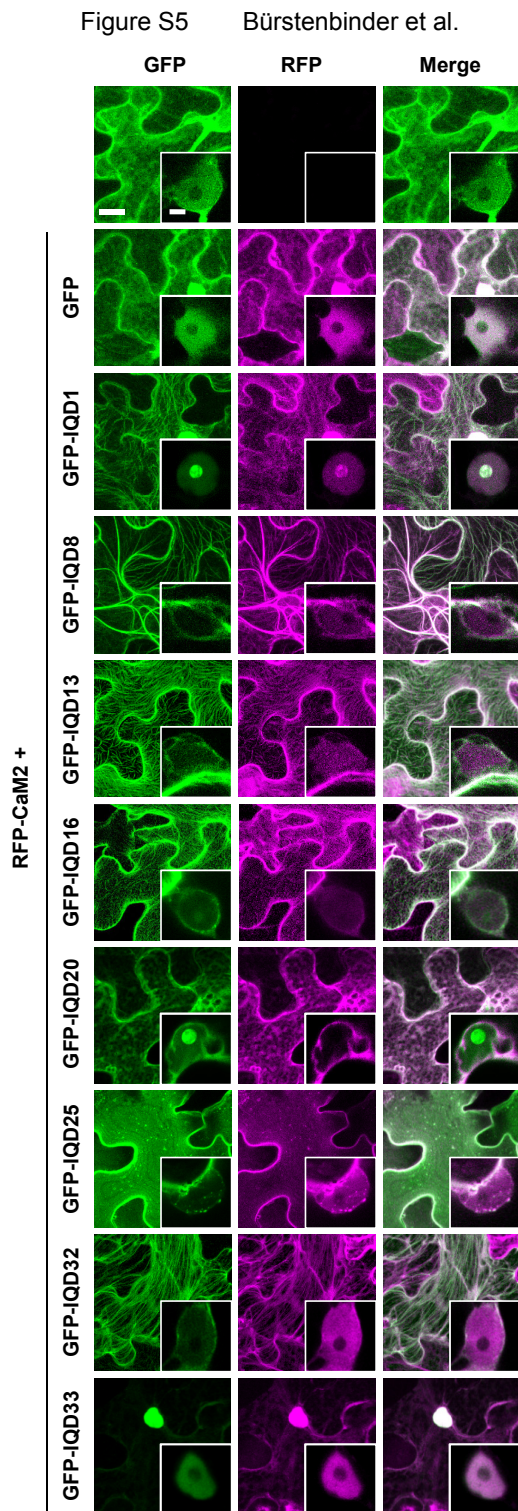
Supplemental Figure 5. Co-Expression Assays of GFP-IQD Fusions with RFP-CaM2
Z-stack images of cells (co)-expressing GFP, or GFP-IQD fusions with RFP-CaM2. Insets, single optical sections of nuclei. Left column, GFP signals, center column, RFP signals, right column, merged images. Scale bars, 20 µm and 5 µm (insets).
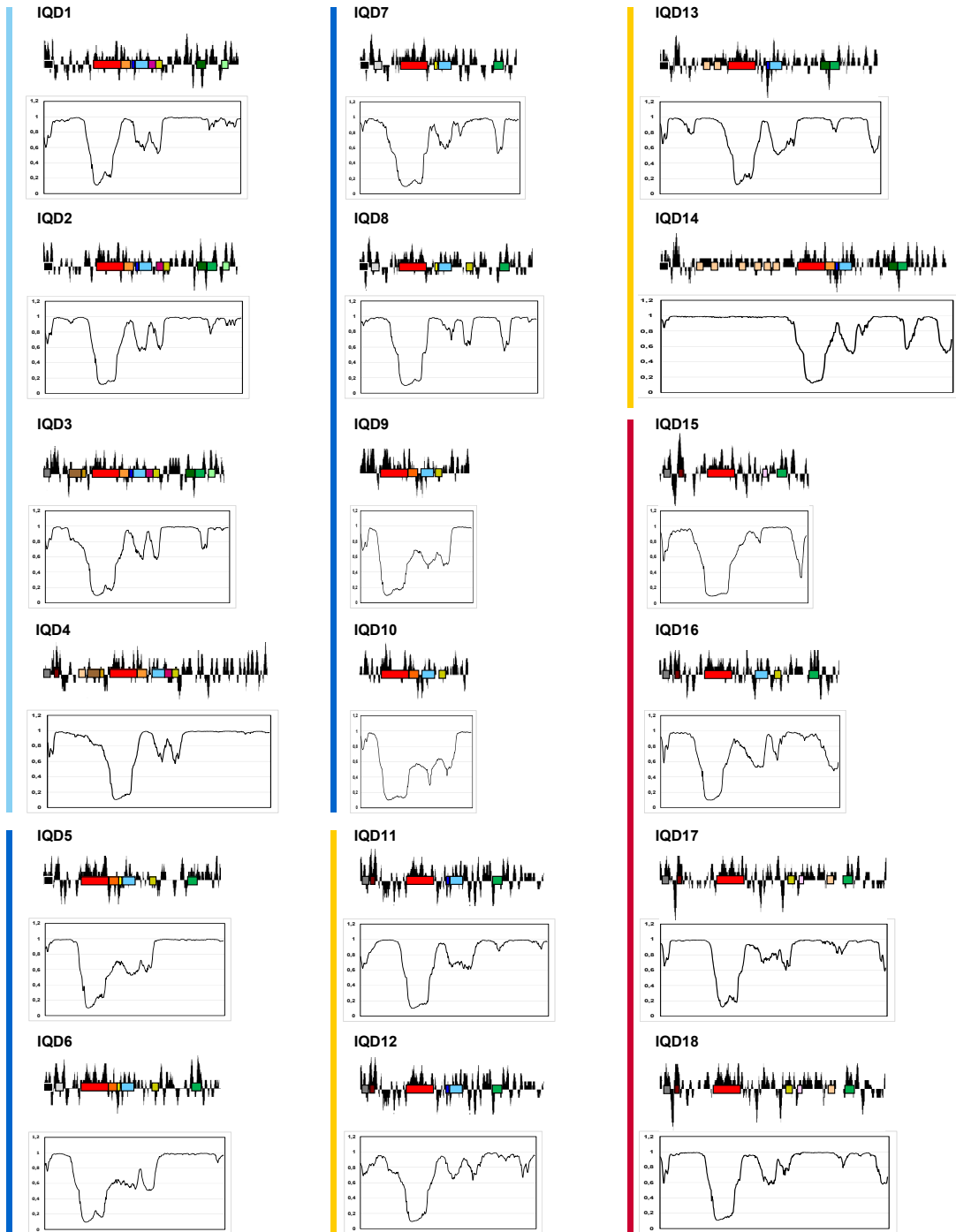
Figure S6   Bürstenbinder et al.

Figure S6   Bürstenbinder et al.
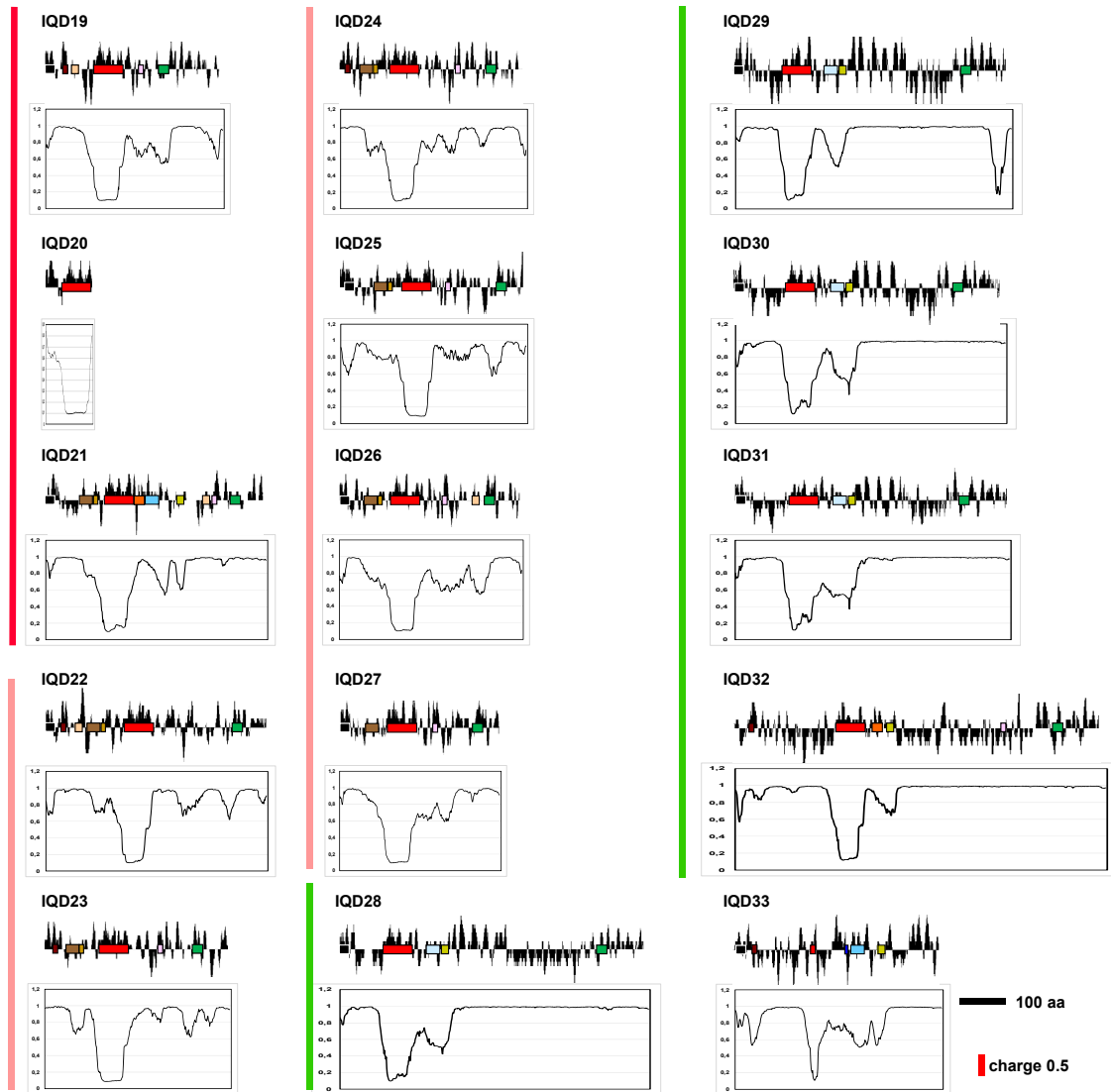
Supplemental Figure 6. Predicted Properties of *A. thaliana* IQD Proteins.
The predicted amino acid sequences of the 33 *Arabidopsis* IQD proteins were analyzed for charge distribution (top) and intrinsic disorder (bottom) using the EMBOSS charge tool (http://emboss.bioinformatics.nl/cgi-bin/emboss/charge) and SPINE-D (Zhang et al., 2012) neural networks (http://sparks-lab.org/SPINE-D/), respectively. The plots (length of IQD proteins) are drawn to scale, and shown according to the position of motifs and domains (red, IQ67 domain) and phylogenetic groups (vertical bars). Mean charges were computed across a window of 5 amino acid residues using the following values: -1 (Asp, Glu), +0.5 (His), +1 (Lys, Arg). The IQ67 domain is likely highly ordered (scores < 0.4), whereas the flanking regions are predicted to be largely disordered. Scores of 0.4-0.7 indicate semi-disorder (semi-collapsed with some secondary structure), and scores of 0.7-1.0 indicate full disorder.

Supplemental Table 1: Prediction of nuclear localization signals, lipidation sites, and signal peptides in *A. thaliana* IQD proteins

Gene Identifier	Protein	Size (aa)	MW [kDa]	GFP-fusion MW [kDa]	Nucleus	Nuclear Localization Signal — NLS (NLStradamus, 4 state HMM, static)	Palmitoylation (S-Acylation) GPS-lipid, medium — Cluster / Position (aa)	CSS-Palm, Medium — Position (aa)	N-Myristoylation GPS-lipid, medium	PlsitP	Farnesyl GPS-lipid, medium	Farnesyl PrePS	Geranylgeranyl GPS-lipid, medium	Geranylgeranyl PrePS	Signal-Blast, SPI	SignalP 4.1, default
At5g09710	IQD1	454	50.5	79.2	yes	329 - NPRKKRPIPSSHKSKSS - 345, 417 - KAPAKKRLSTSASPARKPRSSAPPKVEK - 445	Cluster A, 155		negative	negative	negative	negative	negative	negative	negative	negative
At5g03040	IQD2	461	50.5	79.1	yes				non-consensus	negative	negative	negative	negative	negative	negative	negative
At3g52290	IQD3	430	48.1	76.8	yes				non-consensus	negative	negative	negative	negative	negative	negative	negative
At2g08410	IQD4	527	58.3	85.3	yes	12 - ALSPFKGKKEDKPHKSKKWFGHSRKK - 37	Cluster A,C, 8,13	8,13	negative	negative	negative	negative	negative	negative	negative	negative
At2g22190	IQD5	400	44.5	75.5	no	18 - KDKKNQKPEKPHRKWSFGKQKS - 39, 479 - KKKPP - 483, 488 - PKKR - 491			non-consensus	negative	negative	negative	negative	negative	negative	negative
At3g26180	IQD6	416	46.9	75.5	yes	355 - NPKEFRSR - 361			non-consensus	positive	negative	negative	negative	negative	negative	negative
At1g17480	IQD7	370	41	69.8	no	24 - KGNGKKWKWK - 33	Cluster A, 141		non-consensus	negative	negative	negative	negative	negative	negative	negative
At1g72670	IQD8	414	45.9	74.5	no		Cluster B, 361		non-consensus	negative	negative	negative	negative	negative	negative	negative
At3g23990	IQD9	249	28.5	55.5	yes	12 - RLKKSNQGTEKKKTSAVKPKKGSKKKGT - 39, 58 - KAYKARKSLLRRLKGIARA - 75	Cluster B, 111		non-consensus	negative	negative	negative	negative	negative	negative	negative
At5g15050	IQD10	259	29.6	58.2	yes	229 - KK - 230	Cluster C, 12	12	negative	negative	negative	negative	negative	negative	SP, cleavage site aa 12	negative
At5g13460	IQD11	443	50.8	77.8	no	22 - EKKEKRRWVFFWKLRKKR - 40	Cluster A,B, 158, 170, 370		negative	negative	negative	negative	negative	negative	negative	negative
At4g03960	IQD12	403	46	73	no	20 - KARAEKPHRLLRVWFKFLK - 37	Cluster C, 17	17	MGXXXS/T	negative	non-consensus	negative	negative	not significant	SP, cleavage site aa 24	negative
At3g59690	IQD13	517	58.5	85.5	no	34 - KSENEKKKKGFGKKLR - 50			negative	negative	negative	negative	negative	negative	negative	negative
At2g43680	IQD14	668	74.3	103.1	no	21 - EKLANQEPERKSGKEKKKGFGKLR - 45		497	negative	negative	negative	negative	negative	negative	negative	negative
At5g08380	IQD15	352	40.8	67.8	yes	27 - RINRRIODNDLVEEVEDELHGRPKRRKRRWLFK - 58	Cluster B, 302		non-consensus	negative	non-consensus	negative	negative	negative	negative	negative
At4g10640	IQD16	423	48.7	77.4	yes	27 - KTHHIKETDMKRKEKWGWIF - 46, 363 - PRRRP - 367	Cluster B, 360, 401	401	negative	negative	negative	negative	negative	negative	negative	negative
At4g00820	IQD17	534	60	87	no	42 - KKKEKRWL - 50	Cluster A,B, 510, 511	510	non-consensus	negative	non-consensus	negative	negative	negative	negative	negative
At1g01110	IQD18	527	59.2	86.2	yes	23 - IKKDHSNDVEEDEEKKREKRHWFRKP - 47	Cluster B,C, 377, 505	505	non-consensus	negative	negative	negative	negative	negative	negative	negative
At4g14750	IQD19	387	43.9	70.9	no		Cluster A,B, 28, 59		non-consensus	negative	negative	negative	negative	negative	negative	negative
At5g51380	IQD20	103	11.8	40.5	yes				negative	negative	negative	negative	negative	negative	negative	negative
At4g49260	IQD21	471	52.1	79.1	no				MGXXXS/T	negative	negative	negative	negative	negative	negative	negative
At4g20060	IQD22	543	60.1	87.3	no		Cluster B, 397	396, 397	non-consensus	negative	negative	negative	negative	negative	negative	negative
At5g62070	IQD23	403	44.3	71.3	yes	11 - KKKSDVKAASFDKRR - 25, 76 - QSSPRRRKGK - 86			negative	negative	negative	negative	negative	negative	negative	negative
At5g07240	IQD24	401	45.3	72.3	yes				negative	negative	negative	negative	negative	negative	negative	negative
At4g29150	IQD25	383	41.4	68.4	no		Cluster B, 305	305, 306	negative	negative	negative	negative	negative	negative	negative	negative
At4g16490	IQD26	398	48.7	72	no		Cluster B, 297	346, 348	non-consensus	negative	non-consensus	negative	negative	negative	negative	negative
At1g51960	IQD27	351	39.3	68	no	260 - EAEKKRPKRSKKKPSGT - 276, 283 - SRFTAERNKPNRNVRHKA - 299, 343 - KRSSLSNGSKKKATLRSAEKKKK - 364			non-consensus	negative	negative	p-value 0.278601	CCXXC	p-value 0.198676	negative	negative
At4g14380	IQD28	664	72.8	99.8	no	17 - KKSSKISSLEKGGEKLRSA - 34, 272 - EAEKGKLKRGVRK - 284, 351 - KPRISHKKASLSNGIGKKATRKSAEKKK - 377			MGXXXS/T	negative	negative	negative	negative	negative	negative	negative
At2g02790	IQD29	597	65.3	96.8	no	258 - KKASLRKSQKKFASNPQ - 274, 278 - AEFARPKKSVR - 288	Cluster C, 530		non-consensus	negative	negative	negative	negative	negative	negative	negative
At1g08840	IQD30	572	62.7	89.7	no	278 - ESAKPKKSVR - 288, 465 - SAGKENGKSRKKG - 477	Cluster B, 155, 185	155, 156, 249	MGXXXS/T	negative	negative	negative	negative	negative	negative	negative
At4g74690	IQD31	587	65.2	92.9	yes			253	non-consensus	negative	negative	negative	negative	negative	negative	negative
At1g19970	IQD32	794	86.8	113.6	yes	656 - KARKGKSRGSSSDKRKKKK - 676	Cluster A,C, 9, 15		non-consensus	negative	negative	negative	negative	negative	negative	negative
At5g35670	IQD33	442	49.5	78.2	yes	386 - RKSVPSYKSDRHHHKLOATKSDLODGDTKAKKA - 418	Cluster B, 425	9, 15	non-consensus	negative	negative	negative	negative	negative	negative	negative

B. Supplemental Material

Supplemental Table 2.Gene Identifiers and Clone Information.

Gene name	Gene model	Primer Orientation		Position relative to ATG		Vector
		Fwd	Rev	Start	Stop	
IQD1	At3g09710.1	I108	I109	1	1365	pENTR/dTOPO
IQD2	At5g03040.1	I110	I111	1	1386	pENTR/dTOPO
IQD3	At3g52290.1	I112	I113	1	1293	pENTR/dTOPO
IQD4	At2g26410.1	I114	I115	1	1584	pENTR/dTOPO
IQD5	At3g22190.1	I800	I801	1	1269	pENTR3CUSER
IQD6	At2g26180.1	I118	I119	1	1251	pENTR/dTOPO
IQD7	At1g17480.1	I120	I121	1	1116	pENTR/dTOPO
IQD8	At1g72670.1	I222	I223	1	1245	pENTR/dTOPO
IQD9	At2g33990.1	I124	I125	1	792	pENTR/dTOPO
IQD10	At3g15050.1	I126	I127	1	780	pENTR/dTOPO
IQD11	At5g13460.1	I128	I129	1	1332	pENTR/dTOPO
IQD12	At5g03960.1	I130	I131	1	1212	pENTR/dTOPO
IQD13	At3g59690.1	I132	I133	1	1554	pENTR/dTOPO
IQD14	At2g43680.1	I134	I135	1	2010	pENTR/dTOPO
IQD15	At3g49380.1	I136	I137	1	1059	pENTR/dTOPO
IQD16	At4g10640.1	I138	I139	1	1272	pENTR/dTOPO
IQD17	At4g00820.1	I140	I141	1	1605	pENTR/dTOPO
IQD18	At1g01110.2	I142	I143	1	1584	pENTR/dTOPO
IQD19	At4g14750.1	I144	I145	1	1164	pENTR/dTOPO
IQD20	At3g51380.1	I146	I147	1	312	pENTR/dTOPO
IQD21	At3g49260.1	I148	I149	1	1419	pENTR/dTOPO
IQD22	At4g23060.1	I150	I151	1	1455	pENTR/dTOPO
IQD23	At5g62070.1	I152	I153	1	1212	pENTR/dTOPO
IQD24	At5g07240.1	I154	I155	1	1206	pENTR/dTOPO
IQD25	At4g29150.1	I156	I157	1	1200	pENTR/dTOPO
IQD26	At3g16490.1	I158	I159	1	1170	pENTR/dTOPO
IQD27	At1g51960.1	I160	I161	1	1056	pENTR/dTOPO
IQD28	At1g14380.1	I162	I163	1	1995	pENTR/dTOPO
IQD29	At2g02790.2	I164	I165	1	1763	pENTR/dTOPO

IQD30	At1g18840.1	I873	I874	1	1719	pENTR3CUSER
IQD31	At1g74690.1	I168	I169	1	1764	pENTR/dTOPO
IQD32	At1g19870.1	I170	I171	1	2385	pENTR/dTOPO
IQD33	At5g35670.1	I172	I173	1	1329	pENTR/dTOPO
TUA5	At5g19780.1	A019	A020	1	1353	pENTR/dTOPO
PDLP1	At5g43980.1	A027	A028	1	909	pENTR/dTOPO
TRM1	At3g02170.1	I752	I753	1	2718	pDONR221
Rem6.6	At1g13920	I1378	I1379	1	1038	pDONR207
Rem6.7	At5g61280	I1374	I1375	1	792	pDONR207
ProIQD11	At5g13460	I192	I193	-1	-1253	pENTR/dTOPO
ProIQD14	At3g59690	I198	I199	-1	-1868	pENTR/dTOPO
ProIQD16	At4g10640	I202	I203	-1	-1110	pENTR/dTOPO
ProIQD25	At4g29150	I220	I221	-1	-1090	pENTR/dTOPO
ProIQD8:IQD8	At1g72670.1	I250	I1251	-688 (intergenic)	1757	pDONR221
ProIQD14:IQD14	At2g43680.1	I1468	I1469	-1868	2283	pDONR221
ProIQD16:IQD16	At4g10640.1	I1234	I1235	-1110	2311	pDONR221
ProIQD20:IQD20	At3g51380.1	I1380	I1381	-2201	388	pDONR221
ProIQD25:IQD25	At4g29150.1	I1382	I1383	-1090	1568	pDONR221
ProIQD28:IQD28	At1g14380.1	I1369	I1370	-1520	2480	pDONR221

B. Supplemental Material

Supplemental Table 3. Primer Combinations Used for qPCR, RT-PCR and Genotyping.

		Primer Orientation		Position relative to ATG in genomicDNA		Amplicon length
		Fwd	Rev			
iqd16 genotyping	WT allel	I035	I036	993	1849	857
	Insert	A004	I036	SALK_LB	1849	416-616
iqd25 genotyping	WT allel	I053	I054	-369	537	855
	Insert *iqd25-1*	A004	I054	SALK_LB	537	482-732
	Insert *iqd25-2*	A004	I054	SALK_LB	537	330-530
				Position relative to ATG in CDS		
IQD16 qPCR	At4g10640.1	I409	I410	757	840	84
PP2AA3 qPCR	At1g13320.1	A015	A016	824	897	73
IQD16 RT-PCR	At4g10640.1	I035	I036	564	807	244
IQD11 RT-PCR	At5g13460.1	I088	I089	3	402	400
IQD14 RT-PCR	At3g59690.1	I029	I030	1	643	643
IQD25 RT-PCR	At4g29150.1	I095	I096	56	542	487
ACTIN2 RT-PCR	At3g18780.2	A005	A006	695	1121	427

Supplemental Table 4. Primer Sequences.

Primer	Sequence(5'→3')
I108	CACCATGGTTAAAAAAGCGAAATGGC
I109	TCACGGCGTTCTCTCTGC
I110	CACCATGGGGAAAAAAGCTAAATGGT
I111	TCAGCTGCCTGCTCCGTT
I112	CACCATGGGTAAGAGTTGGTTTTCAG
I113	TTAAGCAAGATTAGTATCTTTCTTAGTAA
I114	CACCATGGGTAAGAACTGGTTAACATGTG
I115	TTAACTGCTACCACCATTTCTTCTC
I800	GGCCATTUATGGGAAGATCTCCAGCTTCT
I801	GGTGATTUCTATGCAAGCCTCTGTTTTATTGG
I118	CACCATGGGTGCTTCAGGGAAAT
I119	TTAACCTCTCGGCTTCTCGA
I120	CACCATGGGTGGGTCAGGAAATT
I121	TTAGCTTCGCTGGCTCTTG
I122	CACCATGGGTGGCTCTGGAAATT
I123	TTAGCCTCTCTGGCTCTTTGC
I124	CACCATGGGTTCTGGGAATTTGATT
I125	TCAAGCACCTGGAAATGACA
I126	CACCATGGGATCTGGATGGCTG
I127	TTATCCGGAACCAGGCTTT
I128	CACCATGGCTAAGAAGAAGGGCTTG
I129	TCATCTCAAGCTGCTCTGCTT
I130	CACCATGGCGAAGAGGAGGTCG
I131	TCAACAATTATGTTGATATGTGGTCA
I132	CACCATGGGGAAGAAAGGAAGTTG
I133	TCACGCAAATCTGTTAAAAGCC
I134	CACCATGGTGAAGAAAGGAAGTTGGTT
I135	TCACACAAATCTGTTAAATTCCTTTC
I136	CACCATGGGGAAAACCGACGGA
I137	TTAGTACTGAAAATCTTCGTGAGCA
I138	CACCATGGCTAAAAAGAACGGAACG
I139	TCACTTTAACCACCGGAGCT
I140	CACCATGGGTAAGAAGAGCGGTTCT
I141	TCATCTTAACCATCGCCTATAATC

B. Supplemental Material

I142	CACCATGGGGAAAAAGAACGGC
I143	TCATCTAAGCCAATTCCTAAAGTCA
I144	CACCATGGGGAAAACTAGCAAATGG
I145	TCAGTACATGTTGTTGTTTCCCTG
I146	CACCATGGCCAACTCCAAACGTT
I147	TTAATGAGAGAGAAGCTGACGAGC
I148	CACCATGGGGAAGAAAGGGAGTG
I149	CTAATGATCATGCCTCCAGC
I150	CACCATGGGAAAAGCGTCACGG
I151	TCAGTACCTATACCCAATTGGCA
I152	CACCATGGGCTTTTTCGGGAGA
I153	CTAAACAAGAAACGAAGAATGCATC
I154	CACCATGGGGTTTCTTTGGAAGACTGTT
I155	CTATTGAAAGAAAAGAGGATTAGAACT
I156	CACCATGAGAAAGAATCTCACAAAATTGAC
I157	TCACCAACGCATCCTACG
I158	CACCATGGGGAAGAGCTGCGAGAT
I159	CTAATTATAGAATCTAAAATCAGTCTCG
I160	CACCATGGGCAGAGCAGCAAGAT
I161	TTAAAGCGGATCACAGGAACA
I162	CACCATGGGGAAAGACTCCTGGTAAATG
I163	TCACCGTTTCCAGTCGGT
I164	CACCATGGGGAAAGACTCCAAGTCCT
I165	TCACTCAGCTTTCGTTGACTCA
I873	GGTGATTUATGGGAAAGCCTGCAAGG
I874	GGCCATTUTCATCTCTTCCACTCTACCGG
I168	CACCATGGGGGAAGTCTACAAAATGGTT
I169	TCACCTCTTCCGCTCTGC
I170	CACCATGGGAAGATCTCCAGCTTCT
I171	AAATCACCTCTGCCATTTTCTATCC
I172	CACCATGGGTGTTACAGGAGGATTAGTC
I173	TTAGGTGCTGCTATTTAGCTTATGTG
A027	CACCATGAAACTCACCTATCAATTCTTCATC
A028	TTTATAAGCATCATATTTATTACTCTTCTTCT
I752	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGTCTGCAAAACTTCTATATAA
I753	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAGCAGAAGCAAACTTCATT

I1378	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGATACCTTAATCAAGCAGAC
I1379	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGAAACAGCATGCATTTC
I1374	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGATAATTTGGTTAAGCA
I1375	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAGTAACACCGAAAGCAGA
I192	CACCACTTCGAAGTGAAACATATGGAAC
I193	TGTTGTACAACCATCACCTGC
I198	CACCATCAGAATCTCTCCACTAAACCTAAT
I199	TTTGCCAACATCACTTTCCTT
I202	CACCGTGTAAACACGTGACAACACCA
I203	ATTGAATTAACGTTTTCTAAAAGCG
I220	CACCACGCAACATTTTGAGTTATTCTTTG
I221	AATGGTTTTAGTTTGGTTTCACTTG
I1250	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGTTGCTCAAAGGAAGGAATAAT
I1251	GGGGACCACTTTGTACAAGAAAGCTGGGTCGCCTCTCTGGCTCTTTGC
I1468	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCatcagaatctctccactaaacctaat
I1469	GGGGACCACTTTGTACAAGAAAGCTGGGTCCACAAATCTGTTAAATTCCTTTC
I1380	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCcagtaacgtagaactcacttccatca
I1381	GGGGACCACTTTGTACAAGAAAGCTGGGTCATGAGAGAGAAGCTGACGAGC
I1382	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCACGCAACATTTTGAGTTATTCTTTG
I1383	GGGGACCACTTTGTACAAGAAAGCTGGGTCCCAACGCATCCTACG
I1369	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCtatgcatctttaggtccatataaatt
I1370	GGGGACCACTTTGTACAAGAAAGCTGGGTCCCGTTTCCAGTCGGTTCTGGT
I035	CGCTGAATCTAACGGTTTTTG
I036	TAGAAGCTCCCTATCATCGCC
A004	ATTTTGCCGATTTCGGAAC
I053	ACGAAGTCGATTTTTCTCACG
I054	AACGCTTTTCTCGCCTTTTAC
I409	CGGAGCCGGTCTTCCCGGAATCA
I410	AGCCATCCATCGGTCAAGCCATTGT
A015	AGCCAACTAGGACGGATCTGGT
A016	CTATCCGAACTTCTGCCTCATTA
I088	GGCTAAGAAGAAGGGCTTGTTC
I089	CTTCAAGGCACGTAGAGCTTTC
I029	TTGGCAAAATGGTGAAGAAAG
I030	GCTTTGAAGAGATGGCTTGTG
I095	GAGCAACGAGGTGGTTCAAG

B. Supplemental Material

I096	GCTTCCATACTCCGGAGAGTC
A005	CAAAGACCAGCTCTTCCATC
A006	CTGTGAACGATTCCTGGACCT

B.5 Extraction of protein profiles from primary neurons using active contour models and wavelets (Misiak et al., Journal of Neuroscience Methods, 2014)

For this work the following supplemental material is available from the journal website at `https://doi.org/10.1016/j.jneumeth.2013.12.009`, Appendix B:

- Matlab figure file illustrating the different steps of the proposed segmentation method

B.6 rhizoTrak: a flexible open source Fiji plugin for user-friendly manual annotation of time-series images from minirhizotrons (Möller et al., Plant and Soil, 2019)

For this work the following supplemental material is available from the journal website at `https://doi.org/10.1016/j.jneumeth.2013.12.009`, section 'Electronic supplementary material':

- Supplemental figures (PDF) showing sample minirhizotron images and visualization options for annotations of root segments

- Details on import and export of RSML in rhizoTrak (PDF)

- Sample root measurements for individual root segments (CSV)

- Sample root measurements in aggregated format (CSV)

Appendix C

Declaration on Oath /
Eidesstattliche Erklärung

I hereby declare in lieu of oath that I have written this thesis on my own and without any inadmissible help from outside. I have not used any other media or materials than the ones referred to in this thesis, and I included all the citations correctly both in word or content.

Hiermit erkläre ich an Eides statt, dass ich diese Arbeit selbständig und ohne fremde Hilfe verfasst, keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

<div></div>

_____   _____
          Date   Dr. Birgit Möller

Appendix D

Short Curriculum Vitae

Dr. rer. nat. Birgit Möller

Born	September 20, 1976
Place	Bielefeld, Germany
Nationality	German

Degrees

2005	Dr. rer. nat.	Institute of Computer Science, Martin Luther University Halle-Wittenberg, Germany
2001	Dipl.-Inform.	Computer Science in the Natural Sciences, Technical Faculty, Bielefeld University, Germany

Scientific Career

Apr 2007 – today	Research Associate Pattern Recognition & Bioinformatics Group, Institute of Computer Science, Martin Luther University Halle-Wittenberg, Germany
Apr 2006 – Mar 2007	Postdoc Underwater Vision Lab, University of Girona, Spain, funded by a research grant in the Postdoc Program of the German Academic Exchange Service (DAAD)
Oct 2002 – Mar 2006	Research Associate Pattern Recognition & Bioinformatics Group, Institute of Computer Science, Martin Luther University Halle-Wittenberg, Germany
Apr 2001 – Sep 2002	Research Associate SFB 360 "Situated Artificial Communicators"/ Applied Computer Science Group, Technical Faculty, Bielefeld University, Germany
Jan 1999 – Nov 2000	Student Assistant SFB 360 "Situated Artificial Communicators"/ Applied Computer Science Group, Technical Faculty, Bielefeld University, Germany

_____   _____

      Date   Dr. Birgit Möller