

UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

Quantification of Endomembrane Phenotypes Using Chemical Genetics and Image Informatics

Permalink

<https://escholarship.org/uc/item/6xt6f1wh>

Author

Ung, Nolan Michael

Publication Date

2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Quantification of Endomembrane Phenotypes Using Chemical Genetics and
Image Informatics

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Plant Biology

by

Nolan Michael Ung

June 2015

Dissertation Committee:

Dr. Natasha V. Raikhel, Chairperson
Dr. Venugopala Reddy Gonehal
Dr. Zhenbiao Yang

Copyright by
Nolan Michael Ung
2015

The Dissertation of Nolan Michael Ung is approved:

Committee Chairperson

University of California, Riverside

Acknowledgement

I would like to thank my dissertation committee, Natasha Raikhel, Zhenbiao Yang and Venu Reddy for their support and guidance throughout my time at UCR. I would especially like to thank my committee chair and PI Natasha Raikhel for taking me under her wing and pushing me to become an interdisciplinary scientist and for having confidence in my abilities and passion. She has taught me how to persevere and always provides a great example of being a visionary scientist. She has provided me with so many opportunities to grow as a scientist and improve myself. She has opened doors for me that few others could have.

I'd like to thank Bir Bhanu and the other Co-PI's, Zhenbiao Yang, Pru Talbot and Victor Rodgers, for accepting me into the Video Bioinformatics IGERT program and giving me the opportunity to become a truly interdisciplinary scientist as well as financial support [DGE-0903667]. Without this program I would not be the scientist with the interests and skills that I have today. They have irreversibly influenced the trajectory of my career and for that I'm very thankful. I would also like to thank all of my other IGERT fellows for always being supportive and helping me with the foreign field of computer programming and computer vision. I would like to single out Asongu Tambo for being my personal computer science guru and mentor. He has spent countless hours teaching me everything I know about Matlab programming and image analysis. With respect to my graduate career, our

collaboration is the single most important opportunity that the IGERT program has given me.

I would also like to thank the NSF GRFP for providing me with funding and opportunities that has allowed me the academic freedom to focus on my research, unburdened by other responsibilities [DGE-0903667]. Because of this, I would also like to thank all of those who influenced me to apply and help revise my essays and shape them into their final form and made this opportunity happen for me.

My projects would not have taken the form they now have if it not for the guidance and influence of Glenn Hicks and David carter at the Botany Core. Glenn has showed me how to ask great questions and to think in a biological context. He have been a great sounding board and someone I can brainstorm with even when I'm not sure where my brain is going at the time. He has been an invaluable mentor and I am grateful for that. David has taught me everything I know about microscopy and has always been there to help analyzing images or me when I'm having problems imaging. He is a master of his craft and I can always rely on him for that and for great conversation during coffee hour. I would also like to thank all of the staff that have made the administrative details of graduate school a breeze. Thank you to Deidra, Kennett, Henry, Mariella, Guille and many others.

The Raikhel lab members past and present have always helped me with the day-to-day challenges how ever small or large. Mein is the backbone of the lab as

has been a rock for me, I can always count on her help and she is always straight with me, which I have always appreciated. She handles the immense responsibility she shoulders with such grace. I would also like to thank Michelle for making my transition into the Raikhel lab so smooth and helping me for that brief period of time our paths overlapped. She was so encouraging and helped solidify my decision to join the Raikhel lab. Chunhua thanks for always being so happy and cheerful in the lab while being a top-notch scientist and setting an example for the kind of post doc that I strive to be. Ruixi, thanks for your helpful conversations and suggestions with experiments and lab issues. I'd also like to thank my students that I mentored, Trevor and Sabrina. Thank you for trusting me and for being so enthusiastic about your projects and science. You give me confidence that others share my passion and that our field will continue to grow.

The Summer of 2013 I was fortunate enough to have an internship with DuPont Pioneer, working with their spectroscopy group. This would have not been possible if it weren't for April Agee, a former graduate student in Natasha's group. She made this wonderful opportunity possible and I am indebted to her for that. I'd also like to thank James Janni and Jian Jin for being great supervisors and teaching me about the industry and what it's like to be a part of a scientific team. I am extremely grateful for everyone I met while in Iowa and that unforgettable opportunity.

The GSMP has been a large part of my graduate career, helping me with my transition from undergraduate to graduate school and allowing me to give back to others as a peer mentor. I have made great friends through this program and I am a true supporter of their mission and cause. I would also like to thank the AGEP program for connecting me with other minorities in science and the MARCU program for with them I would not have decided to go to graduate school.

On a personal note I would like to thank my Grandparents Ernie and Alice Ung and Olga and Aliseo Herrera for their constant support and emphasis on education and hard work. They made so many sacrifices so that I could have this opportunity to pursue my passion and make a difference. I can't thank my parents enough or always encouraging me to pursue my passion and to not give up and for being there for me through the hard times. For always understanding me and my passion for science and encouraging me to pursue it. My parents are my permanent advocates and have always pushed me to be my best while accepting me unconditionally. I'd also like to thank my sisters for always being interested in my life and career and being my cheering squad. Lastly, I would like to thank Samantha, my future wife for being an endless source of encouragement and support. Thank you for going on this journey with me and putting up with the long hours, nights in the lab, working weekends and putting our life on hold so that I could pursue my passion and accomplish this milestone.

Thank you to all who have supported me with financial support, professional opportunities and kind words. I offer my gratitude and thanks for all of your support, guidance and encouragement during the completion of my PhD studies.

ABSTRACT OF THE DISSERTATION

Quantification of Endomembrane Phenotypes Using Chemical Genetics and Image Informatics

by

Nolan Michael Ung

Doctor of Philosophy, Graduate Program in Plant Biology
University of California, Riverside, June 2015
Dr. Natasha V. Raikhel, Chairperson

Recent advances in computer vision and image analysis have enabled biological screens of large volumes. Such high-throughput assays increase the likelihood and scope of discovery ultimately leading to functional analysis of a gene or protein. To increase the efficiency of high-throughput screens, computational tools are essential expedite image analysis, models to make sense of the extracted data, and biological assays to characterize the mutation or small molecule. Chemical genomics, the use of small molecules to inactivate proteins, is an advantageous approach when studying a conserved process such as endomembrane trafficking. Endomembrane trafficking spans kingdoms and is important in defense, development, stress response and other vital physiological process. We have taken numerous approaches to study the quantitative behavior

of the dynamic endomembrane system to accelerate discovery and better understand these complex phenomena. First, We identified six small molecules altering endomembrane trafficking in tobacco pollen; we characterized their effect on trafficking dynamics using video tracking facilitated by commercially available software giving us insight into intrinsic quantitative properties of the endomembrane system. Next, we wanted to create an automated tool to enable automatic phenotypic screening in *Arabidopsis*. EndoQuant is an automated computational tool for automatic sub cellular phenotypic analysis. Once this data was collected we developed a model to predict the biological being disrupted based on the cellular phenotype of a fluorescent marker. Using Gaussian Mixture Model, we were able to successfully predict that a subset of small molecules was disrupting endocytic recycling, leading to better experimental design and faster discovery of bioactive molecules. One particular biologically active molecule in *Arabidopsis* drastically reduced the root length of seedlings. This was found to be a result of cellulose deposition, most likely due to the miss localization of the cellulose synthesis machinery in response to disrupted trafficking membrane. We have analyzed real time membrane dynamics, automated phenotypic analysis, created a predictive model to link a phenotype with a biological process and characterized a small molecule disrupting the transport of a vital protein complex. These tools and methodologies will augment and accelerate the discovery process and our understanding of endomembrane trafficking in plant cells.

Table of Contents

Chapter 1

Video Bioinformatics and Bioimage informatics: The application of computer vision to Plant Cell Biology

1.1	Analysis of Static and Dynamic Data from Biological Systems.....	1
1.2	Segmentation: Detecting Regions of Interest.....	6
1.3	Machine learning allows automatic classification of cellular components	10
1.4	Quantifying temporal dynamics adds a new dimension of data.....	12
1.5	Opportunities for innovation in cellular video bioinformatics.....	14
1.6	References	17

Chapter 2

An Approach to Quantify Endomembrane Dynamics in Tobacco Pollen Utilizing Bioactive Chemicals

2.1	Abstract.....	26
2.2	Introduction	27
2.3	Results	32
2.4	Discussion	46
2.5	Materials and Methods.....	52
2.6	References	57

Chapter 3

EndoQuant: An Image Analysis Package for Automated Quantitative Cellular Phenotyping

3.1	Abstract.....	66
3.2	Introduction.....	67

3.3	Results	71
3.4	Discussion	81
3.5	Materials and Methods.....	85
3.6	References	87

Chapter 4

Prediction of Small Molecules Affecting Recycling Using Phenomics and Machine Learning

4.1	Abstract.....	92
4.2	Introduction	93
4.3	Results	97
4.4	Discussion	107
4.5	Materials and Methods.....	111
4.6	References	113

Chapter 5

A Novel Small Molecule Reduces Anisotropic Growth as a Result of Decreased Cellulose Deposition in *Arabidopsis* Root Cells

5.1	Abstract.....	116
5.2	Introduction	117
5.3	Results	119
5.4	Discussion	131
5.5	Materials and Methods.....	137
5.6	References	139

Chapter 6

Conclusions and Future Directions An Approach to Quantify Endomembrane Dynamics in Tobacco Pollen Utilizing Bioactive Chemicals

- 6.1 An Approach to Quantify Endomembrane Dynamics in Tobacco Pollen Utilizing Bioactive Chemicals 142
- 6.2 EndoQuant: An Image Analysis Package for Automated Quantitative Cellular Phenotyping..... 145
- 6.3 Prediction of Small Molecules Affecting Recycling Using Phenomics and Machine Learning 147
- 6.4 A Novel Small Molecule Reduces Anisotropic Growth as a Result of Decreased Cellulose Deposition in *Arabidopsis* Root Cells 150

List of Figures

Figure 2.1	Morphological phenotypes of pollen expressing Rab2:GFP	34
Figure 2.2	RAE's alter the localization and movement of RAB2:GFP	37
Figure 2.3	RAE phenotypes	39
Figure 2.4.	Movement of RAB2:GFP is inhibited by RAEs.....	40
Figure 2.5	Cumulative distribution function (CDF) of the features measured....	42
Figure 2.6	Correlation of trafficking features.....	45
Figure 3.1	EndoQuant detects cellular structures within cell boundaries	73
Figure 3.2	EndoQuant graphical user interface.....	74
Figure 3.3	EndoQuant is applicable to multiple tissue types.....	78
Figure 3.4	Analysis of EndoQuant Performance.....	80
Figure 4.1	GMM cluster analysis of small molecule phenotypes.....	99
Figure 4.2	Representative images of each phenotypic cluster	100
Figure 4.3	Phenotype confirmation of selected cluster 2 compounds	101
Figure 4.4	BFA washout using Cluster 2 compounds	103
Figure 4.5	Phenotype confirmation of selected cluster 1 compounds	104
Figure 4.6	BFA washout using Cluster 1 compounds	105
Figure 4.7	Molecular structure of cluster 2 compounds.....	106
Figure 4.8	Molecular structure of cluster 1 compounds.....	106
Figure 5.1	D5 causes sever reduction in root length and cell swelling.....	121
Figure 5.2	D5 reduces cellulose abundance in root cells.....	123
Figure 5.3	CESA3 and CESA6 miss-localize in the presence of D5	124
Figure 5.4	D5 does not affect microtubule dynamics.....	126
Figure 5.5	2.4 is a mutant resistant to the effect of D5	128
Figure 5.6	Auxin mutants show resistance to D5	129

List of Tables

Table 2.1 Phenotypes from different concentrations of compounds35

Table 2.2 Quantitative values describing the effect of each RAE.....38

Chapter 1

Video Bioinformatics and Bioimage Informatics: The Application of Computer Vision to Plant Cell Biology

1.1 Analysis of Static and Dynamic Data from Biological Systems

Video bioinformatics is a relatively new field which can be described as the automated processing, analysis and mining of biological spatiotemporal information extracted from videos obtained with varying spatial and temporal resolution (Bhanu, 2009). Advancements in the field of computer vision has given biologists the ability to not only quantify spatial and temporal dynamics but the ability to do so in a semi automatic and automatic manner. Understanding temporal dynamics can provides data previously obscured by the lack of temporal data. The challenges that arise from bioimage informatics become increasingly more complicated with the addition of the time dimension. Both techniques share very similar applications and challenges including detection of regions of interest (ROI's) via segmentation, registering images, subcellular localization determination and dealing with large amounts of image data. Here we will discuss the challenges in plant cell biology that can be addressed using automatic quantitative tools such as image and video bioinformatics and the current shortcomings that need to be

improved upon as we continue to discover and describe dynamic biological phenomena at the cellular level.

Most of the image data collected to date has been interpreted in a subjective sense, allowing for personal interpretation and a loss of objectivity (Peterson and Wolffsohn, 2007). In the pursuit of biological discovery we strive for objectivity and quantitative data that we can manipulate and use to better uncover genuine biological phenomena versus artifacts or biased results. The degree of phenotypes can be continuous and cover a large spectrum, for example, when using chemical genomics to dissect conserved cellular processes (Drakakaki, Robert, Szatmari, Brown, Nagawa, Van Damme, Leonard, Yang, Girke, Schmid, Russinova, Friml, Raikhel and Hicks, 2011). Varying concentrations of bioactive compounds or drugs can illicit proportional phenotypes (Robert, Chary, Drakakaki, Li, Yang, Raikhel and Hicks, 2008). Therefore, the need for quantitative image and video data is essential when interpreting image data on any spatial or time scale.

Ultimately, the quantified data demonstrates the most utility when subjected to statistical analysis. Therefore, it makes sense to quantify enough data to allow for a statistically valuable sample size. This often requires large amounts of data that needs to be quantified. Additionally, chemical or genetic screens have much to gain from using quantitative metrics to screen for valuable phenotypes (Beck, Zhou, Faulkner, MacLean and Robatzek, 2012). To meet these challenges in a practical

manner, quantification needs to be automated. Automation not only provides decreased analysis time, it also allows for greatly reduced inter and intra-user variability. A being able to provide a consistent analysis from sample to sample provides more reliable data. Reliable data is essential to fully understand the true nature of any dynamic subcellular process. Dynamic cellular phenomena such as cell division, lipid dynamics, plant defense processes and cell wall biosynthesis, often require the measurement of various static and dynamic features (Tataw, Liu, Roy-Chowdhury, Yadav and Reddy, 2010; Liu, Elmore, Lin and Coaker, 2011; Sampathkumar, Gutierrez, McFarlane, Bringmann, Lindeboom, Emons, Samuels, Ketelaar, Ehrhardt and Persson, 2013). The automated detection tracking and analysis of these regions of interest summarizes the major goals of video bioinformatics in a cell biological context.

Live cell imaging has become an indispensable tool for discovery throughout the basic and applied sciences. This relatively recent technique has allowed for realtime observation and quantification of dynamic biological processes on the scale of nanometers to meters and milliseconds to days (Domozych, 2012). The advent of GREEN FLUORESCENT PROTEIN (GFP), has enabled this live cell imaging revolution and has subsequently enabled the capturing of in vivo spatial and temporal dynamics (Brandizzi, Fricker and Hawes, 2002). Because of the utility and versatility of GFP and its derivatives, these tools have become ubiquitous in molecular and cell biology generating large quantities of image and video data.

Many of the technical advancements in bioimaging have come from a resolute and prolific collaboration between the biological sciences and engineering. The cooperation of these two disciplines has produced indispensable tools to cell biology such as the laser scanning confocal microscope (Rajadhyaksha, Anderson and Webb, 1999), spinning disc confocal microscope (Nakano, 2002), multiphoton microscope (Meyer and Fricker, 2000), Variable-angle epifluorescence microscopy (VAEM) (Konopka and Bednarek, 2008), and STORM (Rust, Bates and Zhuang, 2006) just to name a few. All of these imaging modalities produce large quantities of complex multidimensional data. Scientists need to work together with engineers to dissect, manage, manipulate and ultimately make sense of the image data collected. Practitioners of both disciplines, while still working to improve the acquisition hardware, are also working together to manage and analyze the large amounts of quantifiable image data.

The traditional way of quantifying image data is to manually draw regions of interest containing the biologically relevant information. To this day, this manual measurement is the most popular method of image quantification. Software tools including ImageJ and distributions of ImageJ such as Fiji are free and readily available (Schindelin, Arganda-Carreras, Frise, Kaynig, Longair, Pietzsch, Preibisch, Rueden, Saalfeld, Schmid, Tinevez, White, Hartenstein, Eliceiri, Tomancak and Cardona, 2012). Subcellular phenotyping is time consuming and impractical when performing high-throughput screens, which are necessary for most cell biologists.

This data load is only increased when analyzing videos. A recent push toward automation has favored the use of automated microscopes, and robots that perform automated high-throughput sample preparation (Hicks and Raikhel, 2009). This has led to the development and implementation of automated semi-automated tools that require modest to little user input (Kuehn, Hausner, Bungartz, Wagner, Wilderer and Wuertz, 1998; Ung, Brown, Hicks and Raikhel, 2012). Automated methods can be more consistent and faster since the user does not have to provide information. However, this lack of user input can also lead to reduced flexibility and robustness. On the other hand, semiautomated methods are flexible and possibly more robust due to user input but can often be slower because the user had to provide prior information to the software. As this analysis becomes more user friendly and practical, the ability to apply a single software tool to multiple biological problems including multi dimensional data, will be favored by the biologist thereby most likely favoring the semiautomated methods.

Bioimage informatics has experienced a recent surge in popularity due to the advent of automated microscopes and the subsequent burst of image data. Engineers had to develop methods to manage and interpret large amounts of image data generated by these automated systems. Bioimage informatics relies on several engineering disciplines including computer vision, machine learning, image processing image analysis and pattern recognition (Eils and Athale, 2003). The application of these methods aid biologists is rapid detection quantification and

classification of biological phenomena. Bioimage informatics is generally concerned with 2-dimensional data, in the x and y planes, though it is possible to deal with 3-dimensional, X, Y and Z, and 4-dimensional data X,Y,Z and frequency domain (Cheung and Cousin, 2011). Using these dimensions, data can be accurately extracted when computational techniques are properly applied. However, much information is lost when the 5th dimension of time is not considered. Here we discuss the importance of including temporal dynamics when quantifying cellular phenomena and the recent advances that were possible due to the collaboration of biologists and engineers.

1.2 Segmentation: Detecting Regions of Interest

Ultimately, biologists want to be able to extract data from the acquired multi-dimensional images. However, the biologist needs to be able to identify those sub-regions within the image that hold the most data and that are therefore more important. As expert biologists, we can accurately identify the interesting regions of an image intuitively. Segmentation is the process of partitioning the regions of an image into segments (Shapiro, 2001). Before we can extract data, we must first detect the objects or regions that are biologically meaningful. Biological images are acquired with various modalities and therefore one segmentation method is not going to be effective for all cases, therefore specialized methods must be applied to

each case on a case-by-case basis. Much progress has been made in the domain of confocal microscopy. Bright fluorophores allow for high contrast images that facilitate robust segmentation. In the realm of plant cell biology, many organelles and protein localization sites resemble bright spots or blobs. This is due to the light diffraction limit which limits the resolution of light microscopy at 250nm, making small objects appear as fuzzy blobs (de Lange, Cambi, Huijbens, de Bakker, Rensen, Garcia-Parajo, van Hulst and Figdor, 2001). Quantifying the number or size of these bright blobs is often done manually and can take several days. Simple segmentation can greatly improve this process, which can then lead to feature extraction in both static and dynamic datasets.

Imaging static 2D images is by far the most popular type of microscopy data to analyses because of its ease and relatively short acquisition and analysis time. The majority of subcellular imaging is focused on the localization of proteins of interest. Using Fluorescent markers fused to proteins of interest and dyes, cell biologists can understand the proteins that are involved with biological processes by monitoring the abundance, size shape and localization within organelles. Organelles are of constant interest to cell biologist because of their diverse and extremely important roles in plant development, homeostasis and stress responses. Automatic tools are being developed and used to quantify protein localization and spatial features of discrete compartments (Sethuraman, Taylor, Pridmore, French and Wells, 2009). Organelles often manifest as punctate dots when imaged using fluorescent confocal

laser scanning microscopy. These dots are then quantified per cell area and features extracted such as area, intensity and number of compartments (Ung, Brown, Hicks and Raikhel, 2012). Salomon et al. used such a tool to quantify the response of various endomembrane compartments to bacterial infection, cold stress and dark treatment (Salomon, Grunewald, Stüber, Schaaf, MacLean, Schulze-Lefert and Robatzek, 2010). Crucial information can also be garnered from the cells themselves. Cell borders can be detected when labeled and size as well as shape information analyzed automatically (Salomon, Grunewald, Stüber, Schaaf, MacLean, Schulze-Lefert and Robatzek, 2010). This information can then be used to track cell growth and development. Segmentation is the first crucial step to extracting quantitative information from live cell imaging data.

Cells exist in 4 dimensions, X,Y,Z and time. If cell biologists want the full complement of information from imaging data we have to consider all 4 of these dimension. Collecting and processing 3D data is computationally more expensive and more difficult to manage but can yield a greater understanding of spatial information. Most confocal microscopes can easily collect data in the Z direction and 3D reconstructions are relatively easy to do now with the capable software. Most of the images captured of dividing plant cell are 2 dimensional leaving out the critical third dimension. Miart et al. Used 3D constructions of a growing cell plate to understand the role of cellulose synthase complexes in cell plate formation by analyzing time lapse video data of fluorscently labeled Cellulose synthase

complexes (Miart, Desprez, Biot, Morin, Belcram, Höfte, Gonneau and Vernhettes, 2013). Although these analyses did not take advantage of automated quantification, the visualization of the 3D cell plate greatly contributed to the understanding of how cellulose synthase complexes (CESAs) are involved in cell plate formation (Miart, Desprez, Biot, Morin, Belcram, Höfte, Gonneau and Vernhettes, 2013). Quantifying temporal dynamics in a study such as this would lend insight into how fast this process happens and perhaps how the population of CESA complexes shifts from a homeostatic role to an actively dividing role.

4 dimensional data including 3-D movies of cellular phenomena, will become more popular as the tools to analyze this data become more sophisticated and more user friendly. Automated 4-D analysis tools are already being used by cell biologists to analyze trichome development (Domozych, 2012). This system extracts the leaf surface, segmenting the mid-plane of the young leaf and detects the developing trichomes using a hough transform which can detect circles (Illingworth and Kittler, 1987). One 3D image is registered to the next 3D image in the time series to maintain consistency and to track and compare its growth over time (Bensch, Ronneberger, Greese, Fleck, Wester, Hulskamp and Burkhardt, 2009). These tools will need to be adapted from analyzing gross morphology to tracking moving cellular structures over time.

1.3 Machine learning allows automatic classification of cellular components

Machine learning is a sub-discipline of artificial intelligence that is focused on the development of algorithms that can learn from given data (Goldberg and Holland, 1988). These methods often require the use of training data and user input to learn how the data should be classified. Upon training the algorithm can then correctly identify the class that each subject should belong to. A simple example is the spam filter on most email accounts that can discern between those messages that are spam and those that are important.

An logical application of machine learning in cell biology was determining the sub cellular localization of fluorescent markers base on extracted features. Traditionally Cell biologists have to co-localize their protein of interest and markers of known localization to determine where the protein is located. Biologists could simply analyze an confocal micrograph with a machine learning program and receive the location of their protein of interest. An additional advantage to the machine learning methods over traditional cellular methods, other than reduced time investment is that these methods provide statistics as to how likely the determined localization is to be true (Hua and Sun, 2001). Though there seems to be reasonable progress in determining sub cellular localization using machine learning, the biological community has yet to adopt the methodology. Prediction of

subcellular localization will streamline experimental design and support traditional co-localization assays.

Machine learning is a powerful tool for gene discovery and organelle dynamics. It can help us uncover relationships and details that we otherwise could not. Because organelle dynamics can be complex and highly variable its valuable to be able to simplify and summarize dynamics. Using Bayesian networks, Collinet et al. Found that endosome number, size, concentration of cargo and position are mediated by specific genetic regulation and not random (Collinet, Stöter, Bradshaw, Samusik, Rink, Kenski, Habermann, Buchholz, Henschel, Mueller, Nagel, Fava, Kalaidzidis and Zerial, 2010). Furthermore, they used this method to discover novel components regulating endocytosis by clustering endocytic phenotypes caused by screening siRNA libraries Collinet 2010}. Statistical analysis was similarly used to summarize and classify organelle movement in Arabidopsis Stomata {Higaki 2012}. The result is an atlas of organelle movement in stomata that can be compared to various conditions. Organelle movement patterns were compared between open and closed stomata revealing differences in ER position in response to stomatal opening. Organelle movement These new findings emphasize the need for statistical methods to manage complex data and present this data in forms we can easily understand and manipulate.

Though we are interested in cell autonomous processes, cells do not exist in a vacuum. We are also greatly interested in how a cell influences the development and function of its neighboring cells. To address this challenge segmentation coupled with machine learning was used to jointly detect and classify cell types in whole tissues. 3dimensional images of propidiumiodide stained roots were used to automatically find cell files in longitudinal and transverse sections using watershed segmentation and a support vector machine to classify cell types (Liu, Schmidt, Blein, Durr, Palme and Ronneberger, 2013). An alternative approach used histological sections of arabidopsis hypocotyls to differentiate tissue layers and predict the location of phloem bundle cells (Sankar, Nieminen, Ragni, Xenarios and Hardtke, 2014). The true utility of these tools will be realized when they are used to compare wild type cell profiles with mutants, possibly being used in large content screening.

1.4 Quantifying temporal dynamics adds a new dimension of data

Once an object has been detected and classified its often very important to follow its movement through time and space. This extremely important problem of tracking has been tackled by many engineers developing the field of computer vision. A multitude of tools are available for tracking cells and organelles, most of these being manual and semi automated (Ung, Brown, Hicks and Raikhel, 2012).

Tracking organelles is difficult because rarely do they have a straight forward movement model. It is because of the diversity and variability of tracking problems that semi automated methods are the most widely used. Common problems include object moving out of the plane of focus when using 2D images. It is because of this issue that 3D movies are such valuable data sets (Carlsson, Danielsson, Liljeborg, Majlöf, Lenz and Åslund, 1985). Therefore, automatically tracking object in a 3D image set is an invaluable tool (Racine, Sachse, Salamero, Fraasier, Trubuil and Sibarita, 2007). Other challenges include, maintaining identity when two objects fuse or break off from one another, and maintaining multiple tracks at the same time. A perfect tracking algorithm would overcome all of these problems while maintaining minimal user input and accurate segmentation.

The purpose of quantifying movement and movement patterns is to gain useful biological insight such as diffusion rates, types of motion including brownian motion, non-brownian motion, confined motion, directed motion or anomalous diffusion (Saxton and Jacobson, 1997). Ung et al. correlated multiple dynamic features which suggested that when tobacco pollen tubes where treated with specific bioactive compounds the contained golgi bodies increased in size and this increase in size was correlated to an increase in signal intensity and a decrease in straightness suggesting that these were possibly multiple fused Golgi and that this

fusion disrupted movement (Ung, Brown, Hicks and Raikhel, 2012). Similar correlations were made by collinet et al. When examining endocytosis in mammalian cells (Collinet, Stöter, Bradshaw, Samusik, Rink, Kenski, Habermann, Buchholz, Henschel, Mueller, Nagel, Fava, Kalaidzidis and Zerial, 2010). These trends are possibly indicative of larger principals. Indeed these data are consistent with hypothesis proposed by hamilton et al. (2007) including conservation of surface area, measurement of volume, flux across a membrane the role of pressure and tension and vesicle fusion. These biological details would not be obtainable without quantitative video analysis.

Although each challenge was presented separately, they are by no means mutually exclusive. The vast majority of image analysis problems require identification of regions of interest before they can be quantified, tracked or classified.

1.5 Opportunities for innovation in cellular video bioinformatics

As video bioinformatics tools become increasingly accurate and biologist friendly they will become more prevalent in biological studies. The future of video analysis is moving toward automatic quantification of cellular dynamics in 4-dimensions (3D time lapse images). The amount of data that can be extracted from 3D movies will increase with the availability and ease of use of software. Biologists

will be able to quantify difference in movement possibly identifying underlying principals of movement and other components essential to cellular dynamics. Once these video analysis tools become more fully automated, it will become more practical to then screen for factors that influence dynamics by performing time lapse video screens. In this manner, biologists will be able to directly screen for changes in cellular dynamics.

Creating the tools to quantify cellular dynamics is futile unless biologists use them to produce data. The pipeline from engineer to the biological community needs to be stronger. This could be enhanced by taking advantage of open source repositories of image analysis tools. A small pool of these repositories currently available and will grow in popularity as the need for these programs becomes greater (Swedlow and Eliceiri, 2009). As we take advantages of quantitative methods w will produce large amounts of quantitative data that has the potential to fuel mathematical models or other future studies (Phillips and Milo, 2009). It is to this end that repositories of quantitative data would be indispensable. Many mathematical models require know numerical parameters to be of use.

As live cell imaging modalities and acquisition methods become more advanced including super resolution methods and as biological systems change, our analysis methods will to have to adapt. In the future these efforts will be spearheaded by a handful of interdisciplinary scientists that will be trained in

biological principals, experimental design, computer programming and image analysis's tool design. Future biologists will have to be well versed in computer programming basics and be able to design tools that are applicable to their specific research topic while having a basic understanding of the appropriate algorithms all while being able to communicate with engineers. Engineers on the other hand, will have to understand biological limitations, know which features are useful, experimental design and acquisition methods.

References

Beck, M., Zhou, J., Faulkner, C., MacLean, D. and Robatzek, S., (2012) Spatio-temporal cellular dynamics of the Arabidopsis flagellin receptor reveal activation status-dependent endosomal sorting. *Plant Cell*, 24(10), 4205-19.

Bensch, R., Ronneberger, O., Greese, B., Fleck, C., Wester, K., Hulskamp, M. and Burkhardt, H., (2009) Image analysis of arabidopsis trichome patterning in 4D confocal datasets. In *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*. p 742-745.

Bhanu, Bir. "IGERT Program." The UC Riverside Integrated Graduate Education Research and Training Program in Video Bioinformatics. University of California, Riverside, 26 Oct. 2009. Web. 24 Mar. 2014.

Brandizzi, F., Fricker, M. and Hawes, C., (2002) A greener world: the revolution in plant bioimaging. *Nat Rev Mol Cell Biol*, 3(7), 520-30.

Carlsson, K., Danielsson, P.-E., Liljeborg, A., Majlöf, L., Lenz, R. and Åslund, N., (1985) Three-dimensional microscopy using a confocal laser scanning microscope. *Optics letters*, 10(2), 53-55.

Cheung, G. and Cousin, M.A., (2011) Quantitative analysis of synaptic vesicle pool replenishment in cultured cerebellar granule neurons using FM dyes. *J Vis Exp*, (57).

Collinet, C., Stöter, M., Bradshaw, C.R., Samusik, N., Rink, J.C., Kenski, D., Habermann, B., Buchholz, F., Henschel, R., Mueller, M.S., Nagel, W.E., Fava, E., Kalaidzidis, Y. and Zerial, M., (2010) Systems survey of endocytosis by multiparametric image analysis. *Nature*, 464(7286), 243-9.

Domozych, D.S., (2012) The quest for four-dimensional imaging in plant cell biology: it's just a matter of time. *Ann Bot*, 110(2), 461-74.

Drakakaki, G., Robert, S., Szatmari, A.M., Brown, M.Q., Nagawa, S., Van Damme, D., Leonard, M., Yang, Z., Girke, T., Schmid, S.L., Russinova, E., Friml, J., Raikhel,

N.V. and Hicks, G.R., (2011) Clusters of bioactive compounds target dynamic endomembrane networks in vivo. *Proc Natl Acad Sci U S A*, 108(43), 17850-5.

Eils, R. and Athale, C., (2003) Computational imaging in cell biology. *J Cell Biol*, 161(3), 477-81.

Goldberg, D.E. and Holland, J.H., (1988) Genetic algorithms and machine learning. *Machine learning*, 3(2), 95-99.

Hicks, G.R. and Raikhel, N.V., (2009) Opportunities and challenges in plant chemical biology. *Nat Chem Biol*, 5(5), 268-72.

Hua, S. and Sun, Z., (2001) Support vector machine approach for protein subcellular localization prediction. *Bioinformatics*, 17(8), 721-728.

Illingworth, J. and Kittler, J., (1987) The adaptive Hough transform. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (5), 690-698.

Konopka, C.A. and Bednarek, S.Y., (2008) Variable-angle epifluorescence microscopy: a new way to look at protein dynamics in the plant cell cortex. *Plant J*, 53(1), 186-96.

Kuehn, M., Hausner, M., Bungartz, H.J., Wagner, M., Wilderer, P.A. and Wuertz, S., (1998) Automated confocal laser scanning microscopy and semiautomated image processing for analysis of biofilms. *Appl Environ Microbiol*, 64(11), 4115-27.

de Lange, F., Cambi, A., Huijbens, R., de Bakker, B., Rensen, W., Garcia-Parajo, M., van Hulst, N. and Figdor, C.G., (2001) Cell biology beyond the diffraction limit: near-field scanning optical microscopy. *Journal of cell science*, 114(23), 4153-4160.

Linda G. Shapiro and George C. Stockman (2001): "Computer Vision", pp 279-325, New Jersey, Prentice-Hall, ISBN 0-13-030796-3

Liu, J., Elmore, J.M., Lin, Z.J. and Coaker, G., (2011) A receptor-like cytoplasmic kinase phosphorylates the host target RIN4, leading to the activation of a plant innate immune receptor. *Cell Host Microbe*, 9(2), 137-46.

Liu, K., Schmidt, T., Blein, T., Durr, J., Palme, K. and Ronneberger, O., (2013) Joint 3d cell segmentation and classification in the arabidopsis root using energy minimization and shape priors. In *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*. p 422-425.

Meyer, A.J. and Fricker, M.D., (2000) Direct measurement of glutathione in epidermal cells of intact Arabidopsis roots by two-photon laser scanning microscopy. *Journal of microscopy*, 198(3), 174-181.

Miart, F., Desprez, T., Biot, E., Morin, H., Belcram, K., Höfte, H., Gonneau, M. and Vernhettes, S., (2013) Spatio-temporal analysis of cellulose synthesis during cell plate formation in Arabidopsis. *Plant J*.

Nakano, A., (2002) Spinning-disk confocal microscopy--a cutting-edge tool for imaging of membrane traffic. *Cell structure and function*, 27(5), 349-355.

Peterson, R.C. and Wolffsohn, J.S., (2007) Sensitivity and reliability of objective image analysis compared to subjective grading of bulbar hyperaemia. *Br J Ophthalmol*, 91(11), 1464-6.

Phillips, R. and Milo, R., (2009) A feeling for the numbers in biology. *Proc Natl Acad Sci U S A*, 106(51), 21465-71.

Racine, V., Sachse, M., Salamero, J., Fraisier, V., Trubuil, A. and Sibarita, J.B., (2007) Visualization and quantification of vesicle trafficking on a three-dimensional cytoskeleton network in living cells. *J Microsc*, 225(Pt 3), 214-28.

Rajadhyaksha, M., Anderson, R. and Webb, R.H., (1999) Video-Rate Confocal Scanning Laser Microscope for Imaging Human Tissues *In Vivo*. *Applied Optics*, 38(10), 2105-2115.

Robert, S., Chary, S.N., Drakakaki, G., Li, S., Yang, Z., Raikhel, N.V. and Hicks, G.R., (2008) Endosidin1 defines a compartment involved in endocytosis of the brassinosteroid receptor BRI1 and the auxin transporters PIN2 and AUX1. *Proc Natl Acad Sci U S A*, 105(24), 8464-9.

Rust, M.J., Bates, M. and Zhuang, X., (2006) Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature methods*, 3(10), 793-796.

Salomon, S., Grunewald, D., Stüber, K., Schaaf, S., MacLean, D., Schulze-Lefert, P. and Robatzek, S., (2010) High-Throughput Confocal Imaging of Intact Live Tissue Enables Quantification of Membrane Trafficking in Arabidopsis. *Plant physiology*, 154(3), 1096.

Sampathkumar, A., Gutierrez, R., McFarlane, H.E., Bringmann, M., Lindeboom, J., Emons, A.M., Samuels, L., Ketelaar, T., Ehrhardt, D.W. and Persson, S., (2013) Patterning and lifetime of plasma membrane-localized cellulose synthase is dependent on actin organization in Arabidopsis interphase cells. *Plant Physiol*, 162(2), 675-88.

Sankar, M., Nieminen, K., Ragni, L., Xenarios, I. and Hardtke, C.S., (2014) Automated quantitative histology reveals vascular morphodynamics during Arabidopsis hypocotyl secondary growth. *eLife*, 3.

Saxton, M.J. and Jacobson, K., (1997) Single-particle tracking: applications to membrane dynamics. *Annual review of biophysics and biomolecular structure*, 26(1), 373-399.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P. and Cardona, A., (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods*, 9(7), 676-82.

Sethuraman, V., Taylor, S., Pridmore, T., French, A. and Wells, D., (2009) Segmentation and tracking of confocal images of *Arabidopsis thaliana* root cells using automatically-initialized Network Snakes. In *Bioinformatics and Biomedical Engineering, 2009. ICBBE 2009. 3rd International Conference on*. p 1-4.

Swedlow, J.R. and Eliceiri, K.W., (2009) Open source bioimage informatics for cell biology. *Trends Cell Biol*, 19(11), 656-60.

Tataw, O.M., Liu, M., Roy-Chowdhury, A., Yadav, R.K. and Reddy, G.V., (2010) Pattern analysis of stem cell growth dynamics in the shoot apex of *Arabidopsis*. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*. p 3617-3620.

Ung, N., Brown, M.Q., Hicks, G.R. and Raikhel, N.V., (2012) An Approach to Quantify Endomembrane Dynamics in Pollen Utilizing Bioactive Chemicals. *Mol Plant*.

Chapter 2

An approach to Quantify Endomembrane Dynamics in Tobacco Pollen

Utilizing Bioactive Chemicals

2.1 Abstract

Tip growth of pollen tubes and root hairs occurs via rapid polar growth. These rapidly elongating cells require tip-focused endomembrane trafficking for the deposition and recycling of proteins, membranes and cell wall materials. Most of the image-based data published to date are subjective and non-quantified. Quantitative and comparative descriptors of these highly dynamic processes have been a major challenge, but are highly desirable for genetic and chemical genomics approaches to dissect this biological network. To address this problem, we screened for small molecules that perturbed the localization of a marker for the Golgi Ras-like monomeric G-protein RAB2:GFP expressed in transgenic tobacco pollen. Semi-automated high-throughput imaging and image analysis resulted in the identification of novel compounds that altered pollen tube development and endomembrane trafficking. Six compounds that caused miss localization and varying degrees of altered movement of RAB2:GFP-labeled endomembrane bodies were used to generate a training set of image data from which to quantify vesicle dynamics. The area, velocity, straightness, and intensity of each body were

quantified using semi-automated image analysis tools revealing quantitative differences in the phenotype caused by each compound. A score was then given to each compound enabling quantitative comparisons between compounds. Our results demonstrate that image analysis can be used to quantitatively evaluate dynamic subcellular endomembrane phenotypes induced by bioactive chemicals, mutations or other perturbing agents as part of a strategy to quantitatively dissect the endomembrane network.

2.2 Introduction

Endomembrane trafficking is a dynamic network of processes that is essential for abiotic stress response, pathogen response, cytokinesis, cell expansion, tip growth and development (Samuels et al., 1995; Wick, et al., 2003; Carter et al., 2004). The endomembrane system consists of discrete membrane-bound compartments that are interconnected within complex trafficking pathways. These pathways are highly coordinated and regulated to ensure proper localization and allocation of proteins, cell wall components and other vital materials for proper growth and development (Roberts 1994; Miller et al., 1997; Fürthauer and González-Gaitán, 2009).

Pollen tubes are single, elongated cells that undergo a specialized form of polar growth known as tip growth, a mode of growth shared by root hairs, trichomes and neurons (Hepler et al., 2001). Tip growth requires high levels of tip-

focused endomembrane trafficking to deposit lipids, proteins and cell wall materials to facilitate rapid growth (Hepler et al., 2001). Pollen tubes must form an elongating tube that travels from the surface of the stigma to deliver the sperm to the ovule (Lord, 2000). This rapid tip growth is facilitated by the coordination of the actin cytoskeleton and a highly active secretory system (Hepler, et al., 2001).

Within the Ras superfamily of proteins, the Rab family of small GTPases is evolutionarily conserved and regulates the budding, transport and fusion of vesicles from intracellular donor compartments to target compartments, thus, mediating intracellular trafficking (Stenmark and Olkkonen, 2001; Nielsen et al., 2008; Ebine et al., 2011). When bound to GDP, RABs are inactive; however, when bound to GTP they are activated and interact with downstream effector proteins (Stenmark and Olkkonen, 2001). Although many RAB proteins are found in yeast, plants and mammals, of the three, RAB2 is found only in mammals and plants (Rutherford and Moore, 2002). The *Arabidopsis thaliana* ortholog RAB B1 and the mammalian RAB2 are involved with vesicle maturation of the anterograde trafficking pathway from the endoplasmic reticulum (ER) to the Golgi stacks and localizes primarily in the Golgi (Rutherford and Moore, 2002). In *Nicotiana tabacum* (tobacco), RAB2 is found predominantly in pollen grains, pollen tubes and root hairs, probably due to the high secretory demands of the two cell types. Transcripts are detected in floral buds and lateral roots, though this might be due to its presence in pollen and root hairs (Cheung et al., 2002). When fused to GFP, RAB2 is observed in Golgi stacks

that are highly active within the germinating pollen grain as well as the elongating tube (Cheung et al., 2002). This localization can be exploited to study endomembrane trafficking in a developing pollen tube. Because the endoplasmic reticulum (ER) and Golgi are major components along the secretory pathway, it can be inferred that if agents such as bioactive small molecules perturb this pathway, global secretion will be significantly impaired, allowing the effect of the small molecules to be identified by microscopy.

Qualitative descriptions of trafficking are useful but very limited in the number of comparisons that can be made and conclusions that can be drawn. Since endomembrane trafficking is a physical phenomenon that can be observed by microscopy, it can be quantified and subjected to statistical analysis. Indeed, our ability to quantify distribution, sizes and rates of endomembrane compartments in many organisms and tissue types has been growing over the past two decades (Pyke and Leech, 1991; Derksen, et al., 1995; Logan et al., 2003; Rink et al., 2005; Pochynyuk et al., 2007; Ketelaar et al., 2008; Sparkes et al., 2008; Kato, et al., 2010; Collinet et al., 2010; Salomon et al., 2010). Earlier studies used various methods to quantify trafficking that were limited by the available technology. Initial investigations utilized transmission electron microscopy and manual measurement of organelles. This was used to infer growth rates based on mathematical models (Ketelaar et al., 2008) or by calculating organelle distribution (Derksen et al.,

1995). Although informative, these studies observed frozen or fixed tissue rather than live cells, which provide a more accurate depiction of cellular processes.

The advent of GREEN FLOURESCENT PROTEIN (GFP) made quantitative analyses of living cells possible. Live imaging in combination with image analysis gave a more comprehensive description of organelle size and distribution (Pochynyuk et al., 2007). Nevertheless, these approaches used still frame images, snapshots in time that omit any temporal dynamics. To account for this, video data can capture movements and morphological changes of cellular compartments in real time while simultaneously measuring several features in animal cells (Collinet et al., 2010) and endomembrane trafficking in plant pathogen infected leaf epidermal cells (Salomon et al., 2010). To visualize and quantify video images from cells displaying rapid growth and highly dynamic trafficking, we used germinating pollen as a model system. Our present study combines quantitative video analysis with bioactive chemicals to rapidly and controllably perturb trafficking, permitting the quantitative and comparative description of compounds affecting intracellular trafficking in pollen grains.

There are several challenges to be overcome in understanding endomembrane trafficking in the sporophyte and gametophyte. Classical genetic approaches are limited by functional gene redundancy because mutation of a single gene can result in a lack of phenotype due to compensation. Alternatively,

when an essential gene is mutated, embryo lethality can result. These obstacles can be circumvented by a chemical genomics approach that uses small molecules to temporarily modulate protein function. Small molecules often bind to proteins, transiently inhibiting proper function. This marriage of synthetic chemistry and biology coupled with the grand scale of genomics is a powerful approach that produces large amounts of useful phenotypic data prior to genetics (Hicks and Raikhel, 2012). However, the power of chemical biology and genetics are both limited by an inability to accurately quantify trafficking processes to obtain objective descriptors to compare the dynamics in the presence of different inhibitors or mutations. For example, due to the high throughput nature of chemical screening, data analysis should be quantitative and automatable to accommodate large image datasets.

To address the pressing need to develop quantitative image analysis and comparative descriptors, we report a chemical genomics approach coupled to semi-automated image acquisition and image analysis to identify chemical compounds with differential effects on endomembrane trafficking. Utilizing these compounds as a training set to develop our approach, we quantified their effects on vesicle traits such as size, velocity and straightness and intensity as quantitative descriptors. Furthermore, we developed a scoring system that permitted a comparison of differential intracellular dynamics induced by each of the

molecules, a critical step toward fully automated image characterization of endomembrane dynamics.

2.3 RESULTS

A pollen-specific chemical screen reveals eight classes of morphological and developmental phenotypes

To discover novel compounds that allow the dissection of endomembrane trafficking pathways and to identify a set of compounds to generate a training dataset for quantification, a chemical genomic screen was used to identify small molecules that affect pollen tube morphology and development. Because of their rapid growth and ease at which vesicle movement is visualized, *Nicotiana tobaccum* (tobacco) pollen was an ideal model system in which to perform this screen. Compounds were screened for those that inhibited germination or altered tube morphology, both of which are dependent upon proper tip-focused vesicle trafficking. In a developing pollen tube, vesicle traffic follows a distinct reverse fountain flow in which cargo moves towards the growing tip on the sides of the tube and cargo moving from the tip moves down the center of the tube (Pierson and Cresti, 1992). Several small molecules were found that disrupted this distinct trafficking pattern.

Previously, more than 46,000 compounds were screened in pollen at a concentration of 50-100 μ M (Robert *et al.*, 2011). Of these, 360 compounds were

found to cause morphological or developmental defects in pollen demonstrating their bioactivity (Robert et al. 2008). Small molecules were categorized as bioactive if they had an observable effect on morphology, development or localization of a fluorescent marker. A small molecule was considered to be bioactive if pollen grains did not germinate or there were defects in tube morphology such as isodiametric tip growth. In this current study, the 360 compounds were serially diluted to 25-50 μ M and 10-20 μ M, depending on the molecular weight of the compound, to determine the minimal effective concentrations and to uncover possible latent phenotypes (Figure 2.1). This permitted us to focus our screens on the resulting 225 compounds that inhibited germination and that remained bioactive at lower concentrations, thereby determining effective concentrations for potential compounds of interest (Supplemental Table1).

Identification of compounds that affect endomembrane trafficking

To test the effect of these 225 bioactive compounds on endomembrane trafficking in tobacco pollen, a secondary screen was performed using a transgenic tobacco line expressing RAB2:GFP in pollen (Cheung et al., 2002) to visualize the potential effects on RAB2 trafficking at two concentrations: 25-50 μ M and 10-20 μ M.

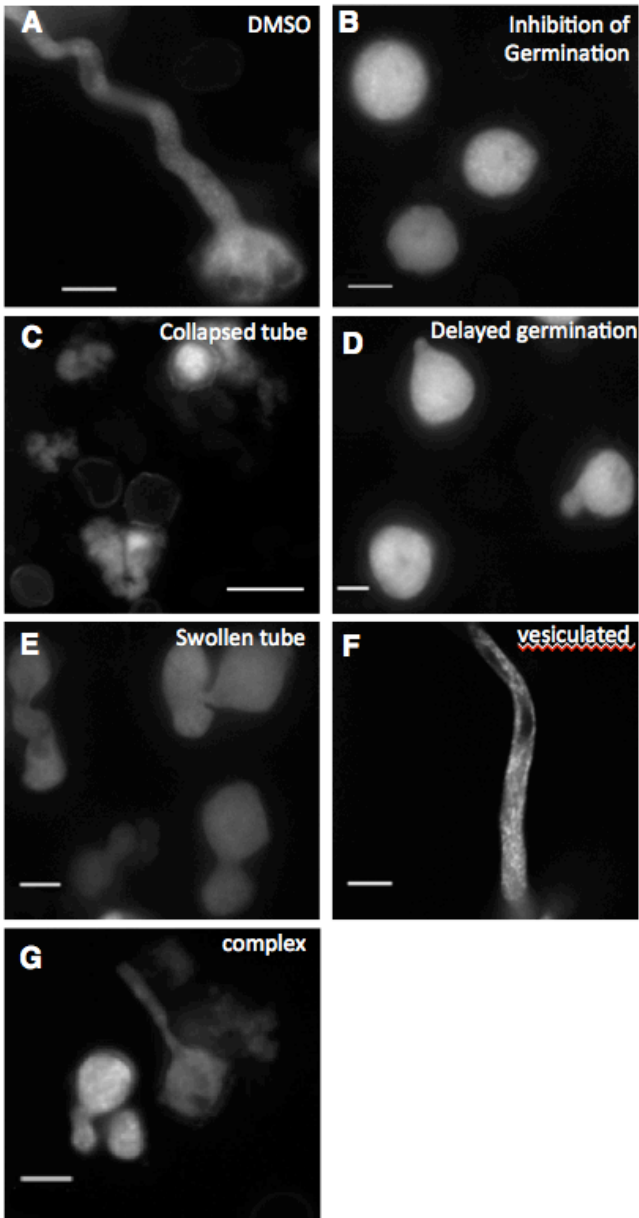


Figure 2.1 Morphological phenotypes of pollen expressing Rab2:GFP. (a) Control treatment treated with DMSO. (b) Inhibition of germination. (c) Collapsed pollen tubes. (d) Short tubes due to delayed growth. (e) Swollen pollen tubes. (f) Vesiculated tubes. (g) A complex phenotype in which multiple phenotypes are demonstrated. All images were taken using a 40x objective. Concentration of compounds ranged from 10 μ M to 50 μ M depending on the compound used. Scale bar = 20mm.

Phenotype	50-100 μ M	25-50 μ M	10-20 μ M
Germination Inhibition	112	76	37
Collapsed Tube	51	27	12
Delayed Growth	40	30	23
Swollen Tube	4	9	5
Vesiculated Tube	8	0	8
Complex Phenotype	27	14	13
TOTAL	290	156	98

Table 2.1. Phenotypes resulting from different concentrations of library compounds. Number of compounds resulting in different phenotypes are indicated.

Images were collected for each compound, analyzed using the Imaris image analysis software package and organized into groups that exhibited the same or similar phenotypes (Figure 2.1). Table 2.1 shows the number of compounds resulting in each phenotype observed. To focus our efforts on compounds that showed the most dramatic effects, a subgroup of six compounds were selected which caused disruptions in the localization of RAB2:GFP associated with

endomembranes (Figure 2.2). The compounds that comprised this subgroup were termed RAB2 effectors (RAEs). Compared to untreated controls, disruptions in the localization of RAB2:GFP were manifested as agglomerations (Figure 2.3.) which we termed “RAB2 bodies”. The bodies were presumed to be derived from Golgi stacks and ER. RAEs 6 and 7 elicited similar phenotypes, suggesting a similar mode of action and structure. Structure-activity relationship (SAR) analysis indicated that both RAE 6 and 7 have a nitro group along with a nitrile group next to a double bond. These compounds disrupted RAB2:GFP movement or localization and therefore, globally altered trafficking in the cell, allowing to measure these changes.

Thirty-second time lapse videos were taken of untreated pollen tubes and pollen tubes treated with the candidate compounds that affected RAB2:GFP localization or movement (Figure 2.4). Although the images of the automated confocal microscope were of lower quality compared to available manual confocal instruments, our eventual goal was to fully automate image capture and analysis; thus, we developed our approach utilizing the automated instrument. Quantitative comparison of the phenotypes induced by the RAEs demonstrated that the semi-automated method can be used to study endomembrane trafficking pathways by disrupting RAB2:GFP movement and localization. RAEs resulted in either moderately or drastically decreased movement.

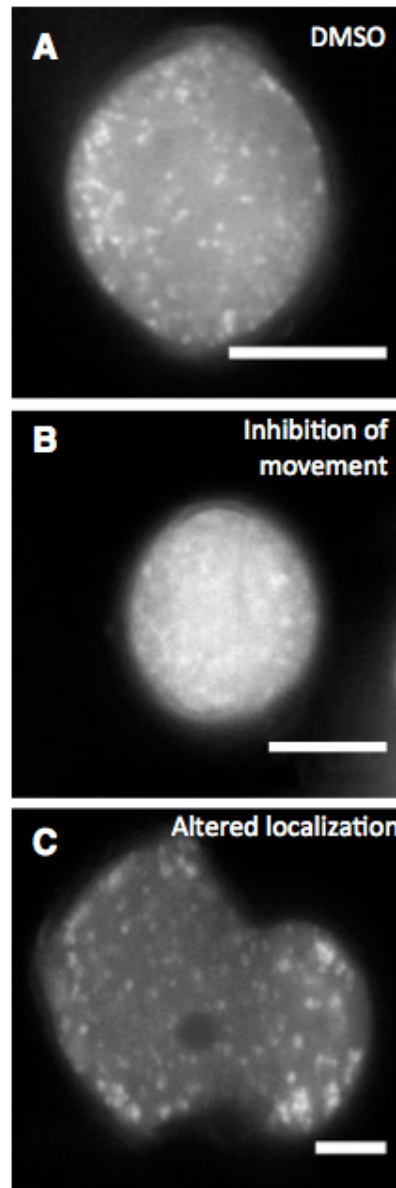


Figure 2.2. RAE's alter the localization and movement of RAB2:GFP. (a) Control DMSO treated tobacco pollen expressing RAB2:GFP. (b) RAB2:GFP localization in a pollen grain and tube following inhibition of trafficking by RAE2. (c) Tobacco pollen showing defects in normal localization of RAB2:GFP caused by treatment with RAE7 . Images are taken using a 63x objective. Scale bar = 10mm.

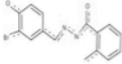
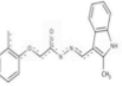
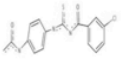
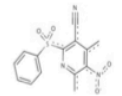
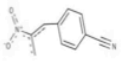
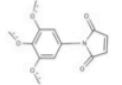
Compound	Structure	Average area of RAB2 bodies (um ²)	% Area of WT RAB2 Bodies	Average velocity of RAB2 bodies (um/s)	% Velocity of WT of RAB2 bodies	Straightness	% Straightness of WT	% Intensity of WT	Difference Index
Untreated	No compound	1.926	100	0.911	100	0.525	100	100	1.00
RAE1 (5271226) Chembridge		4.400	228	0.289	31.7	0.313	60	137	1.88
RAE2 (532951) Chembridge		4.247	221	0.308	33.8	0.532	101	160	2.14
RAE5 (6396311) Chembridge		5.155	268	0.109	11.9	0.122	23	111	2.12
RAE 6 (LAT033E05) LATCA		7.372	383	0.046	5.0	0.107	20	191	4.59
RAE7 (LAT035F05) LATCA		7.730	401	0.048	5.2	0.105	20	218	5.22
RAE8 (LAT045A04) LATCA		6.991	363	0.068	7.4	0.077	15	207	4.37

Table 2.2. Quantitative values describing the effect of each RAE. The name and Identification number of each compound is given along with its structure, the net a relative area, velocity, straightness and intensity of the RAB2 bodies created by it, ultimately culminating in a difference index score.

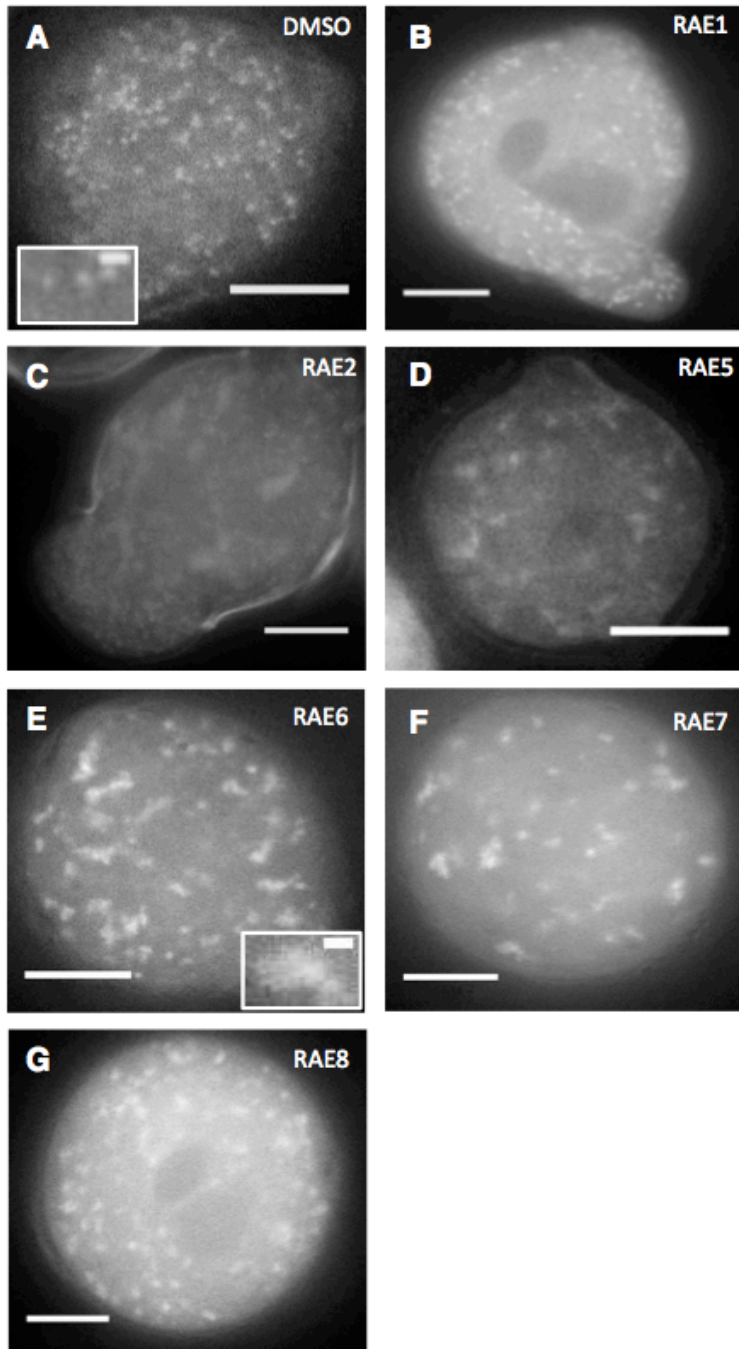


Figure 2.3. RAE phenotypes. RAB2:GFP in pollen treated with (a) DMSO, (b) 36 μ M RAE1, (c) 38 μ M RAE2, (d) 35 μ M of RAE5, (e) 78 μ M RAE6, (f) 32 μ M of RAE7, (g) 28 μ M of RAE7. All images were taken using a 63x objective. Scale bar = 10 μ m. Insets indicate zoomed image of the respective labeled body. Scale bar within inset = 1 μ m.

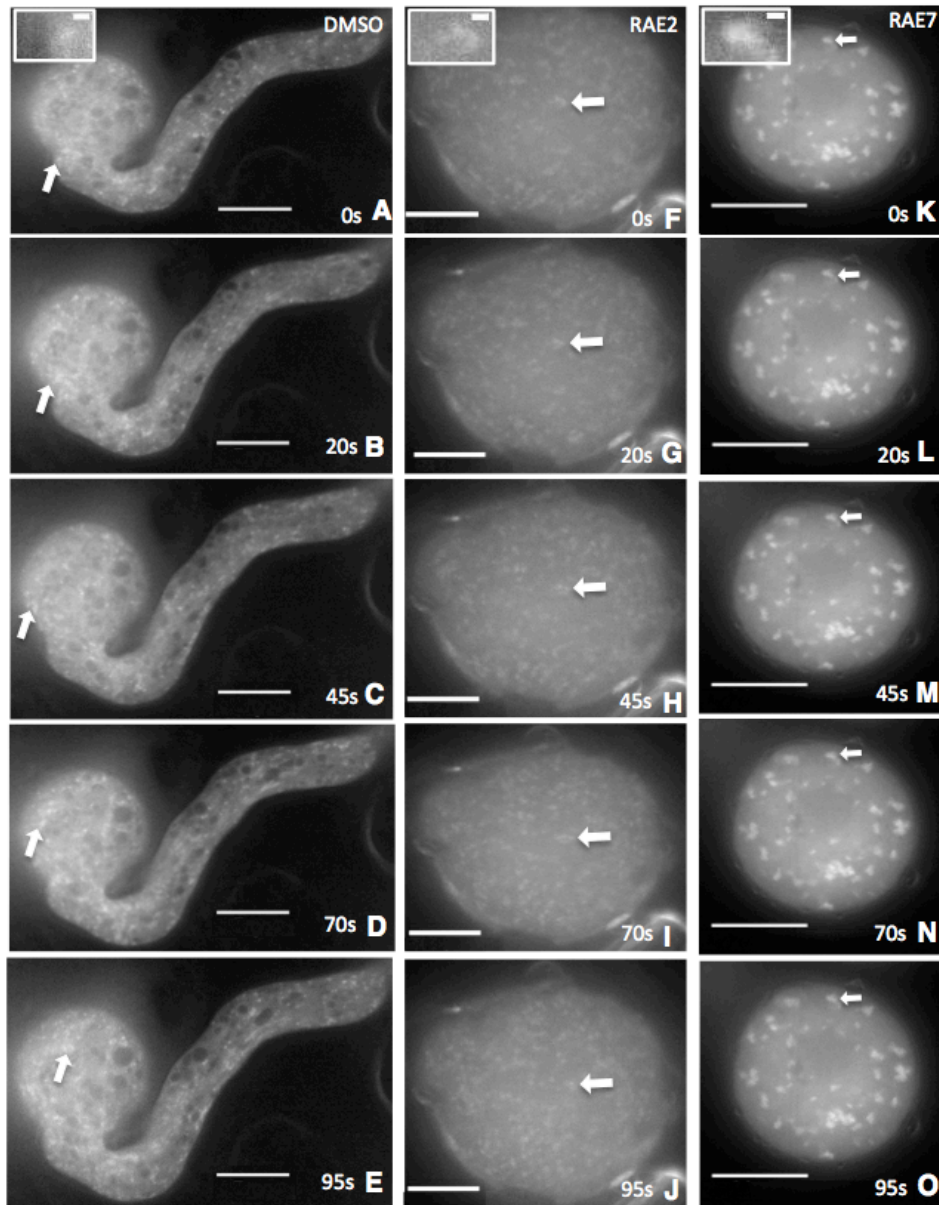


Figure 2.4. Movement of RAB2:GFP is inhibited by RAEs. (a-o) Selected frames from time-lapse videos of pollen tubes. (a-e) DMSO treated RAB2:GFP labeled pollen. Scale bar = 15 μ m. (f-j) RAE2 treated RAB2:GFP labeled pollen. Scale bar = 10 μ m. (k-o) RAE7 treated RAB2:GFP labeled pollen. Scale bar = 20 μ m. Phenotypes are shown at 0, 20, 45, 70, and 95 s after acquisition. Arrows denote the position of the same RAB2 body in each frame. Insets indicate zoomed image of the respective labeled body. Scale bar within inset = 1 μ m.

Quantification of RAB2 bodies reveals distinct groupings and latent effects

The observed phenotypes had varying degrees of severity, which could not be sufficiently described using qualitative descriptors. We, therefore, quantified the area, velocity, straightness of movement, and intensity of the RAB2:GFP agglomerations using the segmentation function in the Imaris image analysis software package. Values were quantified as means (Table 2). However, determining the mean values of RAB2 body measurement in pollen tubes was insufficient to fully understand this system. Since both the Golgi bodies and the RAB2 bodies were highly dynamic and displayed a broad range of sizes, movements, movement patterns and intensities, we presented the area, velocity, straightness, and intensity measurements as cumulative distribution plots (Sparkes et al., 2008) to better describe organelle movement. By permitting a perspective on the variation, these plots provided a more informative comparison of organelle movement (Figure 5). The cumulative distribution plot of the areas indicates a trend that is also inferred by the mean values. In the case of RAEs 1 and 2, the localization of RAB2:GFP was altered causing the formation of RAB2 bodies that were 228% and 221% larger than control Golgi stacks, respectively. RAEs 1 and 2 also had a range of velocity values between approximately 0.03 $\mu\text{m/s}$ and 1.5 $\mu\text{m/s}$, which was about half the range of the untreated Golgi bodies (Figure 2.5A).

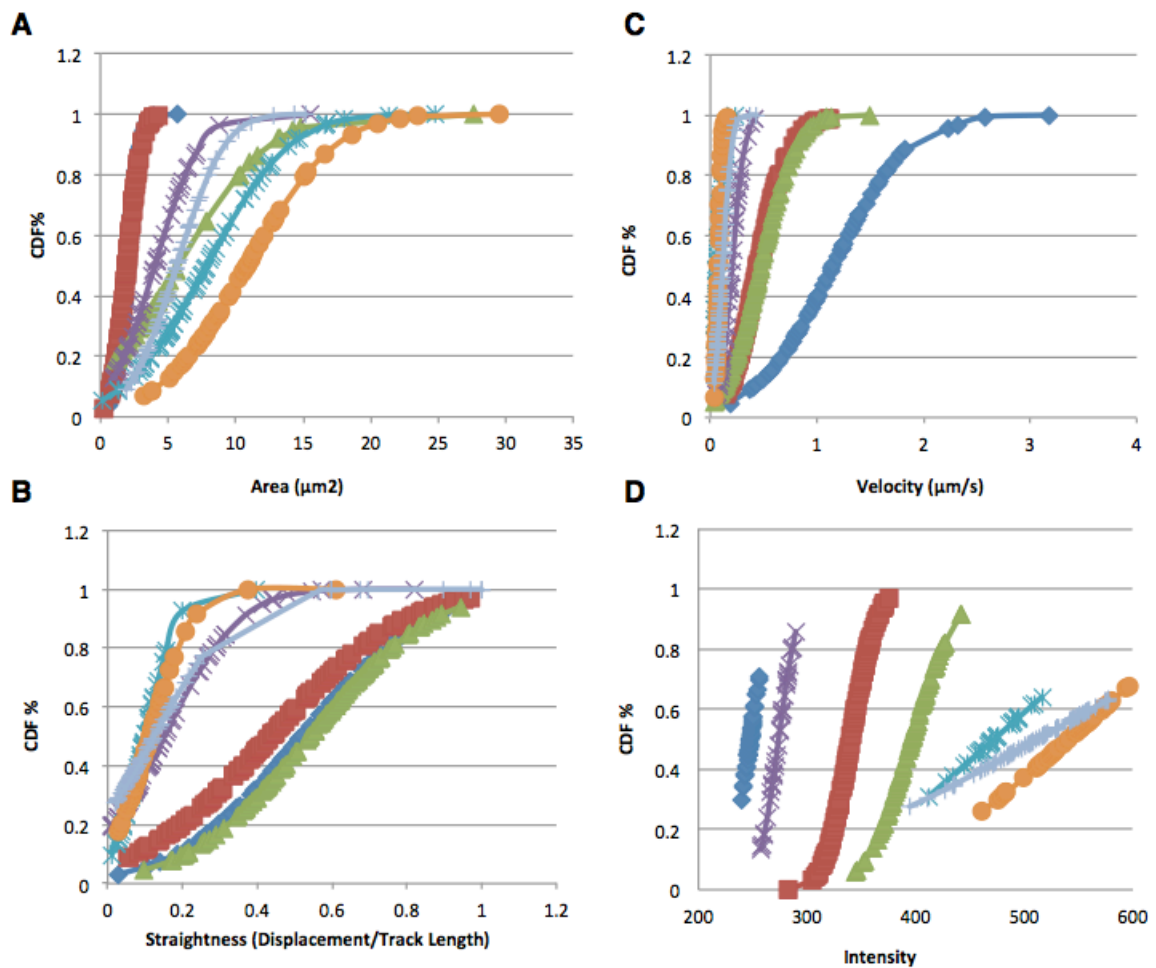


Figure 2.5. Cumulative distribution function (CDF) of the four features measured, area, velocity, straightness, and intensity of all six RAEs plus an untreated control. (a) The CDF for the average area. (b) The CDF for the average velocity. (c) The CDF for average straightness. (d) The CDF for the average intensity. In all plots, blue diamonds represent the Untreated control. Red squares represent RAE1. Green triangles represent RAE2. Purple X's represent RAE5. Teal asterisks represent RAE6. Orange circles represent RAE7. Gray Crosses represent RAE8. All bodies measured were within the pollen grain.

RAEs 5, 6, 7 and 8 resulted in RAB2 bodies that ranged from 268% to 401% larger than the mean of untreated Golgi bodies. They also exhibited lower variance (Figure 56b), indicating that the individual agglomerations were of similar size.

Changes in velocity were also apparent. Initially, RAEs 1 and 2 resulted in slightly reduced speed of the RAB2 bodies, whereas RAEs 5, 6, 7, and 8 dramatically reduced speed. Similar to area measurements, velocity was measured and reported as a percentage of the velocity of the untreated Golgi bodies (Figure 2.5B). RAEs 1 and 2 moved at only 31% and 37% of the velocity of the untreated Golgi bodies, respectively. The quantification of RAEs 5, 6, 7, and 8 supported our initial observation of a dramatic reduction in speed. The RAB2 bodies produced by treatment with the RAEs traveled between 11.9% and 5% of the velocity of untreated Golgi bodies. Upon quantification of the data, we noted that the compounds formed natural groups of weak effectors (RAEs 1 and 2) and strong effectors (RAEs 5, 6, 7 and 8).

Beyond the speed of movement, the pattern of movement was a key component of dynamic organelle behavior. A robust descriptor of movement pattern was the straightness of trajectory of each body. The straightness was determined by the ratio of displacement over the track length. Values closer to one indicated a tendency towards linear movement, whereas those farther from one indicated movements resembling short-distance random or Brownian motion

(Figure 5C). Similar to our previous results, there was a distinct grouping and similarity between RAEs 1 and 2, which had greater values for straightness. This indicated more directed movement, whereas RAEs 5, 6, 7, and 8 stimulated a tendency towards random motion.

Lastly, we quantified the intensity of RAB2:GFP fluorescence (Figure 2.5D). The untreated pollen showed the lowest RAB2:GFP intensity followed closely by RAE 5. RAEs 1 and 2 resulted in an intermediate level of intensity, 137% and 160% of the untreated, respectively. RAEs 6, 7 and 8 resulted in the highest levels of intensities ranging from 191% to 218% of untreated RAB2:GFP intensity levels.

It stands to reason that some of these parameters are most likely correlated. To confirm our hypothesis, the Pearson correlation coefficients for every parameter combination were calculated (Figure 2.6). Area was positively correlated with intensity ($R = .715$) suggesting that as the amount of RAB2:GFP increased with the size of RAB2 bodies. As to be expected, straightness was negatively correlated with velocity ($R = .569$). Bodies traveling in a straight line moved faster than those exhibiting Brownian motion. Less obviously, intensity and straightness were negatively correlated ($R = .571$). In other words, vesicles that moved in a more directed manner tended to have lower intensities. All other parameters showed little to no correlation.

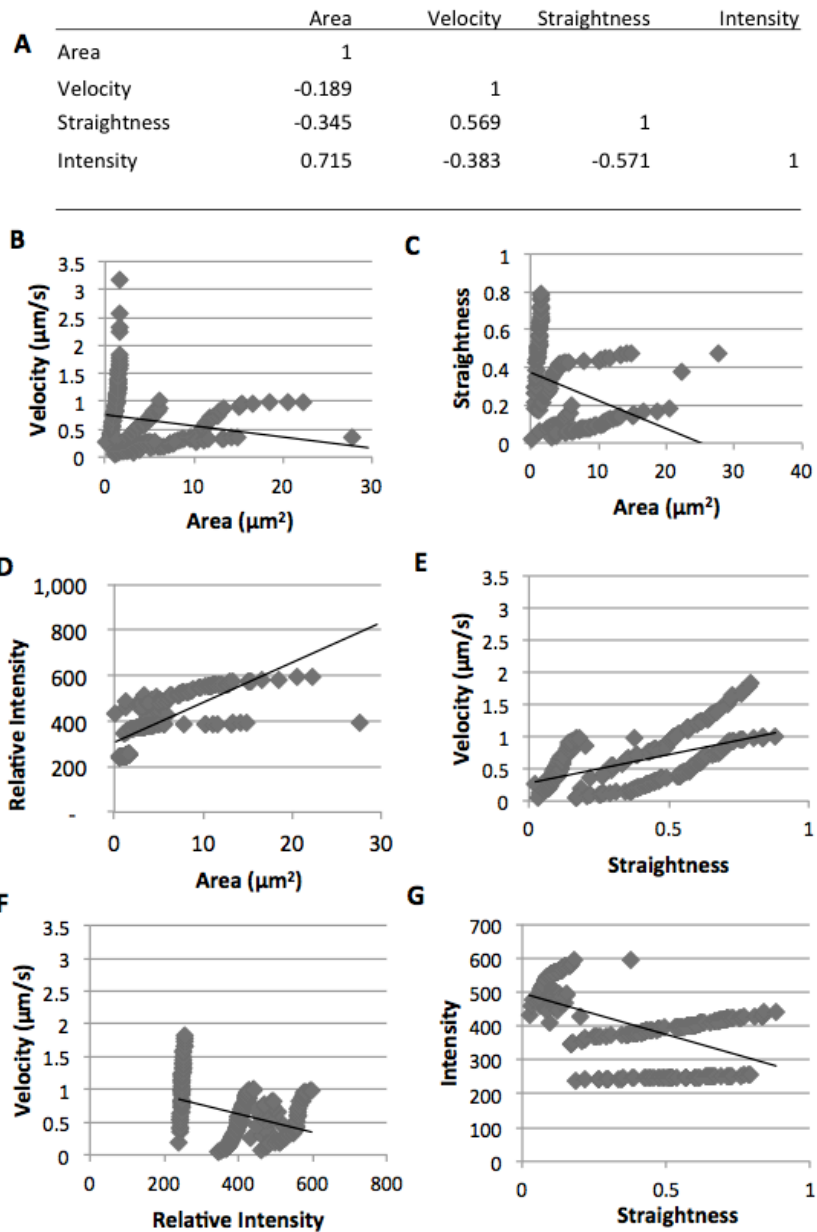


Figure 2.6. Correlation of trafficking features. (a) A correlation matrix showing the Pearson correlation coefficients showing the relationships of all features. (b) The correlation between velocity and area. (c) The correlation between straightness and area. (d) The correlation between intensity and area. (e) The correlation between velocity and straightness. (f) The correlation between velocity and intensity. (g) The correlation between intensity and straightness.

The Difference score provides a quantitative measure of bioactivity

To understand the relative global effects of the RAEs on trafficking, we developed a quantitative score that incorporated all the descriptors of RAB2:GFP dynamics. Our “difference index” was developed to give an integrated quantitative value to describe the degree to which each compound affected the area, velocity straightness, and intensity of each agglomeration (Table 2.2). This straightforward index was the sum of the squares of the normalized value for each measure multiplied by an assigned weighting factor. Because it was difficult to determine if one feature might play a larger role in the identity of a compound, each feature was given equal weight. However, our approach can permit alterations in the weighting of particular phenotypes as necessary.

2.4 Discussion

We utilized tobacco pollen as an efficient system to screen for chemicals that inhibit endomembrane trafficking. In a previous study (Drakakaki, et al., 2011), more than 46,000 chemicals were screened in tobacco pollen. Of these compounds, 360 were found to be bioactive, affecting morphology or development of the pollen tube. Our goal was to utilize this screening system to identify compounds among the 360 that would perturb the secretory pathway in pollen. We reasoned that the tracking of Golgi bodies that were perturbed in movement and other descriptors would offer an ideal model to develop a video imaging

approach to quantify and compare the dynamics of trafficking. Golgi compartments are relatively easy to visualize and are mobile within the pollen grain and tube in this rapidly growing organism. In the past, imaging fixed tissue and single still frame images have limited quantitative studies of endomembrane dynamics. The development of powerful computational tools such as commercially available Imaris image analysis software packages have allowed for the quantification of trafficking parameters in living pollen tubes over time, providing a more accurate measure of *in planta* conditions.

Six compounds were identified that altered the localization and/or movement of RAB2:GFP. Upon quantification of size, velocity, straightness and intensity, the compounds RAE1,2,5,6,7 and 8 formed two phenotypic classes. The first group caused modest changes in size, velocity, straightness, and intensity, whereas the second caused dramatic changes in these descriptors. The first group comprised of RAEs 1,2 and 5 resulted in smaller agglomerations whose movement was modestly reduced compared to the untreated controls. This was reflected in their difference indices, which were 1.88 and 2.14, and 1.12 respectively. Because this is a relative index, the score of the untreated pollen is 1.00. Greater deviations from 1.00 indicate greater relative differences in dynamics compared to the untreated condition. Interestingly, RAE 5 initially grouped with RAEs 6, 7 and 8 but received a score of 2.12, which grouped it with the weaker compounds RAEs 1 and 2. More specifically, whereas RAE 5 was similar to RAEs 6 and 7 in terms of percent

velocity and percent straightness of untreated RAB2 Golgi bodies, it was much similar to RAEs 1 and 3 and terms of percent area of normal RAB2 Golgi bodies and essentially unchanged in percent intensity of normal (Table 2). RAE5 RAB2 bodies were of similar size and intensity as the untreated bodies, but were less mobile and directional.

The second group consisted of RAEs 6, 7 and 8 which gave rise to relatively large agglomerations, drastically reduced movement and directionality, and greater intensity. Their difference index scores of 4.59, 5.22 and 4.37, respectively, indicated that these compounds resulted in overall endomembrane phenotypes with major differences compared to untreated pollen. This type of analysis indicates the value of quantification of vesicle dynamics in a non-biased manner, permitting more precise phenotype classification.

These differences in the degree of effect could be caused by a variety of mechanisms of which we could only speculate here. RAEs 1, 2 and 5 could be targeting proteins that have a minor role in regulating the localization of RAB2:GFP and therefore only cause a minor defect, whereas RAEs 6, 7 and 8 may be targeting proteins that play a major role in regulation of RAB2:GFP localization and endomembrane trafficking thereby causing a dramatic defect. Alternatively, RAEs 1, 2 and 5 may effect trafficking via reversible target binding, whereas RAEs 6, 7 and 8 could bind irreversibly to their target(s) resulting in stronger phenotypes.

Interestingly, RAEs 6 and 7, two of the more potent compounds have a nitrile and nitro group in common. The nitrile group can undergo a nucleophilic reaction that is facilitated by the nitro group to irreversibly bind with proteins (Beck, 1978; Habibi et al., 2011.). Likewise, RAE 8 contains a functional group containing N-ethylmaleimide (NEM). NEM is a Michael acceptor, which adds nucleophiles such as the thiols in cysteine residues and is a commonly used cysteine modifier (Crankshaw and Grant, 2001) which could lead to irreversible target binding. This could very well be the case with RAE8. In fact, soluble N-ethylmaleimide-sensitive factor adaptor protein receptors (SNAREs) are by definition sensitive to NEM suggesting that the target of RAE 8 might conceivably be a specific SNARE (Söllner et al., 1993). Because RAE 8 is predicted to be an irreversible inhibitor, Multidimensional Protein Identification Technology (MudPIT) would be a promising approach for target identification. However, this is beyond our scope to develop an image analysis approach to quantify endomembrane dynamics. Although it is intriguing to consider the reversibility and targets of RAEs, this is beyond our goal of developing a technically robust video-imaging tool to quantify trafficking.

This approach can be used for the rapid characterization and management of small molecules. This method has the ability to streamline chemical genetic screens and aid in experimental design. It is important to note that other automatic or semi automatic image acquisition modalities can be used as well as freely

available image analysis software. When analyzing data the limitations of scale are determined by the resolution of the image and the frame rate at which the video data were collected, not necessarily the image analysis software. One advantage of our approach to classify small molecules is the difference score. This will help facilitate the classification of compounds into meaningful clusters based on bioactivity as a quantitative measure of vesicle dynamics (Drakakaki et al., 2011). Though each feature value was given equal weight, one could increase the weight of any given feature, tailoring this score to any application. However, the difference index can only take into account phenotypes that can be quantified and the number of features measured determines the usefulness and flexibility. In addition, the score is highly dependent on the control phenotype and can only describe changes in preexisting conditions. It is not helpful in describing new phenotypes that were not represented in the control condition.

Among the six RAEs, 6 and 7 are the only previously characterized compounds. RAE 6, also known as 3-Pyridinecarbonitrile, 4,6-dimethyl-5-nitro-2-(phenylsulfonyl), is an inhibitor of severe acute respiratory syndrome coronavirus (SARS-CoV), a main protease (M^{pro}) which is essential to the life cycle of the SARS-CoV (Lu et al., 2006). This suggests that the target in tobacco might be a protease. RAE 7, a substituted pyrrole derivative, is an androgen receptor inhibitor, suggesting that its target in tobacco might be a hormone or other type of receptor

(Matsunaga 2006). Analyzing the known effects of RAE 6 and RAE7 in mammalian systems could shed some light to their possible targets in pollen.

Overall, our study has contributed several important advances in the evaluation of vesicle trafficking dynamics: 1. We have focused on the quantification of video images to capture true real time dynamics *in vivo*; 2. We examined the correlations of individual parameters such velocity and straightness of movement; 3. The summary value score permitted a quantitative and comparative measure of overall vesicle dynamics; 4. The quantification of dynamic behaviors will make it possible to move toward the increased automation of chemical and genetic screens that can be focused directly on changes in vesicle dynamics, permitting the more efficient linkage of cellular and developmental phenotypes.

Quantitative analyses such as these provide a deeper understanding of highly dynamic intracellular movement by providing data that is objective. The difference index gives a single quantitative value that can immediately give insight into the degree of compound effects. A subset of compounds acts transiently, causing a range of phenotypes that vary in degree and duration. Excluding temporal changes excludes potentially important data. Our approach allows quantification and observation of temporal dynamics that occur as a result of transient binding of small molecules.

Full automation of image analysis and quantification of vesicle dynamics will become essential with increased throughput and the dramatically larger resultant image datasets. Quantitative descriptions such as these will facilitate the development of realistic predictive mathematical models of endomembrane trafficking in response to small molecules. Quantitative analysis allows for the identification of the dominant biological forces that contribute to a given biological phenomenon in addition to focusing the biological questions that are subsequently asked (Phillips and Milo, 2009).

Chemical genomics, though still a young field, is starting to gain popularity for dissecting dynamic biological processes. Quantitative analysis will extract comparative information for the elucidation of endomembrane trafficking pathways and their effects on organism development. This connection will allow understanding of the developmental consequences of changes in endomembrane dynamics at the cellular level.

2.5 Methods

Semi-Automated Image Based Screen

Chemical libraries were screened as indicated previously to derive a set of 360 compounds inhibiting pollen germination (Drakakaki et al, 2011). For the present study, chemical stocks were in 100% DMSO and distributed in 96 well clear-

bottom plates at concentrations of 5mg/ml (50-100 μ M), 0.5mg/ml (25-50 μ M), and 0.05mg/ml (10-20 μ M). For the primary screen wild type *Nicotiana tobaccum* pollen was harvested at dehiscence. Thirty anthers from six flowers were harvested and allowed to germinate in 14ml of GM2 medium (18% sucrose, 0.01% boric acid, 5 μ M CaCl₂, 5 μ M Ca(NO₃)₂, and 1 mM MgSO₄ (pH 6.5–7.0). Pollen was then vortexed to release the pollen grains from the anthers. The pollen suspension was added to the 96 well clear-bottom plates and incubated at room temperature in the presence of the appropriate compound on a covered orbital shaker for 3 hours. Images of pollen tubes were acquired using the BD Pathway automated microscope (BD Biosciences) with the 40x and 63x objective. The sample was imaged using 470nm/30 excitation filters and 520nm/30 for GFP.

A custom semi-automated program was created for a 96 well plate using autofocus for the primary screen. Bright field Images were then viewed and screened for defects in morphology as previously described (Robert et al., 2008). The compounds that were found to cause defects in pollen development were screened again using transgenic tobacco pollen expressing the RAB2:GFP fusion protein under the LAT52 pollen-specific promoter (Cheung et al., 2002). Still images as well as 4-5 second time-lapse movies containing 20 frames were captured over a period of approximately 2 minutes. Compounds were screened based on their effect on RAB2:GFP localization and movement. Of the 290 compounds that initially caused defects in pollen tube development, six were

found to consistently affect the localization and movement of RAB2:GFP and were termed RAEs. RAE1 (Benzoic acid, 2-methyl-, 2-[(3-bromo-4-hydroxyphenyl)methylene] hydrazide, Chembridge 5271226) and RAE 2 (Acetic acid, 2-(2-methylphenoxy)-, 2-[(2-methyl-1H-indol-3-yl)methylene]hydrazide, Chembridge 5532951) which both caused small agglomerations and decreased movement of the RAB2 bodies. RAE 5(Benzamide, N-[[[4-(acetylamino)phenyl]amino]thioxomethyl]-3-chloro-, Chembridge no. 6396311), RAE 6 (3-Pyridinecarbonitrile, 4,6-dimethyl-5-nitro-2-(phenylsulfonyl), LATCA no. LAT033E05), RAE 7 (Benzonitrile, 4-[(1E)-2-nitro-1-propen-1-yl]-, LATCA no. LAT035F05), and RAE 8 (1H-Pyrrole-2,5-dione, 1-(3,4,5-trimethoxyphenyl), LATCA no. LAT045A04) caused larger bodies and reduced movement to a greater degree. All confocal images were taken at the same gain and magnification as to not artificially increase or decrease the intensity.

Image analysis

Images were analyzed using Imaris (version 7.3.1, Bitplane, South Windsor, CT) with the 'spots' software package. RAB2:GFP bodies were quantified using the automatic threshold function provided by the 'spots' algorithm builder. The number of bodies measured per pollen grain ranged from 66 to 220. The mask value used was between 13 and 14, which determined the regions that the algorithm would

measure. Segmentation along with region growing was used to quantify the fluorescent RAB2:GFP bodies. The average area of three replicates was recorded in μm^2 . Three biological replicates were used in quantifying the average speed of RAB2:GFP bodies using the 'spots' tracking function. The automatic settings were used for creating the segmentation algorithm. Because of the small size of the bodies measured, the maximum distance traveled between frames was $2\mu\text{m}$. Once the bodies were identified and tracked, an additional filter included in the spots package was used to exclude those bodies that were unable to be tracked more than 15 frames. Area, Velocity, straightness and intensity were represented as cumulative distribution plots. The plots were developed by determining the average of each parameter for each spot over three replicates. The mean and standard deviation for all spots in a movie were calculated and used in determining the z score. This z score also called the CDF% was plotted against the average of the selected parameter.

From these average values, a difference index was created. The Following equation was used to calculate the difference score:

$$\omega_1 a^2 + \omega_2 b^2 + \omega_3 c^2 + \omega_4 d^2$$

Where a is the area of the organelles, or in this case RAB2 bodies, b is the average velocity of all RAB2 bodies, c is the average intensity of all RAB2 bodies and d is the average straightness of all rab2 bodies. Each normalized value was given an

equal weight of .25. The sum of the squares of the normalized values was then found for each feature. The resulting number is the difference index.

2.6 References

Beck, M., Zhou, J., Faulkner, C., MacLean, D. and Robatzek, S., (2012) Spatio-temporal cellular dynamics of the Arabidopsis flagellin receptor reveal activation status-dependent endosomal sorting. *Plant Cell*, 24(10), 4205-19.

Bensch, R., Ronneberger, O., Greese, B., Fleck, C., Wester, K., Hulskamp, M. and Burkhardt, H., (2009) Image analysis of arabidopsis trichome patterning in 4D confocal datasets. In *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*. p 742-745.

Bhanu, Bir. "IGERT Program." The UC Riverside Integrated Graduate Education Research and Training Program in Video Bioinformatics. University of California, Riverside, 26 Oct. 2009. Web. 24 Mar. 2014.

Brandizzi, F., Fricker, M. and Hawes, C., (2002) A greener world: the revolution in plant bioimaging. *Nat Rev Mol Cell Biol*, 3(7), 520-30.

Carlsson, K., Danielsson, P.-E., Liljeborg, A., Majlöf, L., Lenz, R. and Åslund, N., (1985) Three-dimensional microscopy using a confocal laser scanning microscope. *Optics letters*, 10(2), 53-55.

Cheung, G. and Cousin, M.A., (2011) Quantitative analysis of synaptic vesicle pool replenishment in cultured cerebellar granule neurons using FM dyes. *J Vis Exp*, (57).

Collinet, C., Stöter, M., Bradshaw, C.R., Samusik, N., Rink, J.C., Kenski, D., Habermann, B., Buchholz, F., Henschel, R., Mueller, M.S., Nagel, W.E., Fava, E., Kalaidzidis, Y. and Zerial, M., (2010) Systems survey of endocytosis by multiparametric image analysis. *Nature*, 464(7286), 243-9.

Domozych, D.S., (2012) The quest for four-dimensional imaging in plant cell biology: it's just a matter of time. *Ann Bot*, 110(2), 461-74.

Drakakaki, G., Robert, S., Szatmari, A.M., Brown, M.Q., Nagawa, S., Van Damme, D., Leonard, M., Yang, Z., Girke, T., Schmid, S.L., Russinova, E., Friml, J., Raikhel,

N.V. and Hicks, G.R., (2011) Clusters of bioactive compounds target dynamic endomembrane networks in vivo. *Proc Natl Acad Sci U S A*, 108(43), 17850-5.

Eils, R. and Athale, C., (2003) Computational imaging in cell biology. *J Cell Biol*, 161(3), 477-81.

Goldberg, D.E. and Holland, J.H., (1988) Genetic algorithms and machine learning. *Machine learning*, 3(2), 95-99.

Hicks, G.R. and Raikhel, N.V., (2009) Opportunities and challenges in plant chemical biology. *Nat Chem Biol*, 5(5), 268-72.

Hua, S. and Sun, Z., (2001) Support vector machine approach for protein subcellular localization prediction. *Bioinformatics*, 17(8), 721-728.

Illingworth, J. and Kittler, J., (1987) The adaptive Hough transform. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (5), 690-698.

Konopka, C.A. and Bednarek, S.Y., (2008) Variable-angle epifluorescence microscopy: a new way to look at protein dynamics in the plant cell cortex. *Plant J*, 53(1), 186-96.

Kuehn, M., Hausner, M., Bungartz, H.J., Wagner, M., Wilderer, P.A. and Wuertz, S., (1998) Automated confocal laser scanning microscopy and semiautomated image processing for analysis of biofilms. *Appl Environ Microbiol*, 64(11), 4115-27.

de Lange, F., Cambi, A., Huijbens, R., de Bakker, B., Rensen, W., Garcia-Parajo, M., van Hulst, N. and Figdor, C.G., (2001) Cell biology beyond the diffraction limit: near-field scanning optical microscopy. *Journal of cell science*, 114(23), 4153-4160.

Linda G. Shapiro and George C. Stockman (2001): "Computer Vision", pp 279-325, New Jersey, Prentice-Hall, ISBN 0-13-030796-3

Liu, J., Elmore, J.M., Lin, Z.J. and Coaker, G., (2011) A receptor-like cytoplasmic kinase phosphorylates the host target RIN4, leading to the activation of a plant innate immune receptor. *Cell Host Microbe*, 9(2), 137-46.

Liu, K., Schmidt, T., Blein, T., Durr, J., Palme, K. and Ronneberger, O., (2013) Joint 3d cell segmentation and classification in the arabidopsis root using energy minimization and shape priors. In *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*. p 422-425.

Meyer, A.J. and Fricker, M.D., (2000) Direct measurement of glutathione in epidermal cells of intact Arabidopsis roots by two-photon laser scanning microscopy. *Journal of microscopy*, 198(3), 174-181.

Miart, F., Desprez, T., Biot, E., Morin, H., Belcram, K., Höfte, H., Gonneau, M. and Vernhettes, S., (2013) Spatio-temporal analysis of cellulose synthesis during cell plate formation in Arabidopsis. *Plant J*.

Nakano, A., (2002) Spinning-disk confocal microscopy--a cutting-edge tool for imaging of membrane traffic. *Cell structure and function*, 27(5), 349-355.

Peterson, R.C. and Wolffsohn, J.S., (2007) Sensitivity and reliability of objective image analysis compared to subjective grading of bulbar hyperaemia. *Br J Ophthalmol*, 91(11), 1464-6.

Phillips, R. and Milo, R., (2009) A feeling for the numbers in biology. *Proc Natl Acad Sci U S A*, 106(51), 21465-71.

Racine, V., Sachse, M., Salamero, J., Fraisier, V., Trubuil, A. and Sibarita, J.B., (2007) Visualization and quantification of vesicle trafficking on a three-dimensional cytoskeleton network in living cells. *J Microsc*, 225(Pt 3), 214-28.

Rajadhyaksha, M., Anderson, R. and Webb, R.H., (1999) Video-Rate Confocal Scanning Laser Microscope for Imaging Human Tissues *In Vivo*. *Applied Optics*, 38(10), 2105-2115.

Robert, S., Chary, S.N., Drakakaki, G., Li, S., Yang, Z., Raikhel, N.V. and Hicks, G.R., (2008) Endosidin1 defines a compartment involved in endocytosis of the brassinosteroid receptor BRI1 and the auxin transporters PIN2 and AUX1. *Proc Natl Acad Sci U S A*, 105(24), 8464-9.

Rust, M.J., Bates, M. and Zhuang, X., (2006) Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature methods*, 3(10), 793-796.

Salomon, S., Grunewald, D., Stüber, K., Schaaf, S., MacLean, D., Schulze-Lefert, P. and Robatzek, S., (2010) High-Throughput Confocal Imaging of Intact Live Tissue Enables Quantification of Membrane Trafficking in Arabidopsis. *Plant physiology*, 154(3), 1096.

Sampathkumar, A., Gutierrez, R., McFarlane, H.E., Bringmann, M., Lindeboom, J., Emons, A.M., Samuels, L., Ketelaar, T., Ehrhardt, D.W. and Persson, S., (2013) Patterning and lifetime of plasma membrane-localized cellulose synthase is dependent on actin organization in Arabidopsis interphase cells. *Plant Physiol*, 162(2), 675-88.

Sankar, M., Nieminen, K., Ragni, L., Xenarios, I. and Hardtke, C.S., (2014) Automated quantitative histology reveals vascular morphodynamics during Arabidopsis hypocotyl secondary growth. *eLife*, 3.

Saxton, M.J. and Jacobson, K., (1997) Single-particle tracking: applications to membrane dynamics. *Annual review of biophysics and biomolecular structure*, 26(1), 373-399.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P. and Cardona, A., (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods*, 9(7), 676-82.

Sethuraman, V., Taylor, S., Pridmore, T., French, A. and Wells, D., (2009) Segmentation and tracking of confocal images of *Arabidopsis thaliana* root cells using automatically-initialized Network Snakes. In *Bioinformatics and Biomedical Engineering, 2009. ICBBE 2009. 3rd International Conference on*. p 1-4.

Swedlow, J.R. and Eliceiri, K.W., (2009) Open source bioimage informatics for cell biology. *Trends Cell Biol*, 19(11), 656-60.

Tataw, O.M., Liu, M., Roy-Chowdhury, A., Yadav, R.K. and Reddy, G.V., (2010) Pattern analysis of stem cell growth dynamics in the shoot apex of *arabidopsis*. In

Image Processing (ICIP), 2010 17th IEEE International Conference on. p 3617-3620.

Ung, N., Brown, M.Q., Hicks, G.R. and Raikhel, N.V., (2012) An Approach to Quantify Endomembrane Dynamics in Pollen Utilizing Bioactive Chemicals. *Mol Plant.*

Chapter 3

EndoQuant: An Image Analysis Package for Automated Quantitative Cellular Phenotyping

3.1 Abstract

As microscopic imaging technology continues to improve, the ability to quantify observations using computational techniques will greatly increase its utility. The need for increases in assay throughput will require automatic tools to quantify and interpret phenotypes at the level of cellular and subcellular structures. EndoQuant, short for endomembrane quantification tool, is a novel user-friendly image analysis tool for automatic quantification of discrete endomembrane phenotypes from multiple images. It measures features of structures within each cell resulting in the computation of statistics on a per cell basis, giving a biologically powerful metric. Additionally, one can process a large volume of datasets or compare any two images quantitatively. After the analysis is complete, biologists can immediately visualize both data and computed information in subsequent graphs allowing the extracted features to be easily visualized as bar graphs for each cell as well as feature histograms for population analysis. When comparing two images, any two features can be directly compared in a single graph including their standard error bars and as a box plot showing basic statistics including statistical outliers. This

method is applicable to multiple tissue types assuming the cell borders can be properly defined. EndoQuant identifies a mean of 83% true positives and maintains a low false positive rate while reducing processing time by more than 100-fold. EndoQuant is a rapid, accurate and user-friendly automated phenotyping tool that will permit rapid quantitative cellular phenotype analyses and facilitate important biological discovery.

3.2 Introduction

A forward genetics approach for gene discovery requires intensive screening and analysis of large quantities of data, regardless of experimental method (Mayer, 1999; Stockwell, 2000). Indeed, forward genetic screens are often designed to reduce the labor needed to identify individuals of interest by using an obvious macroscopic qualitative or quantitative phenotype as a discriminatory marker (Page and Grossniklaus, 2002). However, when investigating cellular phenomena, cellular phenotypes may not necessarily manifest into obvious and discrete macroscopic phenotypes. Additionally, cellular mechanisms cannot be fully understood until an analysis of microscopic phenomena is incorporated into the forward genetics approach. Imaging and manually analyzing every potentially interesting individual can be highly inefficient, wasting time and resources. Therefore, we have taken advantage of the recent advancements in the computer

vision and image analysis to develop a user-friendly tool which is complete with Graphical User Interface (GUI) to automatically quantify and classify endomembrane features within plant cells. The past years have been marked with a recent trend toward more quantitative methods in plant cell biology and imaging (Matas, et al, 2004; Tataw et al. , 2010; Farhidzadeh, et al. 2012; Mkrtchyan et al. 2013; Liu et al. 2013; Ung et al. 2013) . New tools are needed to easily and efficiently process and analyze large quantities of data produced by high throughput experiments. Computational tools are most commonly user-friendly and generally applicable to most platforms or they are highly technical and specific. EndoQuant attempts to bridge the gap, exhaustively addressing the pressing problem of automatic cellular phenotyping in plant cell biology in a user-friendly manner.

In an effort to dissect the endomembrane system in a systems-wide approach, Drakakaki et al. screened multiple small molecule libraries resulting in the investigation of 46,418 compounds (Drakakaki, et al. 2011). They used an automated primary screen in free-living tobacco pollen to identify compounds disrupting endomembrane processes required for pollen tube growth. It is reasonable to hypothesize that due to evolutionary conservation, molecules that produce phenotypes in pollen could potentially elicit a similar effect in *Arabidopsis* roots. To this end, the authors used several polar membrane markers to dissect recycling at the plasma membrane (PM) in the roots of *Arabidopsis* seedlings. Three

hundred and sixty compounds caused different morphological phenotypes in tobacco pollen presumably as a result of altered endomembrane dynamics as pollen tube growth relies heavily on the endocytosis and exocytosis machinery. Of these three hundred and sixty small molecules, 123 showed altered localization of endomembrane markers in *Arabidopsis* roots, which provide a readily available model for easily visualizing endomembrane trafficking in *Arabidopsis* through the use of fluorescent markers. In total, 15 subcellular phenotypes were identified, the most prevalent being the formation of intracellular bodies of various sizes, subdivided qualitatively into small, medium and large aggregates (Drakakaki, et al. 2011). The placement of each phenotype into a small, medium or large sub-class is descriptive and lacks objectivity. Additionally, in an effort to describe the cellular phenotypes in a biologically relevant manner, the effect of each small molecule was characterized on a per-cell basis.

The challenge of automatically quantifying subcellular structures in eukaryotic cells is hardly new. Most of the computer vision innovations enabling biologists to better study cellular phenomena have been in an effort to develop tools for animal cells, not taking into account unique challenges that exist for plant systems. For example, many segmentation methods involving animal cells rely on an initial estimate of the cell location. Using DAPI stain to mark the nucleus of the cell provides this estimate (Quelhas et al. 2010). These markings are easily detected based on their color. The rest of the cell is detected by employing a region-

growing-type algorithm to detect cell edges. This is not possible in plant cells where the large vacuoles result in tight appression of the nuclei to the PMs. Additionally, plant cells come in many different shapes including rectangular root cells, tube shaped pollen tubes and root hairs and puzzle piece shaped pavement cells; therefore shape based detection would not be generally applicable.

Many freely available open image analysis software programs and commercially available products have been used to automatically detect cell boundaries and endosomal bodies (Bashline, et al. 2013), but they lack the balance of flexibility and specificity needed to address specific challenges in plants including differing cell shapes, multilayers of cells, dispersed Golgi bodies, a large central vacuole, and other challenging architecture while requiring a large amount of user input (Steffens et al. 2014). Existing platforms analyze either a manually designated region-of-interest or the entire image leaving the biologist without a biologically meaningful measurement of specific cellular phenotypes.

Modification of an existing cell detection program designed for animal cells was used to detect pavement cell boundaries in *Arabidopsis*, as well as endosomal bodies, in an effort to automate confocal microscopy screening (Salomon et al. 2010). This approach consisted of two separate scripts, one that detected cell boundaries and one that detected endosomal bodies. However, it described a limited number of features on a per cell basis.

Here, we report a novel user-friendly image analysis tool for automatic quantification of discrete endomembrane phenotypes from multiple images. It measures features of structures within each cell resulting in the biologically relevant computation of statistics on a per cell basis. A large volume of datasets can be processed or any two images can be compared quantitatively. The data can be visualized immediately in graphs allowing the extracted features to be easily visualized for each cell or as feature histograms for population analysis.

3.3 Results

Detection of cell boundaries

Because of the interest in recycling at the PM, Drakakaki et al imaged roots expressing a fusion protein consisting of PIN-FORMED2 (PIN2) fused to green fluorescent protein (GFP) which allowed for the visualization of the PM and endosomal recycling bodies (Drakakaki et al. 2011). In order to provide a biologically useful measure of body number per cell, we first needed to automatically detect the cell regions. Because the PM is clearly labeled in contrast to the background, we needed to detect the boundaries. After pre-processing we used a Laplacian of Gaussian filter on the entire image, highlighting regions of rapid intensity changes such as edges (Chen, et al.1987). The user can define and adjust two parameters associated with this detection method including the window

size and the sigma. The sigma alters the distribution of the weighted pixel values and the window size adjusts the size of the filter that is the number of pixels within the filter. The filtered image was then made binary through the use of Otsu's method creating a black and white image (Otsu, 1975). Holes in the cellular regions were filled-in by using binary morphological operations completing a continuous cellular region (Kong et al. 1996). These regions were used as masks to represent each cell. The contour of each mask was extracted enabling the manipulation of the cell border. The border was then smoothed to remove noisy edges and provide a smooth continuous shape (Figure 3.1A and 3.1B). This parameter can be adjusted by the user in the GUI.

It is important to note that when visualizing a cross-section of a cell in a confocal image all four sides of a cell do not need to be clearly visible to provide clear cell regions. Often, one or more sides are not visible due to being out of the plane of focus or polar PM localization of the fluorescent marker. The cell detection algorithm could reasonably estimate the cell region with only three walls clearly visible. When inferring the boundary of the missing wall, the resulting boundary was unpredictable and not simply a fourth wall, often intruding into the cell or bulging outside of the real cell boundary. Our GUI allowed filter parameters to be altered in order to best capture the border of each cell. Additionally, the borders of the detected cells were overlaid on the subject image with the number of each cell, giving an immediate count of the number of cells in an image. This

allowed the user to determine the cells that were detected correctly and subsequently omit the cells that were improperly detected or of unimportance.

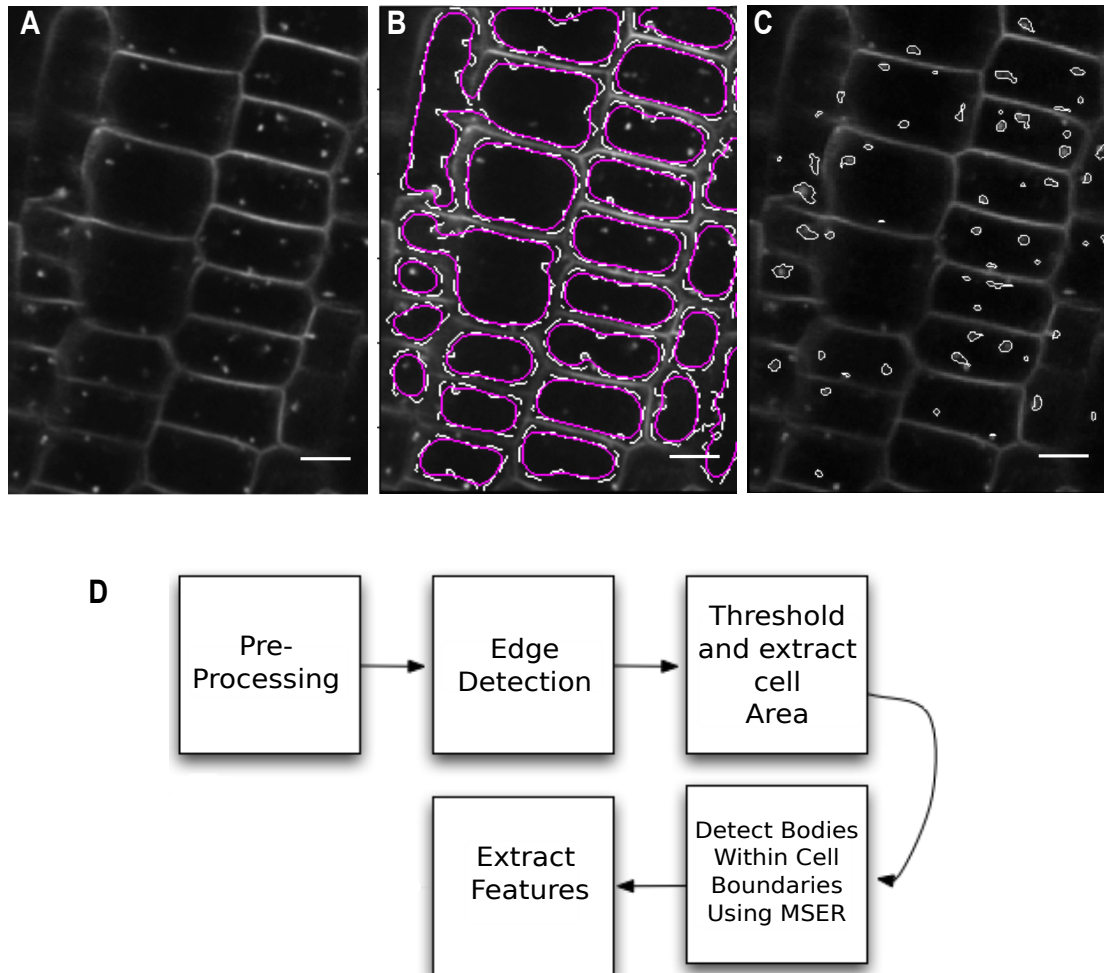


Figure 3.1. EndoQuant detects cellular structures within cell boundaries of PIN2:GFP transformed plants. (A) Original image showing Arabidopsis root cell in the meristematic zone. (B) Cell boundaries detected by EndoQuant. White dotted lines show initially detected regions; solid magenta lines denote smoothed shape. (C) Cellular structures within each cell are detected and outlined in white. (D) Flowchart describing detection and analysis steps in the fourth panel. MSER refers to Maximally Stable Extremal Regions. Scale bars in A-C = 20 μ m.

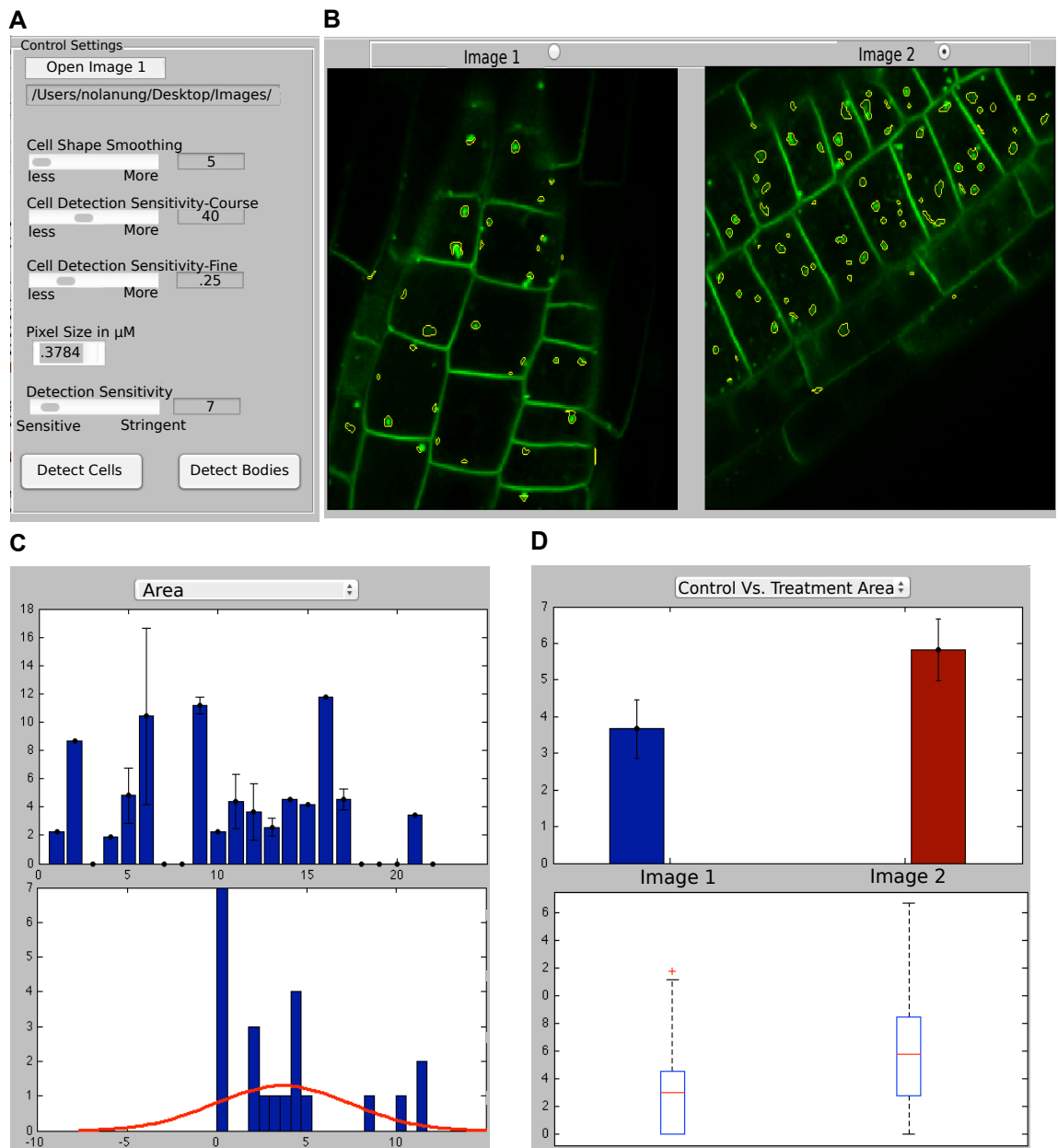


Figure 3.2. EndoQuant graphical user interface. (A) Settings panel. (B) Dual display. (C) Histogram displaying average parameter value per cell (upper) and histogram of values (lower). (D) Histogram (upper) comparing two images (panel B) and box plot showing basic statistics (lower).

Detecting cellular structures within individual cells

Each cell mask detected by the previously discussed method was then used to find cellular structures within each region. Within the cellular region, bright cellular structures were detected using Maximally Stable Extremal Regions (MSER). For the given region, MSER varied an intensity threshold using a user-defined step size, noting the area and location of the segmented objects. A region was considered stable if a change in intensity threshold did not lead to a change in region area. This process continued until all stable regions were determined. Note that MSER was performed for each segmented region obtained from the previous step in the algorithm. As such, the search for stable regions within each cell was independent of other cells (Figure 3.1C). Thirteen features were extracted and presented to the user as average values per cell in a graphical form and as raw numerical data (Supplemental Table 3.1). Each feature was calculated with respect to the number of whole cells present within the two-dimensional image. This provided an intuitive and biologically meaningful metric with which to measure cellular phenotypes. Figure 3D provides a summary of the steps taken to detect the cellular boundaries and structures.

The GUI has multiple useful features.

The GUI was designed with plant biologists in mind. It was crafted to be intuitive and user friendly. One simply needs to select and upload data, adjust three parameters to detect cell borders, one parameter for detection of structures within the cell, and then select two buttons, one for each detection step. Adjustment of three parameters provided flexibility to accommodate image variation (Figure 2A).

The graphical displays offered a number of different visualization options, effectively eliminating the time consuming step of processing and displaying raw data. A more detailed graph can be selected that shows the average value of each cell (Figure 3.2C, upper histogram). Simultaneously, a histogram of values with a best fit, was displayed showing the user a distribution of values (Figure 3.2C, lower histogram). This feature was particularly useful if there were multiple structures labeled or there were two populations of cellular structures; the user was able to track the dynamics of this complex population of cellular entities.

The flexibility of this tool also gives biologists the ability to analyze two images simultaneously and view each image independently or directly compare images, most likely a control and a treatment image (Figure 3.2B). The results of these images were immediately viewed and directly compared with each other by selecting the corresponding graph option, complete with standard error bars (Figure 2D, upper histogram). When this option was chosen, a box and whisker plot was

also displayed directly below the conventional bar graph (Figure 3.2D lower histogram). The box and whisker plot allowed the biologist to view fundamental statistics concerning the two sets of data such as the quartiles, medians and outliers of the dataset.

Once the cells and bodies have been detected and the extracted data has been visualized, the user can then export both the raw numerical data and the graphical data. The raw data is exported in a format compatible with common spreadsheet software for further statistical analysis, and the graphs are exported as image files for inclusion in presentation or publications.

EndoQuant has cross-tissue utility and high accuracy

EndoQuant was amenable to multiple cell types as it only required the defined boundaries and did not take shape information into account. Therefore, it was applicable to images of various cell types including root cells, pavement cells, and pollen tubes (Figure 3.3). It is reasonable to assume that protoplasts and root hairs would also be amenable to this technique. A common lipid or cellulose dye can be used to indicate boundaries in samples that lack boundary definition by labeling the PM or cell wall respectively. The only requirement was that a continuous border must surround the object.

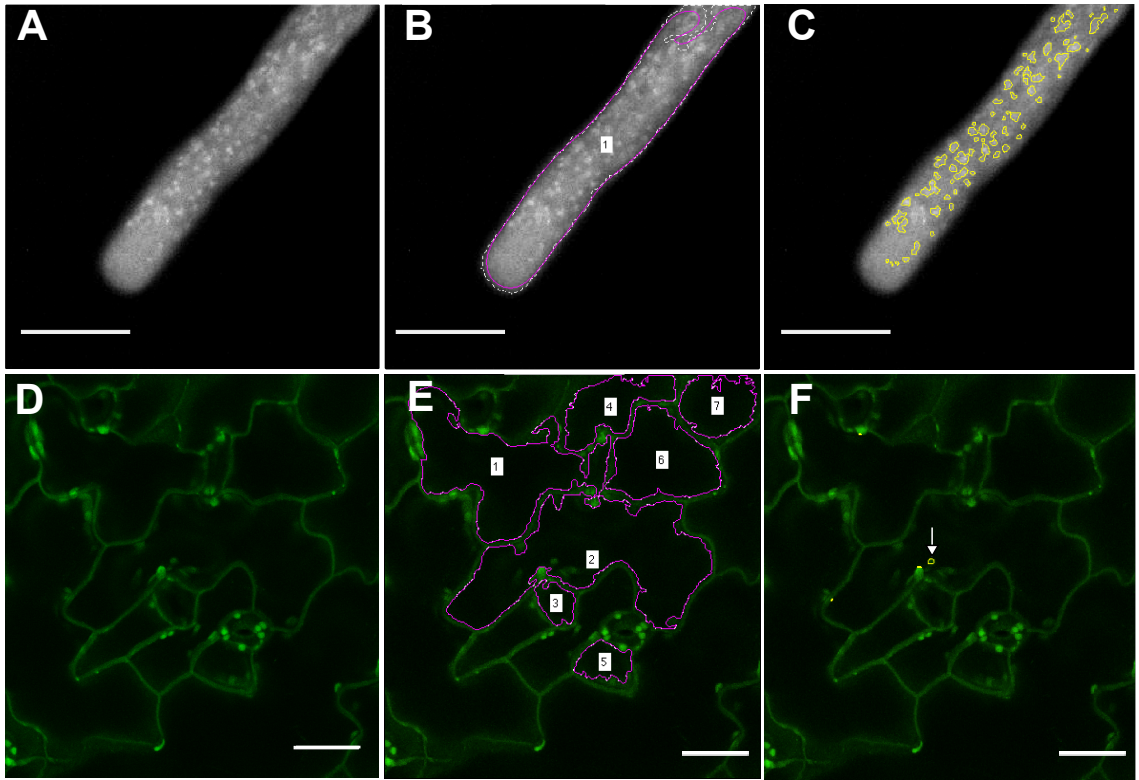


Figure 3.3. EndoQuant is applicable to multiple tissue types. (A-C), Arabidopsis pollen tube expressing ROP1 activity Marker Rab2:GFP. (D-F), Arabidopsis pavement cells expressing the endosomal and PM marker BRI1:GFP (Geldner et al. 2007). (A) and (D), original confocal images. (B) and (E), Detection of cell borders; cell boundaries indicated by magenta outlines. (C) and (F), detection of stable regions indicated by yellow outlines. White arrow indicates outlined region. Scale Bars A-C = 10 μ m, D-F = 25 μ m.

To evaluate the utility of MSER when detecting cellular structures we compared the performance of the combined methods designed to detect the cell boundaries and the structures within those boundaries with the manually detected ground-truth. The resulting receiver operating characteristic (ROC) curve which

was used to show the performance of a binary classifier not only demonstrated the efficacy of our tool but also indicated that on average, the sensitivity parameter was most effective at a delta value of four (Figure 3.4A.). Delta is the user defined step size the MSER algorithm uses when varying the intensity threshold to determine stability. Change in This delta value means that an intensity threshold is varied by a value of 4 and the region does not change; thus it is considered stable. We took advantage of this fact by making this the default value for delta; however, the user has the ability to adjust this parameter to optimize detection for any given dataset. In its current state EndoQuant can detect bodies with an accuracy of up to 85% (n=32 Image fields) while maintaining a low false positive rate. Twenty features are then extracted from these detected regions including the number of structures, area, basic shape information, mean intensity and other relevant features for each cell.

Reduced analysis time is a major motivation for automated image analysis programs, given that manual analysis is highly time consuming. When compared to manual detection with an open source image analysis package, EndoQuant drastically reduced analysis time by more than 100-fold (Figure 3.4B).

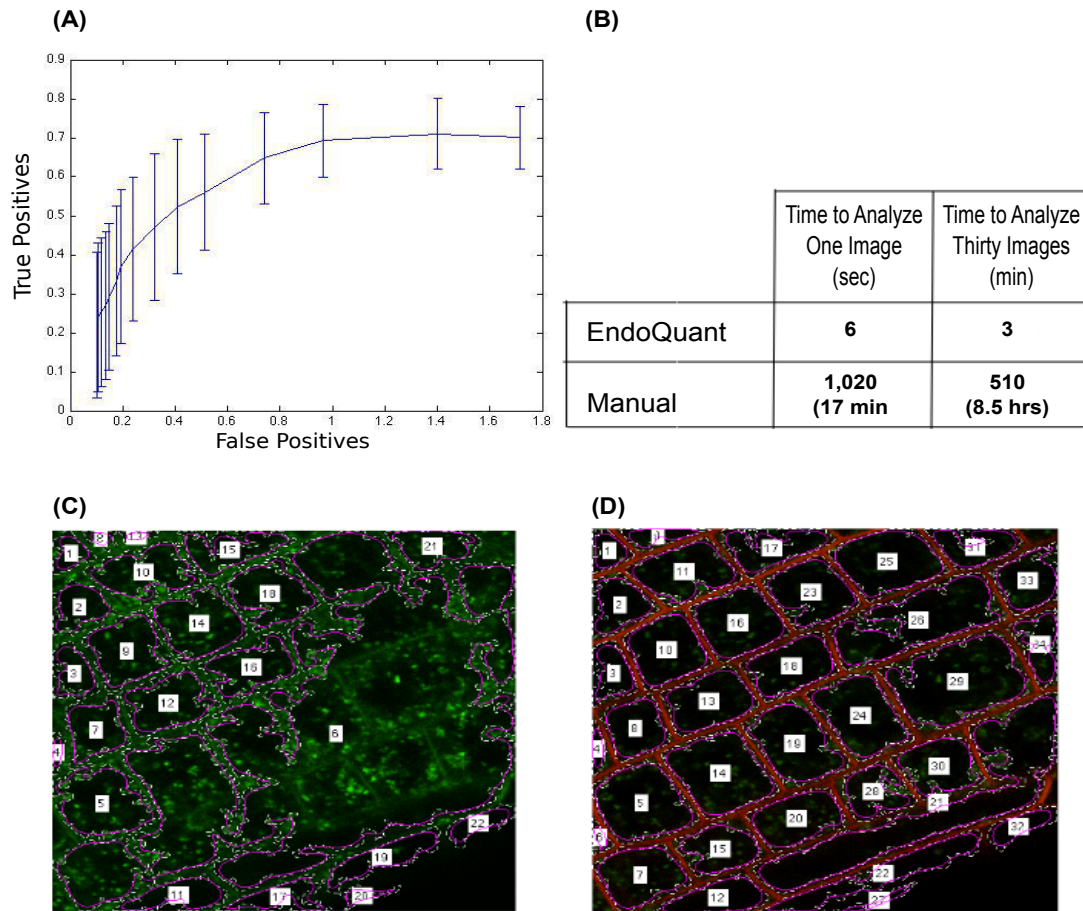


Figure 3.4. Analysis of EndoQuant Performance. (A) ROC curve showing the effect of delta on MSER sensitivity. (B) Table showing analysis time of EndoQuant versus manual detection using ImageJ. (C) The result of EndoQuant’s cell detection algorithm on root tissue labeled with the endosome marker ARA7:GFP (Geldner et al. 2009.) (D) The result of EndoQuant’s cell detection algorithm on root tissue labeled with ARA7:GFP and overlaid with FM4-64 dye labeling of the PM.

Limitations

For EndoQuant to be effective images must have high contrast with clearly defined borders like those of a fluorescently labeled PM or cell wall images acquired using

confocal microscopy. The cell detection algorithm will fail if these borders are inconsistent or of low intensity. Cellular structures within the cell and at the PM have similar intensities; therefore bodies that are very near the cell boundary were occasionally included in the boundary and excluded (see for example cell 28, Figure 3.4D; cell 5 Figure 3.3E). This can be minimized by increasing the smoothing of the cell boundaries; however over-smoothing may result in a loss of the cell corners. A great advantage of our cell detection technique was its ability to estimate cell boundaries when given three walls of a cell. However, when these borders were established, they could be unpredictable without the fourth side, either over or underestimating the true cell boundaries. A manual selection feature helped mitigate this problem by allowing users to select a subset of cells to analyze, thereby reducing the amount of error. These instances of mis-detection of cell boundaries can be compensated for when the image is acquired. If fluorescence signal is weak or absent in the PM, a common cell wall or styryl dye such as FM4-64 that labels these boundaries can be used to enhance the visibility of these boundaries. (Figures 3.4C and 3.4D).

3.4 Discussion

The growing high-throughput nature of biology demands innovative tools to manage and automatically quantify cellular phenotypes in a manner that is rapid,

accurate and user friendly. Manual region drawing is time consuming and inconsistent due to intra- and inter-user variation. Endoquant will not only allow for rapid processing of data, but it uniquely allows for rapid decision making through instant visualization comparison of data and basic statistical calculations. This tool allows high-throughput screens on the single cell level to be practical thereby allowing scientists to access previously unobserved phenotypes. Designed by biologists, for biologists, EndoQuant addresses issues that have not been accounted for by other image analysis suits such as detecting bodies or cellular structures on a per cell basis and providing rapid visualization and subsequent analysis of results.

First, it provides a biologically relevant metric by delivering results as an average per cell. This is achieved by first detecting the cell boundaries indicated by a membrane dye, cell wall dye, or PM localized fluorescent protein. The second step searches within this cellular region to find areas of maximum stability; that is, regions whose area fluctuates the least when the intensity level is increased by δ . The full potential of this method is realized when maximum intensity projections (MIPs) are used as inputs. A MIP is created by projecting the voxels with the highest intensity from every angle throughout the volume onto a two-dimensional image plane. In this case, MIPs provide more spatial data and condense a three-dimensional environment into a more manageable two-dimensional sample.

Like any detection method, increasing the sensitivity of detection will eventually lead to false positives; at high sensitivity the algorithm will detect the PM boundaries and skew the data. However, decreasing the sensitivity severely will result in false negatives, omitting important data. The adjustable parameters and the manual pruning give the biologists control over the ultimate outcome of the data. Parameters should be set to minimize both false positives and false negatives. These should remain consistent when analyzing multiple images to ensure accurate comparative analysis.

Second, when analyzing a single image, the features are displayed immediately as histograms for each cell and histograms for population analysis. This now provides biologists the ability to quantitatively distinguish two populations of compartments based on size, intensity and other parameters. For example, a mutant may cause the mis-localization of a fluorescently tagged protein to two different compartments. Endoquant would allow biologists to quantitatively and rapidly visualize the two distinct compartments. When comparing two images, two features can be directly compared in a single graph including standard error bars and as a box plot showing basic statistics, including statistical outliers. Subsequently, both graphical and raw data can be exported for further analysis and manipulation.

Although there are limitations with using EndoQuant such as having clearly defined borders and regions of interest within the intracellular area that must be separated from the boundary limit, it still represents a major improvement compared to existing methods that do not show the accuracy and biological relevance demonstrated by EndoQuant. In its current state, EndoQuant is available as an open access MATLAB script. However, future versions will be available as an ImageJ plugin. It is common for engineering departments at most universities to use MATLAB and would most likely be available to biologists. Technology such as EndoQuant will allow for high-throughput assays at the cellular level to become practical, permitting the investigation of phenotypes and biological phenomena that do not have an obvious macroscopic phenotype. It will also enable biologists to collect and interpret large amounts of phenotypic data that can be later used to influence experimental designs or be incorporated into models. EndoQuant can be used to analyze organelle dynamics by quantifying changes in endomembrane populations over time. Automated tools like EndoQuant will become increasingly imperative for dissecting cellular phenomena as quantification becomes increasingly essential.

3.5 Materials and Methods

Growth conditions and chemical treatments

Arabidopsis thaliana ecotype Colombia (col-0) seedlings were sterilized then sown on 0.5X MS containing media with 0.8% phytoagar. Rab2:GFP (Cheung et al. 2002), BRI1:GFP and VHAA1:GFP-expressing seedlings were stratified for 48hrs at 4°C then grown on an 18hr light cycle at 22°C. Original chemical screen and methods for imaging PIN2:GFP individuals can be found in Drakakaki et al. (2011). VHAA1:GFP seedlings were treated with 1µM FM4-64 in 0.5X liquid MS media for 10 minutes before imaging. Pollen was collected from tobacco plants and incubated in a pollen germination media for 2 hours. Pollen grains were then transferred on to a microscope slide for imaging (Robert et al, 2008).

Microscopy

The pollen tubes, pavement cells and root cells containing VHAA1:GFP were imaged using a Leica SP5 Confocal Laser Scanning Microscope (CLSM). Excitation wavelengths used were 488nm (GFP) and 543(FM4-64). Rab2: GFP pollen tubes were imaged using the BD Pathway automated microscope. Manufacturer settings were used for detection of fluorescence. All confocal microscopy was done at the Microscopy and Imaging Core at the Center for Plant Cell Biology (CEPCEB) of the Institute for Integrative Genome Biology (IIGB) at University of California at Riverside.

Image analysis

EndoQuant was developed using Matlab and the image analysis toolbox. The subsequent code is included in the Supplementary Material, including the code for the GUI. All code is open source and available upon request. Manual detection for the ROC plot was done using Matlab. The user indicated the centroid of the body by clicking in the center. The centroids of the manual vs. the automatic methods were compared and if they were within a three pixel neighborhood of the manually detected point they were considered accurate. Manual analysis comparing the time of ROI selection was done using FIJI.

3.6 References

Bashline, L., Li, S., Anderson, C.T., Lei, L. and Gu, Y. (2013). The endocytosis of cellulose synthase in *Arabidopsis* is dependent on $\mu 2$, a clathrin-mediated endocytosis adaptin. *Plant Physiol.* 163:150-60.

Chen, J.-S., Huertas, A. and Medioni, G. (1987). Fast convolution with Laplacian-of-Gaussian masks. *Pattern Analysis and Machine Intelligence, IEEE Transactions on.* 584-590.

Cheung, A.Y., Chen, C.Y., Glaven, R.H., de Graaf, B.H.J., Vidali, L., Hepler, P.K., and Wu, H. (2002). Rab2 GTPase regulates vesicle trafficking between the endoplasmic reticulum and the Golgi bodies and is important to pollen tube growth. *The Plant Cell Online* 14: 945.

Drakakaki, G., Robert, S., Szatmari, A.M., Brown, M.Q., Nagawa, S., Van Damme, D., Leonard, M., Yang, Z., Girke, T., Schmid, S.L., Russinova, E., Friml, J., Raikhel, N.V. and Hicks, G.R. (2011). Clusters of bioactive compounds target dynamic endomembrane networks in vivo. *Proc Natl Acad Sci U S A.* 108:17850-5.

Farhidzadeh, H., Hashemi, S.N. and Masoudnia, S. (2012). Phenotypic analysis of *Arabidopsis Thaliana* root plant with improved feature extraction and combining

classifiers approach. In *Artificial Intelligence and Signal Processing (AISP), 2012 16th CSI International Symposium on*. p 452-457.

Geldner N., Hyman D.L., Wang X., Schumacher K., Chory J. (2007). Endosomal signaling of plant steroid receptor kinase BRI1. *Genes Dev* 21:1598 -1602.

Geldner, N., Déneraud-Tendon, V., Hyman, D. L., Mayer, U., Stierhof, Y.-D. and Chory, J. (2009). Rapid, combinatorial analysis of membrane compartments in intact plants with a multicolor marker set. *Plant J* 59(1):169-178.

Liu, K., Schmidt, T., Blein, T., Durr, J., Palme, K. and Ronneberger, O. (2013). Joint 3d cell segmentation and classification in the arabidopsis root using energy minimization and shape priors. In *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*. p 422-425.

Matas, J., Chum, O., Urban, M. and Pajdla, T. (2004). Robust wide-baseline stereo from maximally stable extremal regions. *Image and vision computing*. 22:761-767.

Mayer, .U. (1999). Small Molecule Inhibitor of Mitotic Spindle Bipolarity Identified in a Phenotype-Based Screen. *Science*. 286:971-974.

Mkrtchyan, K., Chakraborty, A. and Roy-Chowdhury, A.K. (2013). Automated registration of live imaging stacks of Arabidopsis. In *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*. p 672-675.

Otsu, N. (1975). A threshold selection method from gray-level histograms. *Automatica*. 11:23-27.

Page, D.R. and Grossniklaus, U. (2002). The art and design of genetic screens: *Arabidopsis thaliana*. *Nat Rev Genet*. 3:124-36.

Quelhas, P., Marcuzzo, M., Mendonça, A.M. and Campilho, A. (2010). Cell nuclei and cytoplasm joint segmentation using the sliding band filter. *IEEE Trans Med Imaging*. 29:1463-73.

Salomon, S., Grunewald, D., Stüber, K., Schaaf, S., MacLean, D., Schulze-Lefert, P. and Robatzek, S. (2010). High-Throughput Confocal Imaging of Intact Live Tissue Enables Quantification of Membrane Trafficking in Arabidopsis. *Plant physiology*. 154:1096.

Schindelin, J., Arganda-Carreras, I. & Frise, E. et al. (2012). [Fiji: an open-source platform for biological-image analysis](#). *Nature methods* 9(7): 676-682.

Steffens, A., Jaegle, B., Tresch, A., Hülskamp, M. and Jakoby, M. (2014). Processing-body movement in Arabidopsis depends on an interaction between myosins and DECAPPING PROTEIN1. *Plant Physiol.* 164:1879-92.

Stockwell, B.R. (2000). Chemical genetics: ligand-based discovery of gene function. *Nature Reviews Genetics.* 1:116-125.

Tambo, Asongu L.; Bhanu, Bir; Luo, Nan; Harlow, Geoffrey; Yang, Zhenbiao, (2014) Integrated Model for Understanding Pollen Tube Growth in Video. *Pattern Recognition (ICPR), 2014 22nd International Conference on*, vol., no., pp.2727,2732

Tataw, O.M., Liu, M., Roy-Chowdhury, A., Yadav, R.K. and Reddy, G.V. (2010). Pattern analysis of stem cell growth dynamics in the shoot apex of arabidopsis. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*. p 3617-3620.

Ung, N. Brown, M.Q., Hicks, G.R., Raikhel, N.V. , (2013) An Approach to Quantify Endomembrane Dynamics in Pollen Utilizing Bioactive Compounds. *Molecular Plant* 6:1202-1213.

Chapter 4

Prediction of Small Molecules Affecting Recycling Using Phenomics and Machine Learning

4.1 Abstract

Advances in automated microscopy and image analysis increasingly require computational methods to manage, process and interpret the large amounts of phenotypic data that are collected from high-throughput assays. We analyzed the quantitative phenotypic data extracted from a large screen of small molecules in *Arabidopsis* roots designed to identify compounds that disrupt endomembrane recycling. Using a Gaussian Mixture Model (GMM), a machine learning technique, we found that the common “body” phenotypes that manifested as a result of the small molecule treatment fit into three clusters based on the size and number of the bodies per cell. Cluster one phenotypes were typified by few small bodies, cluster two by many large bodies and cluster three by few large bodies. A well-characterized molecule that disrupts membrane recycling, Endosidin2 (ES2) was classified into cluster two which suggested that other compounds in cluster two might have similar disruptive properties. A BFA washout was used to validate this prediction and found that indeed all tested compounds in cluster two inhibited trafficking. In contrast all compounds in cluster one did not alter membrane

recycling demonstrating the predictive power of this method. We are now able to predict one biological process affected by a small molecule based on the phenotype without *a priori* knowledge. This will guide future experiments and aid biologists in further characterizing small molecules of interest. Machine learning when applied to subcellular phenotyping can be a powerful strategy for predicting biological function and intelligently designing subsequent experiments.

4.2 Introduction

Accelerated phenotypic analysis and increased throughput are necessary to enhance our understanding of quantitative phenotypic differences and gain the ability to dissect intricate biological phenomena at the subcellular level. Indeed this objective is being addressed by an increased number of automated tools available for phenotypic analysis across model systems within recent years (Mathew et al., 2012; Vogt et al., 2009; Conrad et al., 2004). These tools require the automated processing and extraction of image data, retrieving relevant quantitative biological information using automated computer vision algorithms. The majority of these efforts in plant systems center on macroscopic growth phenotypes for the purpose of supporting breeding and growth studies (Klukas et al., 2014; Hartmann et al., 2011). However, the resulting data must be interpreted and supported by statistical analysis and extensive biological assays to understand the latent process

leading to the observed phenotype. The ability to connect previously uncharacterized phenotypes to biological phenomena will give scientists the ability to rapidly and efficiently identify the underlying biological basis for the corresponding phenotype and greatly aid in experimental design.

Phenotypes can often offer clues that hint at the corresponding biological function that has been modified. For example, a loss of apical dominance in *Arabidopsis* shoots indicates that auxin signaling or transport is disturbed, Brassinosteroid mutants exhibit drastic reduction in size and remain bushy rosettes, and cellulose synthesis defective mutants have drastically reduced seedling growth and swollen cells (Li and Bangerth, 1999; Li et al., 2001; Desprez et al., 2002). Disruptions in biological processes often produce consistent observable phenotypes. We can take advantage of this fact by utilizing machine learning algorithms to aid in the process of analyzing large amounts of phenotypic data produced by high throughput analysis workflows. Machine learning can help overcome this challenge by learning a defined set of rules from user provided examples, training the algorithm, then subsequently applying these rules to classify new data sets, predicting defined properties of these new samples (Sommer and Gerlich, 2013).

Classical genetic approaches consist of causing random mutations in the DNA and observing the resulting phenotype. Some of the challenges that may arise from

this approach and lead to the lack of a phenotype include redundancy, the ability for similar proteins to compensate for the loss of a single protein and lethality, the death of organism due to the loss of a protein that is essential for survival. Chemical genomics takes advantage of small bioactive molecule libraries to probe biological processes, circumventing the challenges of redundancy and lethality. Unlike a classical mutation approach in which the DNA is permanently modified, chemical genomics employs small molecules to bind to and inactivate proteins directly.

We use chemical genomics as an approach to study the endomembrane system, namely, a collection of highly coordinated membrane bound organelles that transport cellular contents from one subcellular location to another. Endomembrane trafficking, the process of transporting cellular cargo within the cell is a highly conserved biological process among eukaryotes (Elias, 2010). Exocytosis or secretion, the process of exporting cellular material to the exterior of the cell, and Endocytosis, the internalization of the outer membrane and external materials, are the prominent trafficking processes vital to a cell's survival. Once vesicle budding at the plasma membrane occurs, beginning the internalization of membrane materials, the resulting vesicle can have be directed to multiple pathways. The recycling pathway returns to the PM. The molecular mechanisms that cooperate to facilitate the recycling process are poorly understood though the process is conserved (Richter et al., 2009). Unlike well studied processes, disruptions in recycling do not have well characterized phenotypes. The ability to

immediately visualize disruptions in recycling would greatly aid in the identification of mutants or small molecules that indicate the discovery of a protein involved with endocytic recycling.

Similar to traditional methods, entire libraries of small molecules need to be screened to highlight those that produce the desired phenotype. This often entails screening and scoring thousands of compounds. However, when exploring poorly understood phenomena, a obvious phenotype is not assigned to a underlying biological process. For example, loss of apical dominance is a phenotype associated with auxin mutants, however this could be due to a decrease in auxin synthesis, transport, or perception. Similarly, many small molecules have been discovered to cause a general endomembrane phenotype of “endomembrane bodies”, anomalous endomembrane derived structures that are a result of disrupting normal endomembrane trafficking processes (Drakakaki et al., 2011). However, the origin and nature of these bodies are challenging to determine solely based on the image data.

Here we describe an approach to predict the biological function of a subset of small molecules using machine learning to classify the subcellular phenotypes produced by each compound. A Gaussian Mixture Model was used to classify the phenotypes into three clusters based on the size of each body and the number of bodies per cell. Known compounds to cause disruptions in endomembrane

recycling were compared against the three clusters. Cluster 2 contained compounds with similar properties to a known endomembrane recycling inhibitor, ES2. When this cluster was validated, all compounds tested inhibited recycling as suggested by a BFA washout assay. This method can be used to rapidly and accurately predict a subset of compounds that appear to be targeting a specific biological process reducing the number of experiments that need to be performed in order to discover compounds that inhibit endomembrane recycling.

4.3 Results

Two-dimensional phenotypic clustering reveals three phenotypic groups

Using EndoQuant, the software package we previously developed to study discrete endomembrane phenotypes, we analyzed the phenotypes of 36 compounds, about one third of the small molecules found to be bioactive in *Arabidopsis* roots (Drakakaki et al., 2011). This sample was chosen due to the high quality of available images. These visual snapshots showed *Arabidopsis* roots labeled with a PIN2:GFP construct: a common marker used to visualize recycling at the PM. These data were deemed suitable for analysis due to the number of cells present, the appearance of clear borders and strength of the flourophore signal.

Previously, compounds were grouped according to the phenotype, most of which exhibited various numbers and sizes of endomembrane agglomerations or

bodies. These Categories were arbitrarily constructed based on subjective and qualitative criteria and were classified into “small” “medium” and “large” categories. Our goal was to create phenotypic groups that were supported by a mathematical model and therefore quantitative in nature. A two-dimensional Gaussian Mixture Model (GMM) was used to cluster the small molecule phenotypes into three groups. A GMM is a probabilistic method that assumes that the data points within a given dataset fit within multiple Gaussian distributions samples (Sommer and Gerlich, 2013). The clustering criteria were based on the body size and the number of bodies per cell (Figure 4.1A). These are the most salient features of the body phenotypes in *Arabidopsis* roots and have been used by previous studies to classify endomembrane phenotypes (Drakakaki et al., 2011).

The clustering algorithm divided the 36 phenotypes into three clusters. Cluster one consisted of small molecules that caused few relatively small bodies per cell (Figure 4.2A). In contrast, cluster two contained phenotypes exhibiting a high amount of relatively large bodies per cell (Figure 4.2B). Lastly, cluster three contained small molecules that caused a small amount of relatively large bodies per cell (Figure 2C). In fact, the average body size of cluster three was greater than that of cluster 2. The GMM proved to be a useful approach in revealing the underlying phenotypic groups within which the predicted recycling-disrupting small molecules naturally occurred.

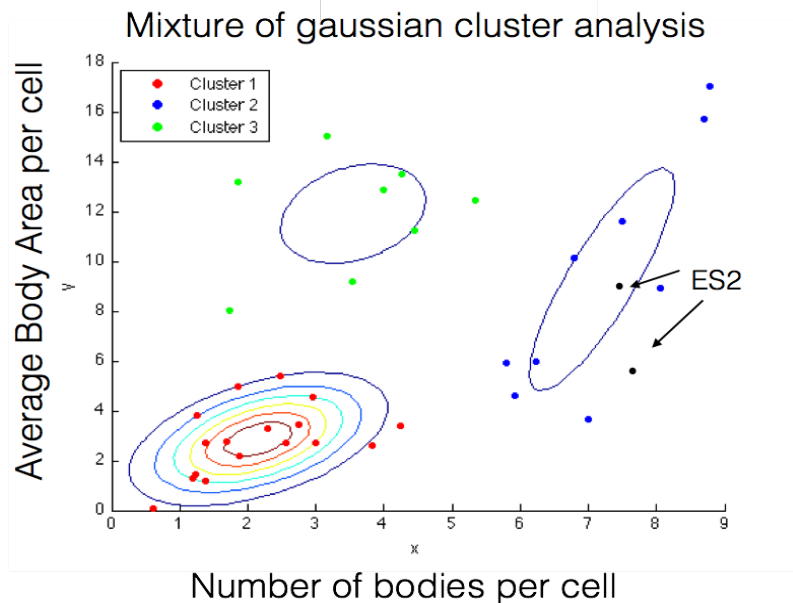


Figure 4.1. GMM cluster analysis of small molecule phenotypes. The X-axis shows the number of bodies per cell. The Y-axis shows the average body area per cell in micrometers. Cluster one is indicated in red, Cluster 2 in Blue and Cluster 3 in green. Black data points show two images of the ES2 phenotype clustered into cluster 2.

Predicting small molecules that target endocytic recycling.

By using this clustering algorithm to determine three groups, it has “learned” the criteria for each group. We hoped that by introducing new data sets, the algorithm will cluster the given phenotype into its corresponding group. When we analyzed the phenotype caused by ES2 a well-characterized inhibitor of recycling, the algorithm sorted it into cluster two. We hypothesized that it might be possible that these compounds caused a similar

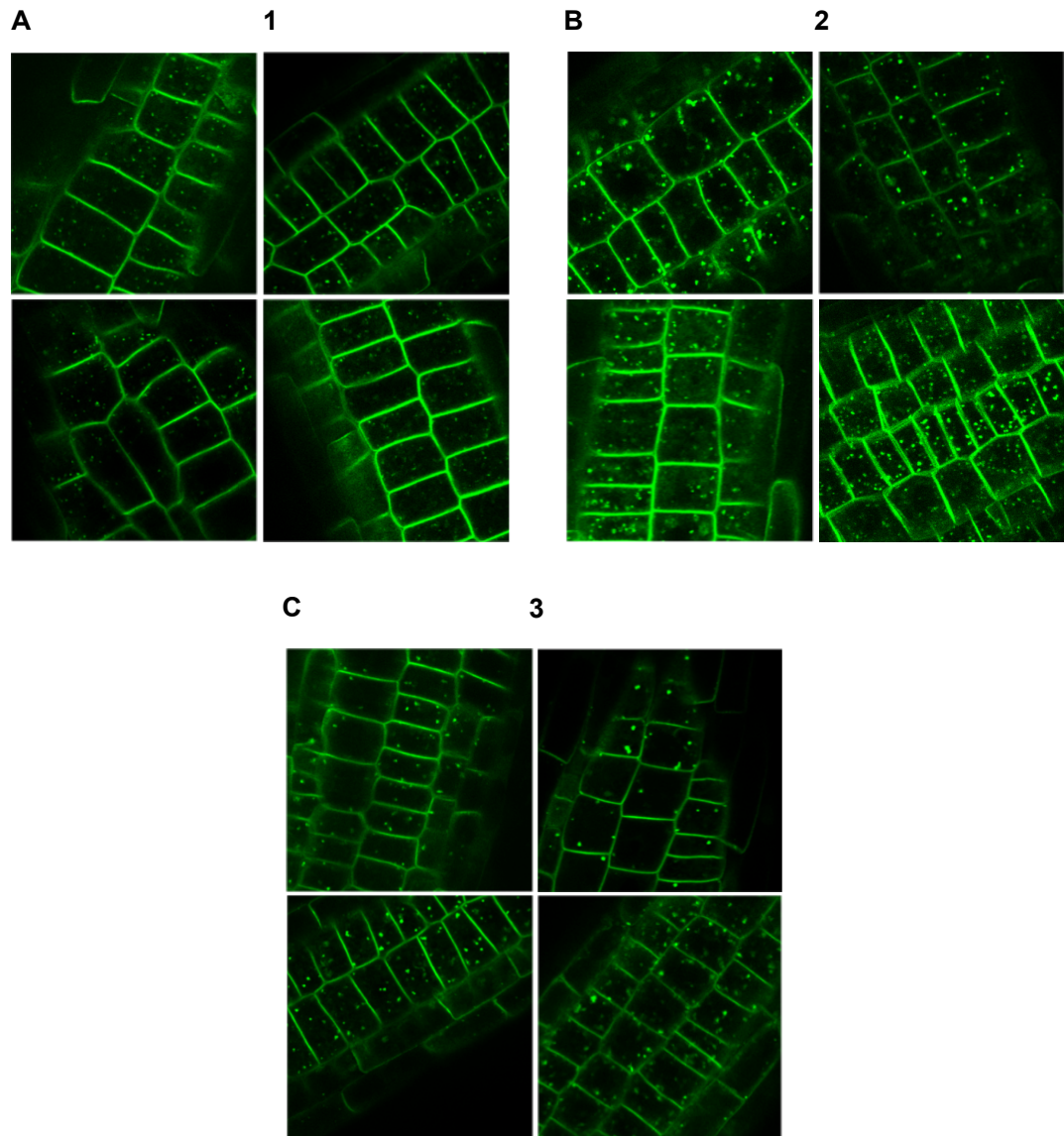


Figure 4.2. Representative images of each phenotypic cluster. (A) Four representative of phenotypes in cluster one exhibiting few small bodies. (B) Four representative images of cluster 2 exhibiting many larger bodies. (C) four representative images of cluster 3 exhibiting few large bodies. Above endomembrane phenotypes are observed using PIN2:GFP a marker for endocytic recycling.

phenotype to ES2, that is, those clustered with ES2 might also be affecting recycling.

To validate this model, we treated *Arabidopsis* roots with four compounds from the cluster two. These compounds were chosen based on their commercial availability. As with the previous images the small molecules were found to produce many relatively large bodies (figure 4.3). A Brefeldin A (BFA) washout assay was used to determine if these small molecules were indeed altering

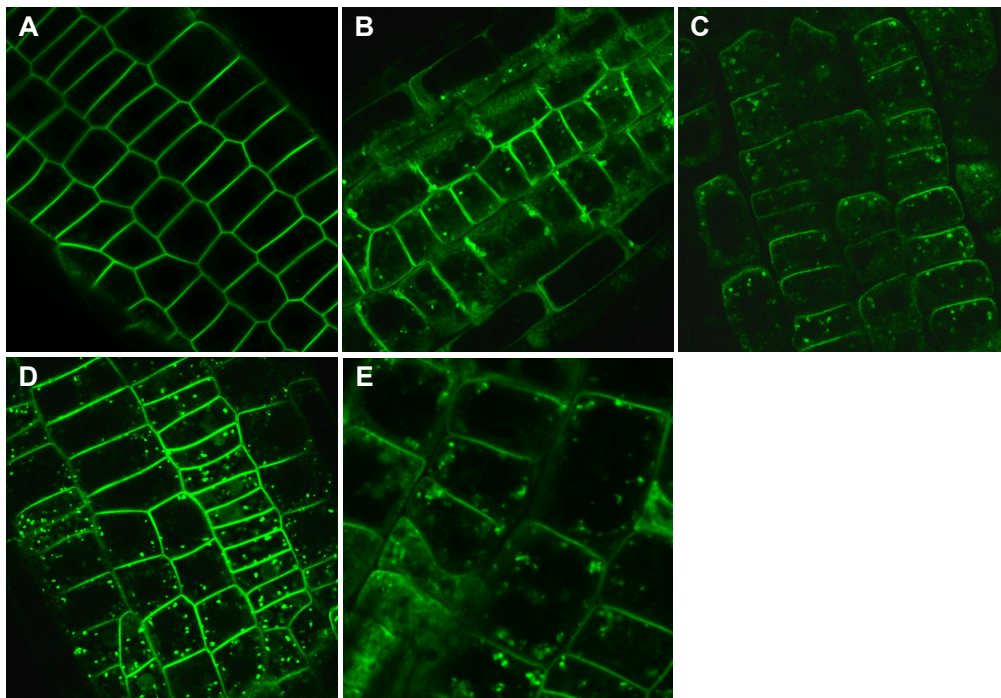


Figure 4.3. Phenotype confirmation of selected cluster 2 compounds. (A) DMSO treated control. (B) compound A7. (C) compound A10. (C) compound B5. (D) compound E4.

recycling dynamics. BFA disrupts the normal trafficking routes from the Endoplasmic reticulum to the Golgi apparatus (Miller et al., 1992). Once this compound is removed the secretory contents of the endomembrane system are allowed to progress beyond the Golgi. If a post-Golgi trafficking process such as recycling is disrupted, the BFA body will persist beyond that of the control. Therefore, recycling is most-likely inhibited if BFA bodies persist in the presence of the cluster two compounds (Hendricks et al., 1992). As predicted, all four of the compounds tested show persistent BFA bodies 80 min after BFA has been washed out in contrast to the DMSO control (Figure 4.4). Interestingly, during the washout phase of the BFA washout assay, each root treated with its respective bioactive compound contain only BFA bodies while others contain BFA bodies in addition to other small bodies (Figure 4.4C).

The fact that all tested compounds could disrupt recycling by itself does not show that our clustering method has any predictive capability. It is possible that all of the analyzed small molecules could have the ability to disrupt endomembrane recycling thus rendering our approach ineffective. To demonstrate that this approach can indeed be used to predict compounds that disrupt endocytic recycling we tested cluster one, the cluster that is the most divergent from the cluster two. If this method is truly predictive then we would expect that cluster one compounds would not disrupt the endocytic recycling process.

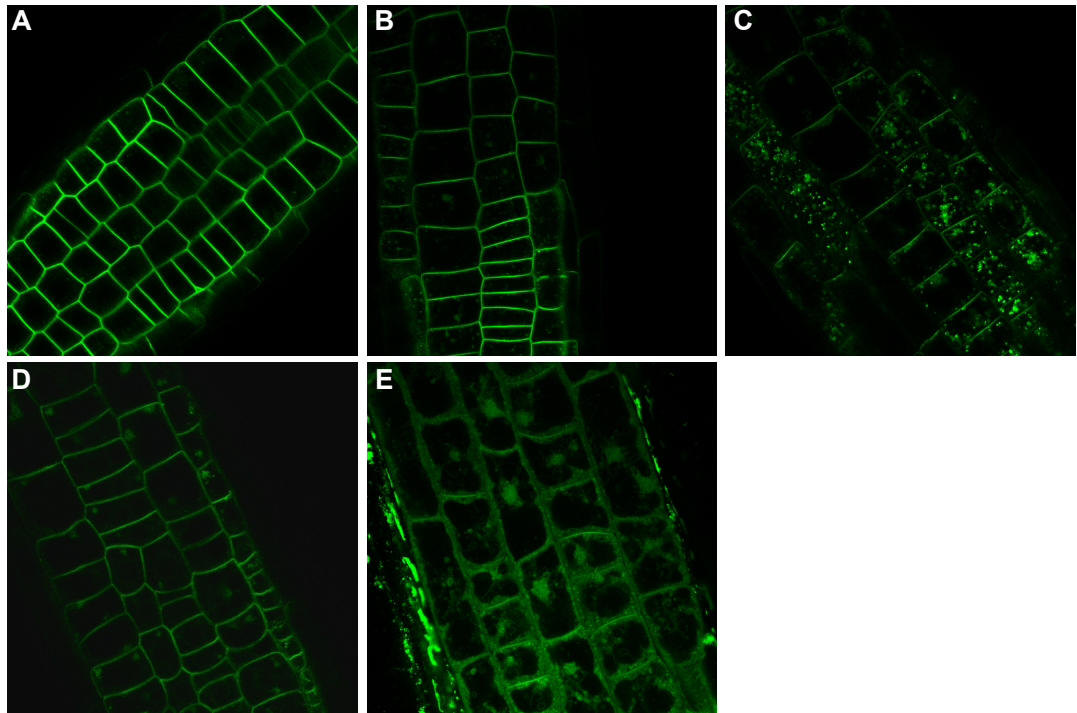


Figure 4.4. BFA washout using Cluster 2 compounds. (A) DMSO treated control. (B) compound A7. (C) compound A10. (C) compound B5. (D) compound E4.

To validate this prediction, we tested the molecules residing within cluster one on the PIN2:GFP marker to confirm their cluster one phenotype typified by the presence of few small bodies (Figure 4.5). Even at relatively high concentrations, these compounds produced few small bodies, confirming their placement in cluster one. To test their effect on recycling and validate our model, the same BFA washout assays was performed with the cluster 1 compounds. We would expect that these compounds do not alter recycling dynamics and therefore would not produce persistent BFA bodies after washout. Indeed the BFA bodies did not persist, indicating the normal function of the recycling machinery (Figure 4.6).

These assays suggest that the clustering method using a Gaussian Mixture Model has the ability to predict the biological function of a small molecule based on the produced phenotype.

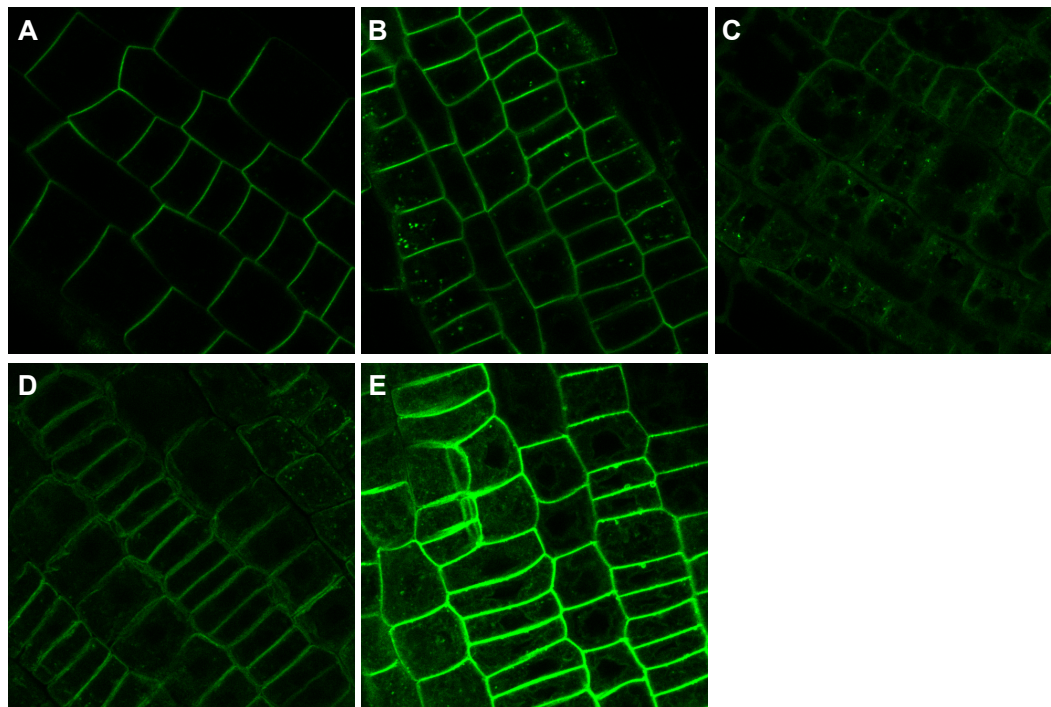


Figure 4.5. Phenotype confirmation of selected cluster 1 compounds. (A) DMSO treated control. (B) compound B5. (C) compound B7. (C) compound C8. (D) compound E11.

The phenotype shared between the members of each group would presumably be caused by the disruption of a similar biological process or even the same protein. It is possible that this could be the result of a similarity in structure. However, when we compared the structures of the compounds within cluster two,

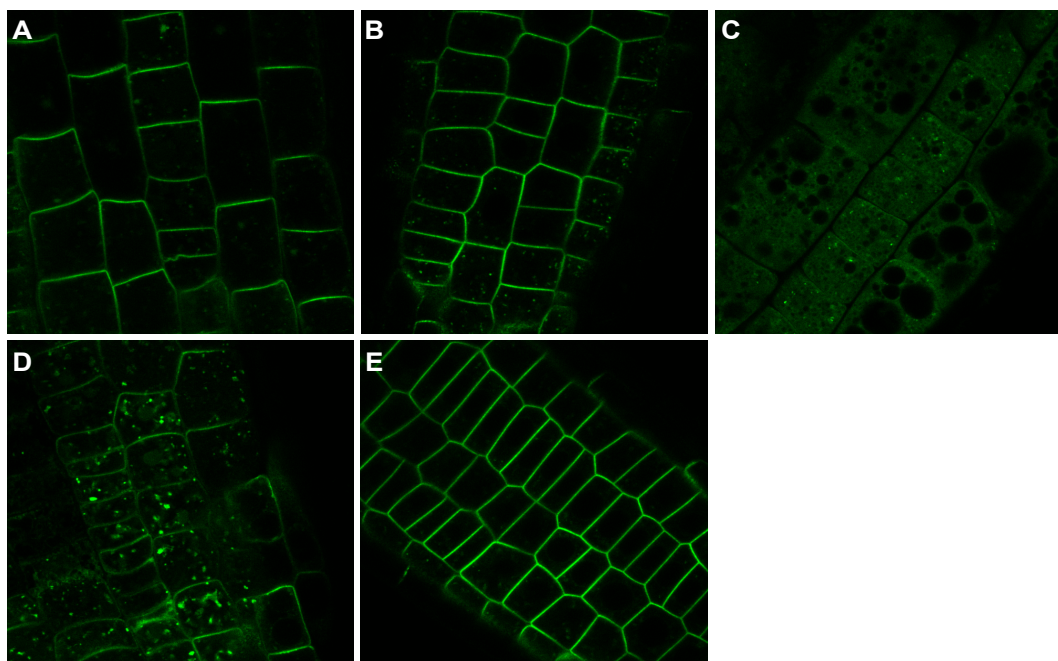


Figure 4.6. BFA washout using Cluster 1 compounds. (A) DMSO treated control. (B) compound B5. (C) compound B7. (C) compound C8. (D) compound E11.

we did not see a structural trend in that all four groups did not share similar functional groups (figure 4.7). Similarly, the small molecules within cluster one do not seem to share and similar functional groups that may suggest that they are acting in a similar way. These differences in structure may suggest that the drug-like molecules aren't targeting the same protein but similar proteins within the same or similar pathways ultimately having the same downstream effect. Similar to cluster two, the small molecules within cluster one do not share any structural commonalities that unify them (figure 4.8). Additionally, commonalities are not observed between either cluster as well, which would be expected. The differences

between cluster one and cluster two validate the predictive capabilities of our method.

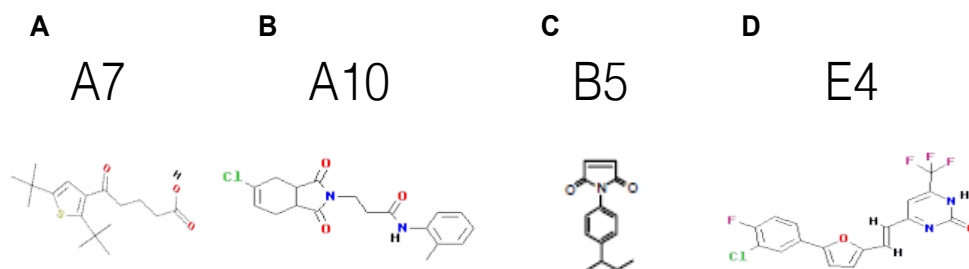


Figure 4.7. Molecular structure of cluster 2 compounds. (A) compound A7. (B) compound A10. (C) compound B5. (D) compound E4.

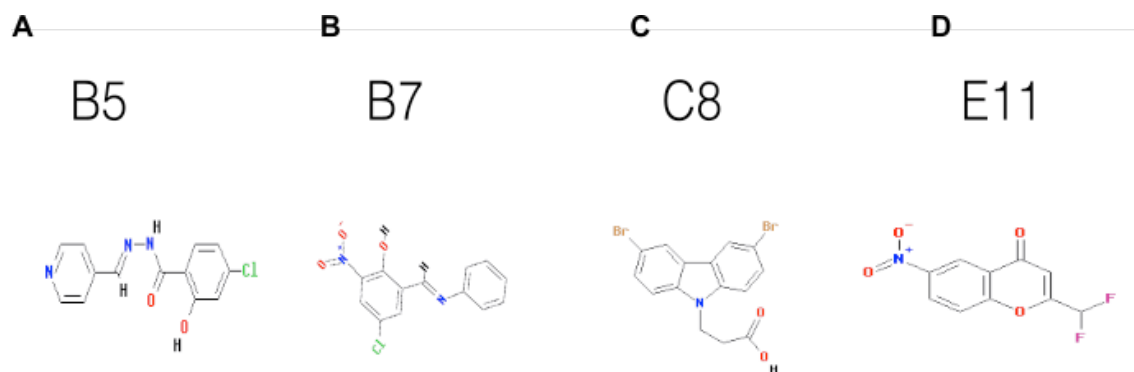


Figure 4.8. Molecular structure of cluster 1 compounds. (A) compound B5. (B) compound B7. (C) compound C8. (D) compound E11.

4.4 Discussion

Chemical genomics is a powerful approach for the dissection of highly conserved processes including endomembrane trafficking and has numerous advantages including tunable control, applicability to multiple systems and the ability to circumvent lethality and redundancy. However, the screens necessary for the identification of useful small molecules are high throughput in nature and require the subsequent identification of the target protein, which can be challenging and labor intensive. To compound these challenges, the resultant amount of data is so large that it is often impractical to manage and analyze it manually. We have described a method that gives us the ability to predict the molecules that are of interest, in this case, those that disrupt endomembrane trafficking using a GMM. This allowed us to classify and predict their function based on the resulting phenotype. As a result, a more intelligent experimental design could be created that would expedite the screening process and enable researchers to focus on a subset of predicted hit compounds.

The GMM showed that the data could be subdivided into three subgroups. Generally every compound elicited phenotype can be grouped into one of three clusters; cluster one has few small bodies, cluster 2 has many large bodies and cluster 3 has few large bodies. Upon first glance, these various cellular phenotypes

are indistinguishable from one another, however, the GMM allowed us to parse out these subsets and reveal differences that would have otherwise gone unobserved. Phenotypes within cluster one showed high association with one another. Cluster two and cluster three show more variation, however both clusters show clear separation from either cluster one or each other. Although this sample size seems to represent the population for phenotypes observed in the secondary *Arabidopsis* screen, increasing the sample size would increase our understanding of the phenotypes produced by these bioactive compounds. Assuming these groupings are genuine we would expect that increasing the sample size would reduce the variation within the clusters while maintaining the variation between them. Our model shows high confidence in the classification of three subgroups of body phenotypes, though increasing the sample size would bring us closer to revealing the true nature of these compounds and their phenotypes.

ES2 has been well characterized an inhibitor of recycling and produces very stable body phenotypes (Drakakaki et al., 2011). When multiple images of ES2 were analyzed and included in the GMM clustering model, it was classified into cluster two. We then hypothesized that it might be possible that the other compounds that produced similar phenotypes to ES2 may also be affecting the endomembrane recycling process. However, like any computational model, our GMM and clustering approach are a simplification of extremely complex circumstances and needs to be validated by testing the predicted results using a

reliable biological test. The BFA washout assays were used to confirm that selected compounds from cluster two did indeed inhibit the endomembrane recycling process and that compounds in cluster one did not, thereby confirming the predictive power of this approach. Due to the limited availability of some of these compounds only four from both cluster one and two were tested. As with any constrained model, increasing the sample size would increase the ability to better evaluate the accuracy of this method. All compounds tested behaved as predicted, compounds in cluster two showed lingering BFA bodies upon washout showing inhibition of the recycling pathway. In contrast, all of the compounds in cluster one did not show lingering BFA bodies and therefore demonstrates that they do not inhibit recycling dynamics. This does not however, rule out the possibility that they are effecting recycling in other ways such as increasing recycling or re-routing recycled materials. Interestingly, during the washout phase of the BFA washout assay, when each root is being treated with its respective bioactive compound, some cells contain only BFA bodies and other contain BFA bodies and other small bodies. This might indicate that the persistent bodies are in a BFA insensitive compartment and that the small molecule is disrupting recycling by effecting this BFA insensitive compartment. Upon validation our model has the ability to truly predict those compounds that can disrupt recycling and those that cannot.

Many endomembrane phenotypes manifest as punctate structures of various sizes and abundances making this approach applicable to multiple scientific

pursuits. Organelles, early endosomes for example, differ in size and abundance from mitochondria, ER bodies, and other compartments (Mukherji and O'Shea, 2014). The GMM approach or some variation of it would likely be able to differentiate between organelles, supporting the evidence of co-localization studies. When added to the pipeline of a high throughput assay this method would also be able to identify those body phenotypes that deviate from the normal phenotype, identifying mutants of bioactive compounds. It is important to note that this model was developed using body size and number per cell. However, other parameters could be used to create different two-dimensional models or additional features could be added to the existing model to produce multi-dimensional models, allowing for a more accurate classification. With minimal modifications, this approach could be used for many diverse applications.

Large amounts of quantitative data need to be managed and interpreted efficiently and rapidly. Machine learning approaches such as the described GMM approach allowed biologists to make sense of phenotypic data while gaining the ability to predict the underlying corresponding biological process based on complex phenotypes. Using the information provided, biologists seeking to identify compounds that disrupt recycling can now focus on further testing and characterizing four compounds rather than testing the full three hundred and sixty compound library, saving time, resources and expediting discovery. Methods like this will enable cell biologist to more effectively discover target genes and

corresponding small molecules as the technology evolves and assays become increasingly higher in throughput and complexity.

4.5 Materials and Methods

Gaussian Mixture Model and clustering

Images were collected from the previous screen of three hundred and sixty compounds (drakakaki). 40 Images were chosen based on quality and analyzed using the software package EndoQuant (Ung et al. In press). The computation and modeling was done using MatLab by Mathworks version R2010b. The function `gmdistribution` was used to construct the GMM fitting the data to three Gaussian distributions. Subsequently, The function `cluster` was then used to cluster the resultant data from the GMM.

Seedling growth conditions

Arabidopsis thaliana ecotype Colombia (col-0) seedlings were sterilized then sown on 0.5X MS containing media with 0.8% phytoagar. Seedlings were stratified for 48hrs at 4°C then grown on an 18hr light cycle at 22°C. Original chemical screen and methods for imaging PIN2:GFP individuals can be found in Drakakaki et al. (2011).

BFA Washout Assay and Microscopy

PIN2:GFP seedlings were incubated in a solution of .5X liquid MS media containing 10 μ M BFA for 1 hour. The seedlings were then dipped in .5X MS media and then transferred to a well containing .5x MS liquid media and one of the cluster one and cluster two compounds or DMSO for 80 min. The seedlings were then transferred to microscope slides for imaging. Seedlings were imaged using a Leica SP5 Confocal Laser Scanning Microscope (CLSM). Excitation wavelengths used were 488nm (GFP). Manufacturer settings were used for detection of fluorescence. All confocal microscopy was done at the Microscopy and Imaging Core at the Center for Plant Cell Biology (CEPCEB) of the Institute for Integrative Genome Biology (IIGB) at University of California at Riverside.

4.6 References

Conrad, C., Erfle, H., Warnat, P., Daigle, N., Lörch, T., Ellenberg, J., Pepperkok, R., and Eils, R. (2004). Automatic identification of subcellular phenotypes on human cell arrays. *Genome Res* 14, 1130-36.

Desprez, T., Vernhettes, S., Fagard, M., Refrégier, G., Desnos, T., Aletti, E., Py, N., Pelletier, S., and Höfte, H. (2002). Resistance against herbicide isoxaben and cellulose deficiency caused by distinct mutations in same cellulose synthase isoform CESA6. *Plant Physiol* 128, 482-490.

Drakakaki, G., Robert, S., Szatmari, A.M., Brown, M.Q., Nagawa, S., Van Damme, D., Leonard, M., Yang, Z., Girke, T., et al. (2011). Clusters of bioactive compounds target dynamic endomembrane networks in vivo. *Proc Natl Acad Sci U S A* 108, 17850-55.

Elias, M. (2010). Patterns and processes in the evolution of the eukaryotic endomembrane system. *Mol Membr Biol* 27, 469-489.

Hartmann, A., Czauderna, T., Hoffmann, R., Stein, N., and Schreiber, F. (2011). HTPPheno: an image analysis pipeline for high-throughput plant phenotyping. *BMC Bioinformatics* 12, 148.

Hendricks, L.C., McClanahan, S.L., Palade, G.E., and Farquhar, M.G. (1992). Brefeldin A affects early events but does not affect late events along the exocytic pathway in pancreatic acinar cells. *Proceedings of the National Academy of Sciences* *89*, 7242-46.

Klukas, C., Chen, D., and Pape, J.M. (2014). Integrated Analysis Platform: An Open-Source Information System for High-Throughput Plant Phenotyping. *Plant Physiol* *165*, 506-518.

Li, C.-J., and Bangerth, F. (1999). Autoinhibition of indoleacetic acid transport in the shoots of two-branched pea (*Pisum sativum*) plants and its relationship to correlative dominance. *Physiologia Plantarum* *106*, 415-420.

Li, J., Nam, K.H., Vafeados, D., and Chory, J. (2001). BIN2, a new brassinosteroid-insensitive locus in *Arabidopsis*. *Plant Physiol* *127*, 14-22.

Mathew, M.D., Mathew, N.D., and Ebert, P.R. (2012). WormScan: a technique for high-throughput phenotypic analysis of *Caenorhabditis elegans*. *PLoS One* *7*, e33483.

Miller, S.G., Carnell, L., and Moore, H.H. (1992). Post-Golgi membrane traffic: brefeldin A inhibits export from distal Golgi compartments to the cell surface but not recycling. *J Cell Biol* 118, 267-283.

Mukherji, S., and O'Shea, E.K. (2014). Mechanisms of organelle biogenesis govern stochastic fluctuations in organelle abundance. *Elife* 3, e02678.

Richter, S., Voss, U., and Jürgens, G. (2009). Post-Golgi traffic in plants. *Traffic* 10, 819-828.

Sommer, C., and Gerlich, D.W. (2013). Machine learning in cell biology - teaching computers to recognize phenotypes. *J Cell Sci* 126, 5529-539.

Vogt, A., Cholewinski, A., Shen, X., Nelson, S.G., Lazo, J.S., Tsang, M., and Hukriede, N.A. (2009). Automated image-based phenotypic analysis in zebrafish embryos. *Dev Dyn* 238, 656-663.

Chapter 5

A Novel Small Molecule Reduces Anisotropic Growth as a Result of Decreased Cellulose Deposition in *Arabidopsis* Root Cells

5.1 Abstract

One of the fundamental differences between animal and plant cells is the development of a cellulosic cell wall. The cell wall provides structural support, mechanical defense against pathogens and it has been implicated in signaling and is essential for growth. Cellulose is synthesized by cellulose synthase complexes (CSC's) that are transported to and inserted into the plasma membrane. Exactly how the CSC's reach the membrane is not well understood. We adopted a chemical genomics approach to identify compounds that altered root length and endomembrane trafficking. D5, a novel small molecule was shown to drastically decrease root length and caused cells to swell, a classic indicator of an inability to produce cellulose. Subcellular localization of CSC subunits were unable to reach the PM when plant seedlings were treated with D5 which seems to be preventing proper CSC localization to the PM without disturbing cytoskeletal dynamics. An EMS mutant was found to be resistant to D5 reversing the cell swelling phenotype associated with the D5 treatment. Understanding the mode of action of D5 will perhaps give us some insights into the transport of CSC's to the PM.

5.2 Introduction

One of the fundamental differences between plant and animal cells is the production and maintenance of a cell wall. Cell walls surround the plasma membrane (PM) of the cell providing structural support, protection and even contribute to signaling and sensing the plant's environment. Cell walls can be found in multiple taxa and are similar in function yet differ in composition. Cell walls found in fungi typically have cell walls that are largely composed of chitin, whereas the main constituent of plant cell walls is cellulose (Lipke and Ovalle, 1998). Cellulose, a 1,4- β -linked glucan found in cell walls is found in the form of microfibrils that interact with other polysaccharide chains to add structural and tensile strength. These cellulose microfibrils are synthesized by hexameric protein complexes known as the cellulose synthase complex (CSC) (Emons and Mulder, 2000). This rosette-like complex is secreted through the Golgi Trans Golgi network (TGN) and SYP61 labeled endosomes to the PM where it is attached to microtubules via Cellulose Synthase Interacting protein 1(CSI) which links CSC's to microtubules. These CSC's then track along microtubules to produce cellulose microfibrils (Bashline et al., 2014). The finer details of how the CSC's travel from the TGN to be finally inserted in the PM is still poorly understood.

Chemical genetics is a growing approach to study gene and protein function using small molecules. This approach uses biologically active compounds to

transiently inactivate proteins in order to study their functions. By using small molecules we can circumvent challenges that are associated with the more classical mutagenesis screen such as genetic redundancy and lethality. Additionally, chemical perturbagens enable tunable control allowing the degree of severity of the phenotype to be modified by adjusting the concentration of the small molecule. By studying the effect of the small molecule we can start to understand the role of the protein or biological process that the small molecule is targeting.

A screen of 46,000 small drug-like molecules in search of endomembrane perturbagens yielded three hundred and sixty compounds that inhibited pollen germination (Drakakaki et al., 2011). Pollen tube growth is highly dependent upon endomembrane trafficking including both exocytosis and endocytosis. Screening for molecules that inhibit pollen tube germination enriched the library for compounds that disrupt endomembrane trafficking. Subsequently these compounds were tested for activity in *Arabidopsis* roots (Drakakaki et al., 2011). One hundred and twenty two of them were found to be active in *Arabidopsis* roots and appeared to effect the recycling of plasma membrane markers.

Several small molecules are known to inhibit cellulose synthesis and are often used as herbicides such as isoxeben and 2,6-Dichlorobenzonitrile (DCB) (Sabba and Vaughn, 1999). Mutations in the cellulose synthase subunit CESA6 were found

to confer resistance to Isoxeben, shedding light on its biological target and transforming Isoxeben into a useful tool for studying cellulose synthase (Desprez et al., 2002). Although the exact mechanism of action is not known, it seems that DCB reduces cellulose synthase mobility (DeBolt et al., 2007). In one case, a small molecule directly inhibits cellulose synthesis and the other alters the dynamics and possibly the ultimate destination of the CESA complexes.

A novel compound, here on referred to as D5, was found to dramatically reduce the root length of Arabidopsis seedlings due to a reduction of cellulose delivery to the PM. The results of these experiments provide insight into the link between cellulose synthesis and auxin homeostasis and establish D5 as an inhibitor of proper cellulose deposition.

5.3 Results

D5 Inhibits Root Growth due to cell swelling

To identify cellular factors involved in the maintenance and integrity of the Early Endosome (EE), the 122 small molecules that perturbed the recycling of known plasma membrane (PM) localized proteins, were assayed for bioactivity by observing the phenotypic effect on a EE localized fluorescent marker, SYP61:CFP. Because the EE is a hub for recycling, retrograde trafficking and anterograde trafficking, the EE is likely to have redundant mechanisms in place to maintain its

proper function. Therefore, chemical genomics is an appropriate tool to circumvent these problems of redundancy and lethality to uncover novel molecular components. D5 is a small molecule that seemingly disrupted the trafficking of SYP61:CFP, a phenotype that was supported by the previous screen, hence its inclusion in the PMRA library, a collection of small molecules affecting proper endomembrane trafficking in *Arabidopsis* roots. The previous study suggested that D5 causes small aggregates of PIN2:GFP to form in the PIN2:GFP marker line. The *growth* phenotype however, was stable and dramatic.

Those compounds that have been found to disrupt recycling have also been found to reduce root length. Therefore, reductions in root length can be used as a possible indicator for disruption of endomembrane trafficking. When grown on D5 containing MS media, *Arabidopsis* seedlings showed dramatic reductions in root length and an increase in root width when compared to seedlings grown on DMSO containing media (Figure 5.1 A and B). Closer examination revealed a swollen root tip and extremely disrupted cellular organization (Figure 5.1 C and D). Canonical meristematic, transition, elongation, and differentiation zones were reduced to the point of being indistinguishable. While the D5 treated roots still seemed to produce root hairs, the number was dramatically reduced and many of them burst. In addition to the root phenotypes, the plants were clearly stressed due to the accumulation of anthocyanins in the cotyledons (Figure 5.1 D).

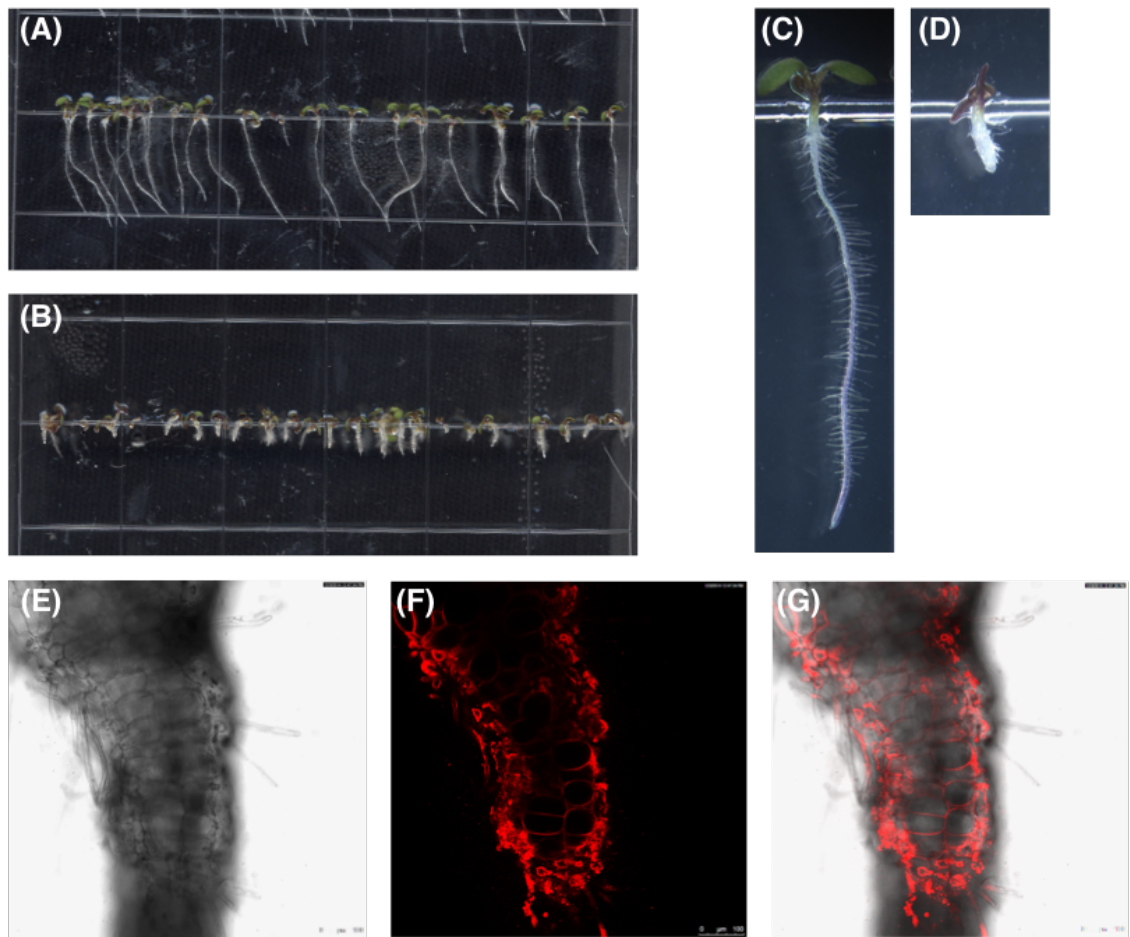


Figure 5.1. D5 causes severe reduction in root length and cell swelling. (A) fourteen day-old seedlings grown on MS media containing DMSO. (B) Fourteen day old seedlings grown on MS media containing 20 μ M D5. (C) Fourteen day-old seedling grown on MS media containing DMSO imaged at 2X magnification. (D) Fourteen day old seedlings grown on MS media containing 20 μ M D5 imaged at 2X. (E) Brightfield image of a D5 treated root. (F) D5 treated root stained using the styryl dye FM4-64. (G) merged image.

Examination of the cellular morphology showed that the cells had not simply arrested their development but were swollen well beyond the average *Arabidopsis* root cell size. Seedlings were then treated with the styryl dye FM4-64 to specifically label the PM confirming the boundaries of the distended cells. Under normal conditions swelling is very tightly controlled and only used to expand cells when growing. Swollen cells may indicate an inability to divide or defects in the structure of the cell wall.

Cellulose content is reduced in D5 treated root cell walls

This small swollen root phenotype is strikingly reminiscent of Isoxaben treated seedlings, a proven inhibitor of cellulose synthase subunits (Desprez et al., 2002). The inability to produce cellulose synthase is consistent with the swollen cell phenotype due to the inability to withstand the outward forces of turgor pressure exerted on the weakened cell walls.

Calcafluor, a common fluorescent dye that is used to visualize cellulose and chitin was used to visualize the amount of cellulose within the root. Root cells were stained with two concentrations of Calcafluor and imaged, a bright signal indicating a higher concentration of cellulose. D5 treated roots clearly show a reduced amount of cellulose indicated by the reduced pixel intensity shown by the image and the histogram (Fig 2 (B) and (D)).

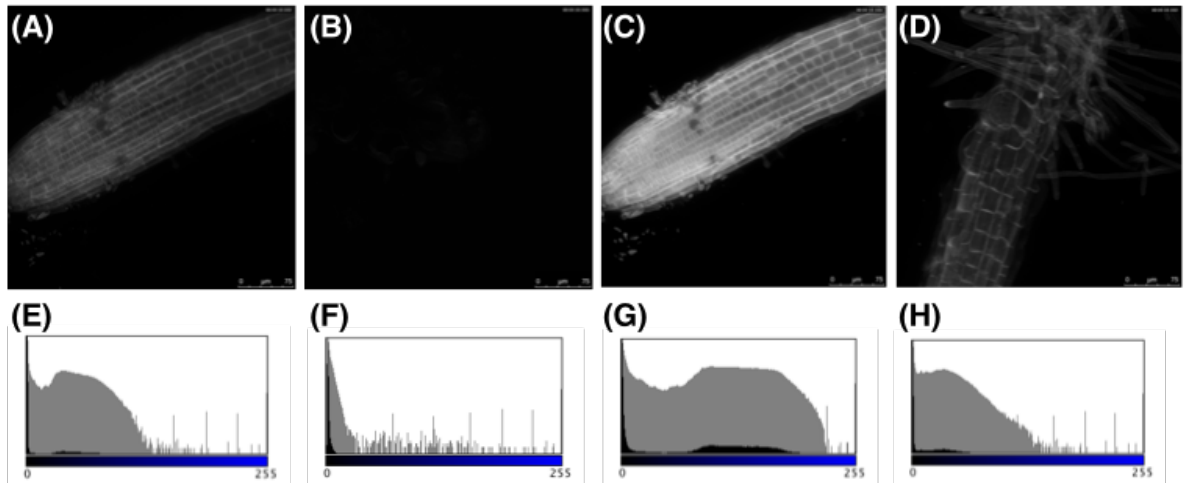


Figure 5.2. D5 reduces cellulose abundance in root cells. (A) and (B) *Arabidopsis* root stained using 9 μ M calcofluor white. (A) treated with DMSO for 24hrs and (B) treated with D5 20 μ M for 24hrs. (C) and (D) *Arabidopsis* root stained using 30 μ M calcofluor white. (C) treated with DMSO for 24hrs and (D) treated with D5 20 μ M for 24hrs. (E) The corresponding intensity histogram for (A). (F) The corresponding intensity histogram for (B). (G) The corresponding intensity histogram for (C). (H) The corresponding intensity histogram for (D).

To confirm this we examined the localization of Cellulose synthase complex subunits CESA3 and CESA6 fused to a GFP fluorescent marker. Untreated roots showed CESA3:GFP signal in both intracellular compartments and the cell wall in both the meristematic and elongation zones. Upon 24hr treatment of roots with D5, the CESA3:GFP signal in the intracellular compartments was unchanged, however, the signal in the cell wall disappeared. In the elongation zone, the disappearance of CESA3:GFP signal was accompanied by noticeable swelling of

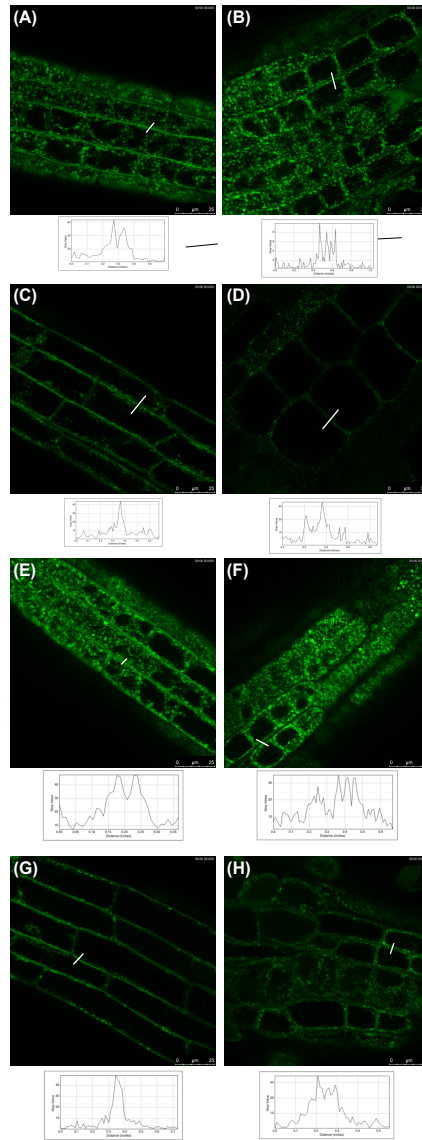


Figure 5.3. CESA3 and CESA6 miss-localize in the presence of D5. (A) - (D) CESA3:GFP in the expansion zone, (A) and (B) and transition zone, (C) and (D) or Arabidopsis roots. (A) and (C) show seedlings treated with DMSO for 24hrs and (B) and (D) show seedlings treated with 20 μ M D5 for 24hrs. (E) - (H) CESA6:GFP in the expansion zone, (E) and (F) and transition zone, (G) and (H) or Arabidopsis roots. (E) and (G) show seedlings treated with DMSO for 24hrs and (F) and (G) show seedlings treated with 20 μ M D5 for 24hrs. White bars indicate intensity profile line. Below each panel is the corresponding intensity profile of the line in each image.

root cells. The cell swelling phenotype observed previously at 6 days is observable at high magnification as early as 24hrs. CESA6:GFP showed a similar phenotype in the presence of D5 in both the meristematic and elongation zones to that of CESA3:GFP. These data suggest that the intracellular trafficking of the CESA subunits is not effected but that they are not reaching the cell wall, which might explain why the cells are behaving as is cellulose synthase activity is diminished.

Cytoskeleton Dynamics are not affected by D5

Mutant *Arabidopsis* seedlings that have disrupted microtubule activity can also exhibit cell swelling and isotropic growth similar to that of the D5 treated seedlings, due to the lack of the structural support of cellulose microfibrils (Sugimoto et al., 2003). This is due to the fact that microtubules are tightly associated with cellulose microfibrils, the product of cellulose synthase complexes. To test D5's ability to alter microtubule dynamics, we treated seedlings with either DMSO or D and observed the microtubule organization in hypocotyl cells expressing TUB:GFP. Similar results were seen in root cells (data not shown). Cells were treated with oryzalin as a positive control for microtubule depolymerization. Oryzalin treated cells showed bright punctate spots as well as a reduction of microtubules spanning the width of the cell, indicating microtubule depolymerization.

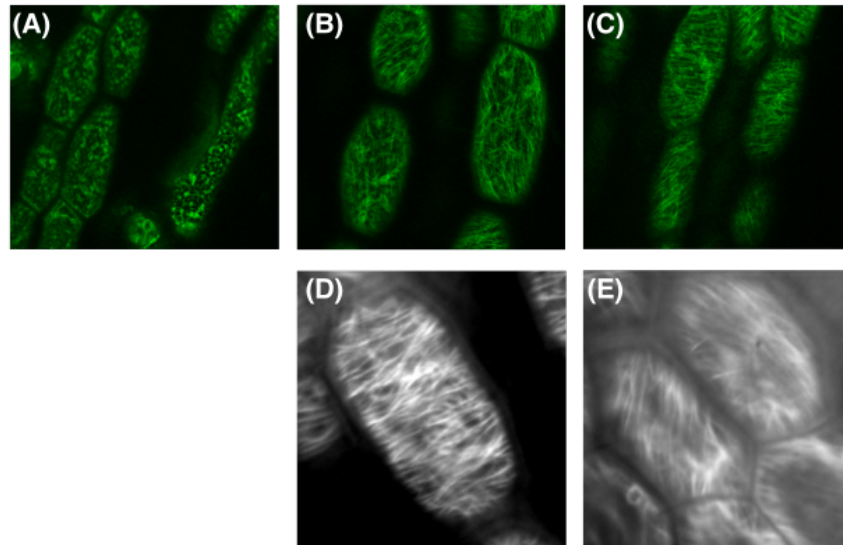


Figure 5.4. D5 does not affect microtubule dynamics. A TUB:GFP seedlings treated with (A) 10 μ M Oryzalin as a positive control (B) DMSO as a negative control and 20 μ M D5. (D) and (E) show Averaged frames of a time course showing MT dynamics of (D) DMSO and (E) D5 treated seedlings.

Cells treated with DMSO and D5 show intact microtubules and a similar organization, indicating that D5 most likely does not have microtubule depolymerizing activity.

Because single images would not show if D5 was a microtubule stabilizing agent, video of microtubule dynamics were taken using a spinning disk confocal microscope. Fig 4 (D) and (E) shows the average of all the frames taken of these videos. The blurry nature of both images indicate microtubule movement. These videos showed dynamic movements of TUB:GFP labeled microtubules, indicating that D5 neither causing the depolymerization of microtubules nor the stabilization

of microtubules, therefore D5 must be acting in an alternative way to prevent the delivery of CESA subunits to the cell wall. As with microtubules, actin microfilament dynamics were not altered when observed by spinning disk microscopy. Microfilaments were neither depolymerized or stabilized indicating that D5 was not altering microfilament dynamics (data not shown).

The 2.4 mutant shows resistance to D5

Using a mutagenesis screen to identify individuals that are resistant to a drug like molecule is a common method for the identification of genes required associated with the drug target (Raikhel and Pirrung, 2005). To this end we performed a resistance screen using an EMS mutagenized population using root length as a macroscopic discriminatory marker to select for resistant individuals. EMS mutant 2.4 was selected as it showed resistance to D5 manifesting as increased root length, larger cotyledons and reduced anthocyanin production. The increase in root length of resistant individuals was statistically significant (Fig 5.5E). Upon closer examination it became clear that the root tips have normal developmental zones, highly organized cell files and reduced isotropic growth (Fig 5.5A - 5.5D).

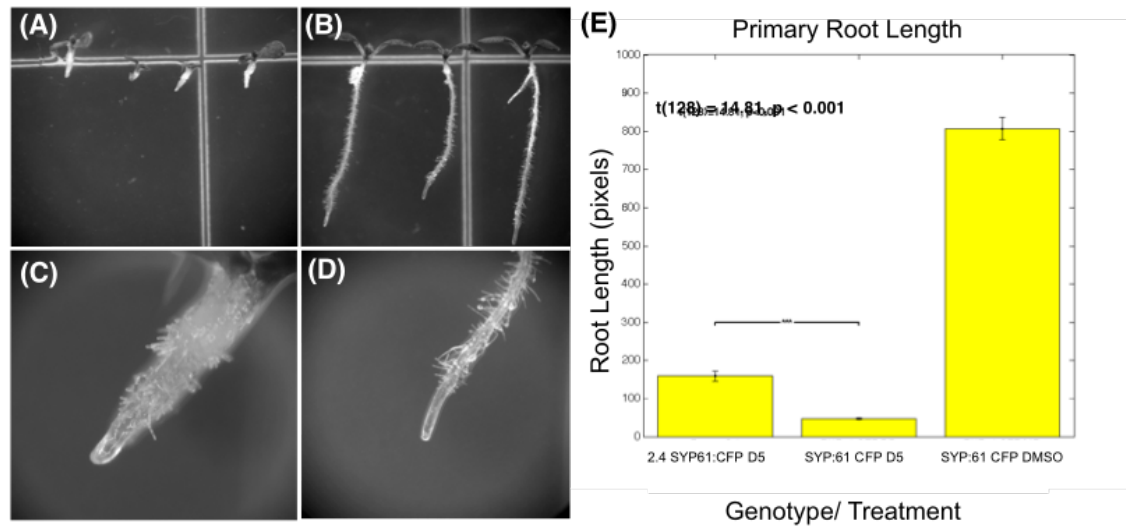


Figure 5.5. 2.4 is a mutant resistant to the effect of D5. Images of D5 treated WT seedlings (A and C) and 2.4 resistant mutant seedlings (B and D) at 2x (A and B) and 6.5x (C and D). (E) The root length of D5 treated 2.4 mutant and WT seedlings and DMSO treated seedlings. 2.4 SYP61 root length is significantly longer than WT.

When grown on the control MS media, the 2.4 mutant seedlings do not show any obvious developmental phenotype at two weeks when compared to the WT control grown under the same conditions. However, when grown to maturity, flowering *Arabidopsis* plants show a loss of apical dominance, a phenotype associated with reduced auxin production or perception (data not shown) (Cline, 1997).

Auxin mutants show resistance to D5

Because mature *2.4 mutant* plants exhibited a loss of apical dominance, a phenotype associated with reduced auxin production or perception, we grew auxin

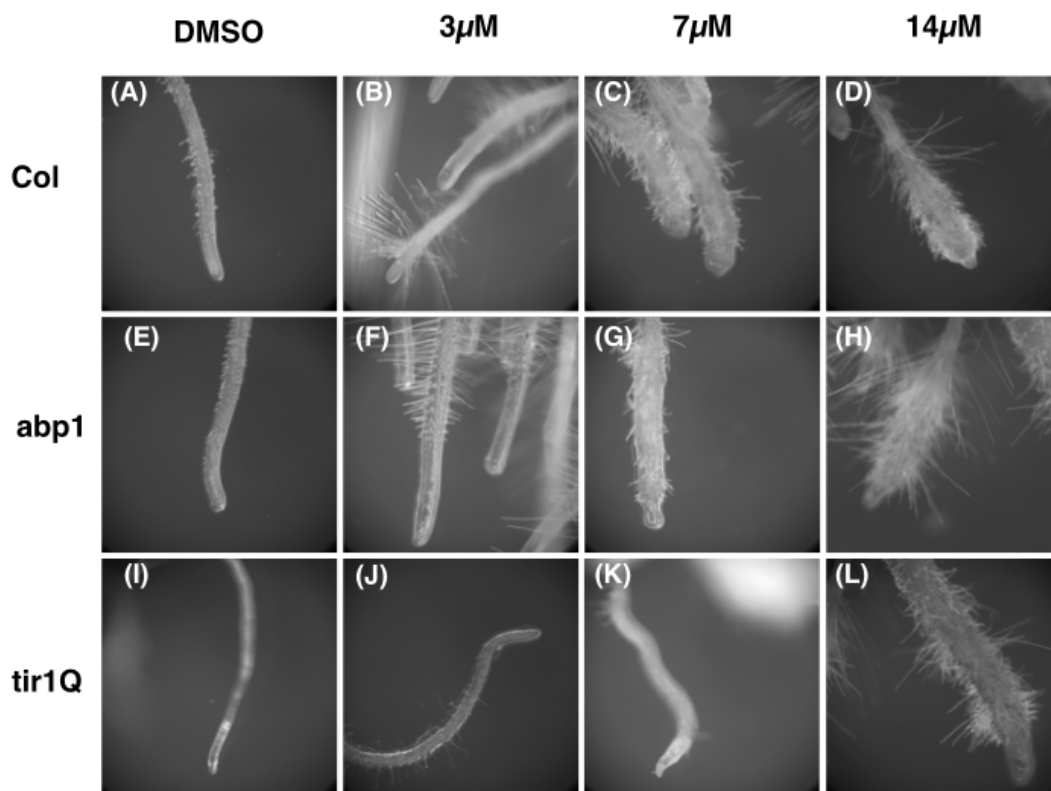


Figure 5.6 Auxin mutants show resistance to D5. (A-D) col WT grown on D5. (E-H) *abp1* mutants grown on D5. (I-L) *tir1Q* quadruple mutant grown on D5. (A,E,I) seedlings treated with DMSO. (B,F,J) seedlings treated with 3 μ M D5. (C,G,K) seedlings treated with 7 μ M D5. (D,H,L) seedlings treated with 14 μ M D5.

perception mutants *abp-1* and *tir1Q* on MS media for four days, then transferred them to media containing various concentrations of D5 to determine if these known auxin perception mutants were resistant to D5. ABP1 and TIR1 are both auxin receptors governing two separate auxin response pathways, ABP1

mediates cell wall remodeling and cell expansion at the PM and TIR1 in the nucleus though they have overlapping functions as well (Paque et al., 2014).

Abp1 mutants did not show any resistance to D5 and behave similarly to Col seedlings when grown on D5 on increasing concentrations of 3 μ M, 7 μ M, and 14 μ M (Figure 5.6A – 5.6H). Both *abp1* mutant and WT seedlings showed very little effect on the 3 μ M concentration, other than a modest increase in the occurrence of root hairs. Swelling of the root tip begins to become apparent when grown on 7 μ M concentrations of D5. At 14 μ M of D5, root tips swell increasing root tip width and root hair density (Figure 5.6D and 5.6H). In contrast to *col* and *abp1* genotypes, the *tir1Q* quadruple mutants showed a slight resistance to D5 at lower concentrations. The *tir1Q* mutants showed little to no swelling when grown on 3 μ M and 7 μ M concentrations of D5 (Fig 5.6J and 5.6K). These results show that D5 may be acting within the TIR1 pathway and not the ABP1 pathway giving us the ability to chemically dissect these pathways. However, it is important to note that the observed resistance of the *tir1Q* mutants is weaker than that of the 2.4 mutants when grown on a higher concentration (Figure 5.6L). Additionally, the *tir1Q* mutant phenotype at maturity is more severe than that of the 2.4 mutant. This suggests that the auxin pathway may not be the only pathways affected.

5.4 Discussion

D5's dramatic growth phenotype is caused by reduced cellulose.

Unlike their animal counterparts, plant cells have evolved a rigid cell wall made out of polysaccharides that provide structure and protection. This cell wall is mainly comprised of the organic polymer cellulose, which is synthesized by hexameric protein complexes termed cellulose synthase complexes. Arabidopsis seedlings grown on D5 a novel small molecule show significant developmental defects associated with reduction in cellulose synthesis including dramatic reduction in root length, reduced cotyledon size, swollen cells, and excessive anthocyanin production. Anthocyanin production is a well-documented response to abiotic stress (Paque et al., 2014). It is possible that this could be a direct response of cell wall signaling, however, it is more likely that this is in response to reduced water and nutrient uptake, due to malformed roots and reduced number of root hairs. If D5's effect on cellulose synthesis was global and non-specific, we would expect root hairs to swell as well, however, this is not observed. The number of root hairs is diminished but this could just be due to the fact that the total number of cells has decreased. This indicates that D5 specifically effects cellulose production of non-root hair cells.

Interestingly, this compound was discovered in a screen that inhibited pollen germination so it is reasonable to conclude that D5 affects the cell wall in pollen

tubes, root cells, but not in root hairs cells. Although, pavement cell morphology was not observed, cotyledon size and health was reduced, most likely indicating a cell expansion and division defect similar to that seen in the roots. This indicates that D5 is selectively inhibiting cellulose production in cell type specific manner. Similar selectivity is found in other cellulose inhibitory compounds such as isoxaben and DCB (Sabba and Vaughn, 1999). However, CESA3 and CESA6 mutants do not show any resistance to D5 indicating that D5 is not directly targeting these subunits (data not shown). Phenotypic analysis therefore points to the lack of cellulose as the causal factor.

Indeed reduced cellulose as visualized by calcafluor staining showed that Arabidopsis roots treated with D5 for 24hrs showed reduced levels of cellulose. There exists a delicate balance between the internal and external forces of the cell. This delicate balance is what maintains the cells shape. Cells intentionally manipulate this balance in order to grow. Enzymes intentionally loosen the bonds between the fibers that compose the cell wall allowing the force of the internal turgor pressure to exceed that of the cell wall allowing the wall to expand. Under normal conditions this is a highly controlled process and the degree of swelling is minimal. Upon treatment with D5 the amount of swelling in root cell is dramatic and isotropic. This change in shape and size is apparent as early as 24hrs.

Impaired CESA delivery to the cell wall is cytoskeletal and endomembrane independent.

Our understanding of cellulose synthesis by cellulose synthase complexes and the trafficking of these complexes have suddenly grown over recent years thanks to advancements in live cell imaging and microscopy. It is known that cellulose synthase complexes are trafficked from the ER to the PM through the Golgi and syp61 compartments (Drakakaki et al., 2011). Additionally, CESA complexes attached to and track along microtubules synthesizing cellulose microfibrils (Paredes et al., 2006). However, gaps in our knowledge still exist in the fine details of the delivery of cellulose to the PM which then allows for the synthesis of cellulose microfibrils. When we visualized the localization of CESA3 and CESA6 we did not find any obvious defects in the intracellular trafficking of these subunits. All intracellular compartments were intact and no major differences in localization when compared to the untreated control. However, it was very apparent that the GFP labeled CESA subunits were not reaching the PM. It is possible that the target of D5 could be responsible in the finer details of this process of delivering the CESA subunits to the PM either acting as an adaptor protein or in a specific membrane bound compartment downstream of SYP61 endosomes delivering the CESA complexes to the PM.

Because microtubule dynamics are not altered, yet the CESA complexes are not getting to the membrane, D5 must be acting within this narrow functional range. It is most likely inhibiting the association of the CESA subunits with the membrane. However, we cannot definitively rule out that this compound could also be an inhibitor of cellulose synthase directly by binding to one or more of the subunits while it associated with the membrane. *ixr-1* and *ixr-2* both isoxaben resistant mutants, were grown on D5 containing media with no change in response when compared with the WT seedlings (data not shown). As these are both point mutations, we cannot rule out the possibility that D5 is directly targeting CESA6 or CESA3 or both. It would be reasonable to assume that if D5 could directly bind to CESA3 and/or CESA6 this binding could result in a conformational change that alters the fluorescent properties of the GFP tag.

The exocyst is a large octameric , conserved protein complex that facilitates the fusion of exocytic vesicles with the PM. Exocyst mediated CSC insertion is one possible mechanism. It hypothesized that the exocyst is important in depositing CSC's in the membrane though this has not been definitively shown. This is an enticing possibility since other compounds within the 122 found to be active in *Arabidopsis* also disrupt exocyst function or proteins associated with the exocyst (unpublished data). However, those compounds that disrupt exocyst function and therefore fusion with the membrane, induced the formation of anomalous endosomal bodies which are not observed in D5 containing compounds.

D5 seems to phenocopy other cellulose synthase inhibitors yet does not share any structural similarity. It also seems to be much less potent than isoxeben or DCB which are active at nanomolar concentrations in contrast to D5 which is active in micromolar concentrations. Isoxeben and DCB are both powerful herbicides, D5 might be a useful addition to this arsenal in a climate when plants are constantly evolving resistance to such pesticides.

Auxin might be increasing cell wall extensibility

Both, the auxin-mutant-like phenotype of the 2.4 D5 resistant mutant and the fact that the tirQ mutant shows slight resistance suggests that D5 might be targeting a protein involved in auxin production or perception. The structure of D5 does not resemble the structure of auxin or any other auxin analogs. Additionally, D5 does not induce the expression of DR5:GFP constructs indicating that D5 is most likely not an auxin analog (data not shown). Furthermore, exogenous auxin does not produce a swollen phenotype to the extent of that of D5. Relatively high concentrations of auxin are known to inhibit cell elongation in roots, as in the mechanism known to cause the gravitropic response. However, recently it was found that changes in auxin homeostasis suppress the cell swelling phenotype in *Arabidopsis* roots commonly associated with defects in cell wall function (Steinwand et al., 2014).

The acid growth hypothesis postulates that auxin increases the acidity of the cell wall, thereby activating expansins, enzymes that disrupt the bonds holding cellulose and hemicellulose in the cell wall and allowing the cell wall to expand under the force of the turgor pressure within the cell (Hager, 2003). Therefore it is possible that without auxin, the extensibility of the cell wall is reduced, preventing the cell expansion that normally accompanies defect in cellulose synthesis. Interestingly, ABP1 is likely to play an important role in this by regulating ATPase H⁺ pumps and K⁺ pumps at the PM (Hager, 2003). Therefore, if this was the only factor, we would have expected *abp1* mutants to show resistance to D5, however, this was not the case.

In contrast *tir1Q* mutants do show some resistance to D5 indicating that the swelling phenotype might be more dependent on the TIRQ pathway than the ABP1 pathway of auxin sensing. Both the ABP1 and TIR1 auxin sensing pathways are thought to have some effect on Cell wall remodeling and integrity (Paque et al., 2014) (Tsang et al., 2011). D5 may provide a way to selectively disrupt the TIR1 signaling pathway leading to cell wall remodeling. It is possible that it does this by regulating the delivery of CESA subunits to the cell wall itself, which we know D5 can inhibit. It is possible that D5 is perturbing a component of TIR1 mediated cell wall expansion. Sequencing the 2.4 resistant plants to determine the nature of their resistance will be vital in understanding the activity of D5 and its potential target(s).

5.5 Materials and Methods

Seedling growth conditions

Arabidopsis thaliana ecotype colombia (col-0) seedlings were sterilized, then sown on .5XMS containing media with .8% phytoagar. Seedlings were stratified for 48hrs at 4°C then grown on an 18hr light cycle at 22°C. Treated seedlings were grown on media containing 20µM D5. Seedlings screened for resistance to D5 were grown for fourteen days before being imaged and scored for root length.

Chemical Treatments Microscopy and Image analysis

Four to seven-day-old seedlings were treated with Calcofluor, FM4-64, D5, Oryzalin at the concentrations previously mentioned. Seedlings were transferred from media plate to .05X liquid MS media and treated for 2hrs for D5, 10 for FM4-64, and 24hrs for calcofluor. CESA3:GFP and CESA6:GFP seedlings were treated for 24 hrs then imaged using the Leica SP5 LSCM. Cytoskeleton images were taken using the Yokagawa spinning disk confocal microscope. Excitation wavelengths used were 488nm (GFP) and 543(FM4-64). Calcafluor was visualized using the UV laser. Manufacturer settings were used for detection of fluorescence. Whole plate images were taken using a desktop scanner at 1000dpi. Single seedling images were taken using a Nikon dissecting scope. Mature whole plants were imaged using a Rebel DSLR camera. Fiji was used to for subsequent image analysis.

Resistance screen

For the primary screen 50,000 mutagenized *Arabidopsis* seedlings in the SYP61:CFP background were screened for D5 resistance on plates containing 20 μ M D5. Seedlings were grown for fourteen days, imaged using a desktop scanner and resistant individuals were rescued to soil and grown to maturity. The candidates were grown to maturity, seeds were collected and screened again. Those that were confirmed resistant were crossed with Landsberg to create the mapping population for sequencing.

5.6 References

- Bashline, L., Li, S., and Gu, Y. (2014). The trafficking of the cellulose synthase complex in higher plants. *Ann Bot*
- Cline, M. (1997). Concepts and terminology of apical dominance. *American Journal of Botany* *84*, 1064-4.
- DeBolt, S., Gutierrez, R., Ehrhardt, D.W., and Somerville, C. (2007). Nonmotile cellulose synthase subunits repeatedly accumulate within localized regions at the plasma membrane in *Arabidopsis* hypocotyl cells following 2,6-dichlorobenzonitrile treatment. *Plant Physiol* *145*, 334-38.
- Desprez, T., Vernhettes, S., Fagard, M., Refrégier, G., Desnos, T., Aletti, E., Py, N., Pelletier, S., and Höfte, H. (2002). Resistance against herbicide isoxaben and cellulose deficiency caused by distinct mutations in same cellulose synthase isoform CESA6. *Plant Physiol* *128*, 482-490.
- Drakakaki, G., Robert, S., Szatmari, A.M., Brown, M.Q., Nagawa, S., Van Damme, D., Leonard, M., Yang, Z., Girke, T., et al. (2011). Clusters of bioactive compounds target dynamic endomembrane networks in vivo. *Proc Natl Acad Sci U S A* *108*, 17850-55.

Drakakaki, G., van de Ven, W., Pan, S., Miao, Y., Wang, J., Keinath, N.K., Weatherly, B., Jiang, L., Schumacher, K., et al. (2011). Isolation and proteomic analysis of the SYP61 compartment reveal its role in exocytic trafficking in Arabidopsis. *Cell Res*

Emons, A.M.C., and Mulder, B.M. (2000). How the deposition of cellulose microfibrils builds cell wall architecture. *Trends Plant Sci* 5, 35-40.

Hager, A. (2003). Role of the plasma membrane H⁺-ATPase in auxin-induced elongation growth: historical and new aspects. *J Plant Res* 116, 483-505.

Lipke, P.N., and Ovalle, R. (1998). Cell wall architecture in yeast: new structure and new challenges. *J Bacteriol* 180, 3735-740.

Paque, S., Mouille, G., Grandont, L., Alabadí, D., Gaertner, C., Goyallon, A., Muller, P., Primard-Brisset, C., Sormani, R., et al. (2014). AUXIN BINDING PROTEIN1 links cell wall remodeling, auxin signaling, and cell expansion in arabidopsis. *Plant Cell* 26, 280-295.

Paredez, A.R., Somerville, C.R., and Ehrhardt, D.W. (2006). Visualization of

cellulose synthase demonstrates functional association with microtubules. *Science* 312, 1491-95.

Raikhel, N., and Pirrung, M. (2005). Adding precision tools to the plant biologists' toolbox with chemical genomics. *Plant Physiol* 138, 563-64.

Sabba, R.P., and Vaughn, K.C. (1999). Herbicides that inhibit cellulose biosynthesis. *Weed Science* , 757-763.

Steinwand, B.J., Xu, S., Polko, J.K., Doctor, S.M., Westafer, M., and Kieber, J.J. (2014). Alterations in auxin homeostasis suppress defects in cell wall function. *PLoS One* 9, e98193.

Sugimoto, K., Himmelspach, R., Williamson, R.E., and Wasteneys, G.O. (2003). Mutation or drug-dependent microtubule disruption causes radial swelling without altering parallel cellulose microfibril deposition in *Arabidopsis* root cells. *The Plant Cell Online* 15, 1414-429.

Tsang, D.L., Edmond, C., Harrington, J.L., and Nühse, T.S. (2011). Cell wall Integrity controls root elongation via a general 1-aminocyclopropane-1-carboxylic acid-dependent, ethylene-independent pathway. *Plant Physiol* 156, 596-604.

Chapter 6

Conclusions and Future Directions

6.1 An Approach to Quantify Endomembrane Dynamics in Tobacco Pollen Using Bioactive Compounds

One of the critical objectives of small molecule and genetic screens in plants is to gain the ability to observe and quantify subcellular phenotypes that can be eventually related to plant development to gain functional information. This requires the ability to generate subcellular video and static images and analyze the resultant data in an ultimately high throughput manner. We have made progress towards this goal by quantifying vesicle dynamics in plants via the analysis of confocal video images. We have chosen Tobacco pollen as a model system due to its high dependence on rapid trafficking for proper development and the ability to study the effects of small molecules in this free-living organism. We screened a library of 238 molecules found to inhibit pollen germination though have little to no effect on trafficking in *Arabidopsis* roots. To define a subset of molecules that specifically altered endomembrane trafficking in pollen, a marker for the ER to Golgi trafficking pathway, the Ras-like monomeric G-protein RAB2:GFP, was used to confirm the disruption of the endomembrane pathway and exclude other

processes that would effect pollen tube growth such as cell wall or cytoskeleton function. Six compounds were found to repeatedly alter normal RAB2:GFP and were termed RAB effectors (RAE's).

We focused on quantifying the phenotypes of these compounds in video images to capture true real time dynamics *in vivo* of the Rab2:GFP labeled compartments. We examined the correlations of individual parameters such as velocity, straightness of movement, area of the bodies and intensity of the organelles. We found that Area was positively correlated with intensity suggesting that a the amount of RAB2:GFP signal increased with the size of RAB2 bodies. Additionally, straightness was negatively correlated with velocity indicating that bodies traveling in a straight line moved faster than those that had less direct movement patterns. By studying the way these compounds effected localization we were able to quantify and infer relationships of fundamental properties of the endomembrane system. These intrinsic properties of organelle dynamics need to be further explored using many more parameters. Additionally, a principal component analysis or similar method could be used to determine the most important contributing factors. Once more is understood about these intrinsic properties and they are quantified. They could then be integrated into a mathematical model that could be used to describe trafficking patterns, leading to a better understanding of endomembrane trafficking, dynamics.

In an effort to quantitatively compare the sub cellular phenotype effect of one small molecule to another, we implemented a novel metric called the difference index. This quantitative summary value permits a comparative measure of overall vesicle dynamics. This can be used when evaluating phenotypes and designing experiments, targeting those with similar scores. This could be an indication of similar underlying mechanisms.

The six RAE's, when fully characterized may shed light on pollen specific mechanisms of polar growth and endomembrane dynamics. Not only are these small molecules valuable tools to quantify endomembrane dynamics but they are a preselected group of chemical that warrant further study. This opportunity is ripe for the taking by any interested group and will serve as a valuable resource for biologists interested in pollen biology and polar growth.

We have taken an important step in the quantification of dynamic behaviors that will make it possible to move toward the increased automation of chemical and genetic screens that can be focused directly on changes in vesicle dynamics, permitting the more efficient linkage of cellular and developmental phenotypes.

6.2 EndoQuant: An Image Analysis Package for Automated Quantitative Cellular Phenotyping

Innovative tools to manage and automatically quantify cellular phenotypes in a manner that is rapid, accurate and user friendly will become increasingly necessary as biological screens grow in throughput. EndoQuant is a novel user-friendly image analysis package designed by biologists for biologists that will not only allow for rapid processing of data, but it uniquely allows for rapid decision making through instant visualization comparison of data and basic statistical calculations. EndoQuant addresses the issue of comparing quantitative features of organelles between cells and between individual organisms and is applicable to multiple plant cell types. We do this by quantifying the phenotypes on a per cell basis, a biologically relevant approach. The entirety of these useful algorithms are packaged in a user-friendly graphical user interface that is intuitive and designed for biologists that will be freely available to the community.

Data, once collected is not of much use to biologists until its is processed and interpreted. The next vital step would be to use this method to quantitative compare multiple phenotypes. This analysis would be exceptionally powerful when coupled with a high throughput screen for the identification of endomembrane phenotypes. Indeed, EndoQuant would not only increase the practicality of high throughput subcellular screens but also aid in distinguishing

subtle differences between cellular phenotypes, identifying those that might have otherwise been overlooked.

EndoQuant has the capacity to analyze temporal dynamics if it is given a time series. One would be able to track the changes in population size and number and any other of the features measured. However, to be more completely analyze dynamics, bodies would need to be tracked from one position to another. This would be the next step for this type of analysis package. The single cell analysis coupled with tracking would truly be unique among analysis software. Similarly, Endoquant is limited to two dimensions when analyzing a single image slice or attempting to infer 3 dimensions from a compressed two dimensional slice when analyzing MIP's. To move toward a scenario that is closer to nature, three dimensional reconstructions need to be analyzed. More information can be garnered by increasing the dimensionality of the analysis however, With higher dimensionality comes increased complexity. Future modifications of this tool will need to include these additional dimensions.

Technology such as EndoQuant will allow for high-throughput assays at the cellular level to become practical, permitting the investigation of phenotypes and biological phenomena that do not have an obvious macroscopic phenotype. It will also enable biologists to collect and interpret large amounts of phenotypic data that can be later used to influence experimental designs or be incorporated into

models Automated tools like EndoQuant will become increasingly imperative for dissecting cellular phenomena as quantification becomes increasingly essential. The rapid linkage of cellular and developmental phenotypes will depend on the automatic quantification of subcellular phenotypes, making it possible to move toward the increased automation of chemical and genetic screens.

6.3 Prediction of small molecules using phenomics and machine learning

Biological assays are becoming increasingly high throughput in nature with the advent of the age of “-omics”. In the Case of Chemical Genomics, small molecules are used to perturb and study conserved biological processes such as membrane recycling are prone to lethality and redundancy. The targets of these small molecules are often unforeseeable and require an extensive amount of characterization and experimentation to uncover. We now have the ability to predict those molecules that are of interest to us, in this case, those that disrupt endomembrane trafficking using a GMM to classify and predict their function based on the resulting phenotype. This machine learning approach will enable us to intelligently design experiments expediting the screening process, focusing our efforts on a subset of predicted hit compounds. Generally every phenotype can be grouped into one of three clusters cluster one has few small bodies, cluster 2 has many large bodies and cluster 3 has few large bodies. Though increasing the

sample size would bring us closer to revealing the true nature of these compounds and their phenotypes, our model shows high confidence in the classification of three subgroups of body phenotypes. ES2, a well-characterized inhibitor of recycling was found to group with cluster 2. We hypothesized that it might be possible that the other compounds that produced similar phenotypes to ES2 may also be affecting the endomembrane recycling process. The BFA washout assay was used to confirm that selected compounds from cluster two did indeed inhibit the endomembrane recycling process and that compounds in cluster one did not, thereby confirming the predictive power of this approach. Upon validation our model has the ability to truly predict those compounds that can disrupt recycling and that that cannot. Machine learning approaches such as the described GMM approach allowed biologists to make sense of phenotypic data while gaining the ability to predict the underlying biological process being disrupted based on complex phenotypes.

Through validation this method has proven to be quite reliable and useful, the experiments were on a relatively small sample size. Increasing the sample size will truly test this method's predictive capacity. Other compounds that are known to inhibit specific biological pathways should also be clustered with this data set. If they are classified into an existing cluster, the other members of the same cluster should be validated to see if they also are involved with a similar biological pathway. Within the same theme, cytoskeletal inhibitors could be used to disrupt

trafficking and classified by the GMM. Compounds that exhibit a similar phenotype might also be altering cytoskeletal dynamics and would allow researchers interested in endomembrane specific events to avoid these compounds. Alternatively, these compounds might be of interest to those investigating the cytoskeleton and its many complex functions.

To those studying endomembrane recycling and trafficking, cluster 2 is most exciting. Cluster 2 is enriched for compounds exhibiting phenotypes consistent with ES2 and have shown to alter recycling. These compounds will be of interest to begin to characterize and understand how they reduce the rate of recycling and disrupt trafficking. Without this method, 360 compound would need to be tested using the BFA washout which would be impractical with the amount of time a resources need to complete this task. However, implementing this GMM in this manner has reduced this large pool to four drug-like molecules that are known to inhibit recycling thus more efficiently selecting compounds to characterize and pursue based on our interests.

Machine learning has given us the ability to predict the biological function disrupted by a small molecule simply by analyzing the subcellular phenotype. This algorithm will expedite discovery and give biologists the ability to better concentrate their efforts on compounds that show the most promise according to our predictive computational model.

6.4 A Small Molecule Reduces Anisotropic Growth in Root Cells as a Result of Reduced Cellulose Deposition

The organic polymer cellulose is synthesized by hexameric protein complexes known as cellulose synthase complexes and is the main polysaccharide component of plant cell walls. D5 is a novel small molecule shown to significantly increase developmental defects associated with reduction in cellulose synthesis. However, D5 is incubated with plants that have mutations in CESA3 and CESA6, known subunits of the cellulose synthase complex do not show any resistance to D5 indicating that D5 is not directly targeting these subunits (data not shown).

Cellulose deposition was visualized by calcafluor staining. Arabidopsis roots treated with D5 for 24hrs showed reduced levels of cellulose. This could indicate several possible scenarios in which cellulose is not being synthesized. Two hypothesis are that it is either directly binding to and inhibiting one or more of the CESA subunits or it is indirectly preventing proper function by binding to another protein needed for function or proper localization. Cellulose synthase complexes are delivered to the PM through the endomembrane system from the ER, through the Golgi and SYP61 compartments ultimately arriving at the PM {Drakakaki 2011}. Using video imaging of cytoskeletal dynamics we found that microtubule function are not altered, yet it is very apparent that the GFP labeled CESA subunits were not

reaching the PM when observed by live cell imaging and confocal microscopy. Unexpectedly, we do not observe an endomembrane phenotype when CESA3:GFP or CESA6:GFP plants are treated with D5. Therefore, we hypothesize that D5 must be acting within this narrow functional range possibly inhibiting the association of the CESA subunits with the PM. The target of D5 may be responsible for the finer details of this process, delivering the CESA subunits to the PM acting either as an adaptor protein or in a specific membrane bound compartment downstream of SYP61 endosomes.

A forward genetic resistance screen was used in the Identification of the target of D5. A mutant designated 2.4 was resistant to the effects of D5 showing increased root length when compared to the WT. Additionally, this mutant showed loss of apical dominance, a trait tightly associated with a decrease in auxin signaling. Both, the auxin-mutant-like phenotype of the 2.4 D5 resistant mutant and the fact that the tir1Q mutant shows slight resistance suggests that D5 might be targeting a protein involved in auxin production or perception. The structure of D5 does not resemble the structure of auxin or any other auxin analogs. Additionally, D5 does not induce the expression of DR5:GFP constructs indicating that D5 is most likely not an auxin analog (data not shown). Tir1Q mutants do show some resistance to D5 indicating that the swelling phenotype might be more dependent on the TIR pathway than the ABP1 pathway of auxin sensing. D5 may provide a way to selectively disrupt the TIR1 signaling pathway leading to cell wall remodeling.

However, recently it was found that changes in auxin homeostasis suppress the cell swelling phenotype in *Arabidopsis* roots commonly associated with defects in cell wall function (Steinwand 2014). Sequencing of the 2.4 mutant line will provide clues and insight into the nature of this resistance to D5 and possibly reveal hints as to the target of D5 and how it is disrupting proper cellulose synthesis.