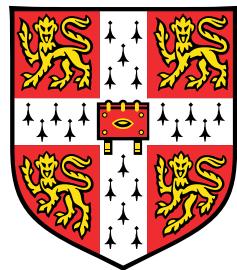


*In vivo dynamics of the *Arabidopsis thaliana* aerial stem cell niche*



Henrik Åhl

Supervisors: Dr José Teles

Professor Henrik Jönsson

Department of Applied Mathematics and Theoretical Physics
University of Cambridge

This dissertation is submitted for the degree of
Master of Philosophy in Computational Biology

Hughes Hall

August 2017

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 18,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Henrik Åhl
August 2017

Abstract

Understanding plant development has implications for fields as wide-ranging as crop yield improvement and regenerative medicine. Aerial development in the plant is driven primarily by the developing center known as the shoot apical meristem (SAM), at the central zone (CZ) of which a small collection of stem cells reside and divide in order to produce cells for regeneration of tissue and new organ formation. The stem cells at the apex express the CLAVATA-3 (CLV3) gene, making the expression of which a direct identifier of the stem cell phenotype. Modern imaging tools and segmentation software have recently enabled extensive capabilities to both accurately measure cells in the SAM, as well as to follow their dynamics over time. In this thesis, we show how a small “true” stem cell niche is maintained in the SAM over time, and that the center of this region is not located at the geometric apex, possibly for phyllotactic priming. We also point at possible regulatory means for the SAM to maintain robustness in expression for the apical cells.

Table of contents

List of figures	ix
List of tables	xi
Nomenclature	xiii
1 Introduction	1
1.1 The shoot apical meristem of <i>Arabidopsis thaliana</i>	1
1.2 Modelling biological systems	2
1.3 Regulatory Mechanics of Plant Stem Cells	4
1.3.1 Molecular tuning determines cell phenotype	4
1.3.2 Developmental Regulation in the Apical Meristem	5
1.4 The role of plant stem cell modelling	6
2 Methodology	7
2.1 Raw data	7
2.2 Processed image data	8
2.3 Data processing	10
2.3.1 Data analysis tools	10
2.3.2 Data filtering	10
2.4 Models of Gene Regulatory Networks	10
2.4.1 Mathematical Formulation	10
2.4.2 Stochastic Simulations	12
2.5 Network modelling approach	15
3 Results	17
3.1 Regulation of the CLV3 domain induces periodicity when perturbed	17
3.2 CLV3 expression is stable for apical cells	18
3.3 Functional clustering implies dsRED technical artefacts	21
3.4 Long-lived cells are amassed centrally in the deep tissue	25
3.5 The CLV3 apex does not coincide with the geometric	28

4 Discussion	31
4.1 NPA dilution induces a periodic response	31
4.2 Distributional cues hints at possible regulatory mechanisms	32
4.3 <i>In vivo</i> tracking sheds light on steady regulation of the true stem cell niche	32
4.3.1 Longevity could be induced by the WUS domain	33
4.3.2 Non-centric CLV3 apex for directed development	33
5 Outlook	35
5.1 Significance of results	35
5.2 Expansion of modelling framework	35
5.3 Data investigation	36
5.4 New experiments	37
References	39
Appendix A Data quality and errors	41
A.1 Tracking quality	41
A.2 Segmentation quality	43
A.3 Missing timepoints	43
Appendix B Supplemental figures	45
Appendix C Software Descriptions	49
C.1 Costanza	49
C.2 MARS-ALT	49
C.3 Organism	50
C.4 extractoR	51
C.5 TissueViewer	52

List of figures

1.1	Graphical illustration of <i>Arabidopsis thaliana</i> and the <i>Shoot Apical Meristem</i>	2
1.2	Modelling biological systems: an illustration	4
1.3	Regulation of apical stem cells	5
2.1	Confocal microscopy data	8
2.2	Segmented data	9
3.1	Number of CLV3 nuclei over time	17
3.2	Raw image CLV3 domain fluctuations	18
3.3	Periodogram of nuclear trajectories	18
3.4	Link between number of nuclei and number of divisions	18
3.5	CLV3 distributions	19
3.6	Epidermis model	20
3.7	Cues of epidermal regulation	21
3.8	Membrane volume clustering	22
3.9	Division rate clustering	22
3.10	Cell line trajectories	24
3.11	Approximation of dsRED decay	25
3.12	Layer-wise age distribution	26
3.13	Clustering of high-longevity cells	27
3.14	Comparison between geometric and genetic apex	28
3.15	Paraboloid time evolution	29
3.16	Example of paraboloid shape evolution	29
A.1	Tracking quality	42
A.2	Membrane segmentation quality per layer	43
B.1	NPA dilution causes primordia formation	45
B.2	Nuclear volumes at apex	46
B.3	Detrended nuclear trajectories	46
B.4	Age distribution, all plants	47

B.5 Non-overlapping apices	48
C.1 extractoR flowchart	51

List of tables

2.1 Filtering settings	11
2.2 Consolidation methods for duplicate nuclei	11

Nomenclature

Acronyms / Abbreviations

AT *Arabidopsis thaliana*

CLV3 CLAVATA-3

CZ Central zone. The region harboring stem cells in the SAM.

dsRED Red fluorescent protein. Used for nuclear CLV3 staining.

GRN Gene regulatory network

KAN1 KANADI-1

L1 Layer-1. The outermost cell layer of the SAM.

L2 Layer-2. See L1.

L3+ Layer-3 and other deep-tissue cells. See L1.

OC Organising center

ODE Ordinary differential equation

PZ Peripheral zone

RAM Root apical meristem

RM Rib meristem

SAM Shoot apical meristem

SDE Stochastic differential equation

SLCU Sainsbury Laboratory at the University of Cambridge

STM Shoot meristemless

WUS WUSCHEL

YFP Yellow fluorescent protein. Used for membrane staining.

Chapter 1

Introduction

1.1 The shoot apical meristem of *Arabidopsis thaliana*

Plant stem cells are organised by two developing centra – the Shoot Apical Meristem (SAM) and the Root Apical Meristem (RAM). The SAM is the region responsible for development of all aerial organs in the plant, and relies on a tightly orchestrated regulatory system in order to ensure a stable and functional developmental process. This includes aspects of cell proliferation and specification, as well as an ability of the plant to maintain and regulate the stem cell identity of the cells at the very apex of the shoot [4]. As opposed to the RAM, which has two stem cell pools in the inside of the root, the SAM maintains a single stem cell pool centered at the apex. It also lacks the root's cap, which protects the stem cells on the inside of the root, whereas these in the shoot are directly exposed to the plant's surroundings. Because of its highly exposed stem cell niche, the plant's regulatory system must therefore be organised in such a way that it can withstand significant perturbations, both due to molecular noise and changes in environmental conditions [12].

The stem cells at the SAM contribute to the development of new organs and tissue by dividing frequently at the top and from there mechanically being pushed out of the apex in order to differentiate, as illustrated in fig. 1.1. The steady maintenance of the stem cell niche allows for a constant production and supply of cells that the plant utilises during both growth and repair of damaged tissue, and is therefore completely essential for the plants survival [12].

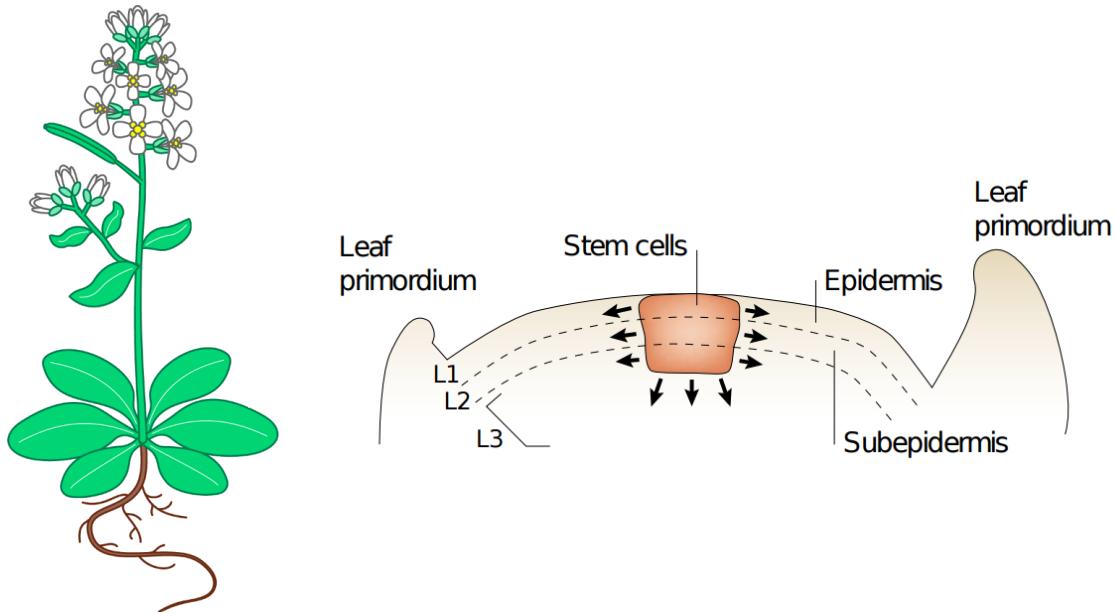


Fig. 1.1 Graphical illustration of model plant *Arabidopsis thaliana* and the shoot stem cell niche. The SAM consists of a dome-like domain where a small pool of stem cells are maintained throughout the life-span of the plant. Over the course of development, the cells are continuously pushed out of the stem cell niche out into the stem or periphery, where they differentiate. Structurally, the shoot is separated into three distinct layers that are of functional importance, here denoted L1, L2 and L3. Figure adapted from Clark [4].

In a simple outline of the SAM, its structure can be said to consist of four core regions: 1) The *central zone* (CZ), which harbors the aerial *stem cell niche* of the plant; 2) The *organising center* (OC), which is located beneath the CZ and acts as a signalling hub for many of the genes driving development in the SAM; 3) The *peripheral zone* (PZ), where cells form organs and new tissue through differentiation; and 4) the *rib meristem*, which make up the bulk of the SAM, and consist of the cells making up the stem. In addition to these regions, the SAM is also often separated into the different layers of the dermis, denoted L1 for the epidermal layer, L2 for the sub-epidermal layer, and L3 for the inner ground and vascular tissues. For cells in both L1 and L2, proliferation happens orthogonally to the shoot surface, i.e. so that cell lineages are preserved within the L1 or L2 correspondingly, with few exceptions. In contrast, this is not the case for the L3, where cells more freely can divide in all directions [28]. In addition, it has been suggested that the epidermis is involved in both promoting and restricting shoot development [10], adding to the notion of coordination and regulation between the different cell layers in order to accurately direct plant growth.

1.2 Modelling biological systems

From a theoretical point of view, organismal development can be considered in the framework of being a *complex system*. This is particularly useful in understanding the molecular pathways that ultimately end up determining cellular functionality and the general development of organisms. In a

systems biology setting, molecular and mechanical interactions are treated as abstract entities, each representing some fundamental part of the whole system in question, much like how machinery can be explained by its separate cogs and gears working together. Specifically in a molecular setting, the descriptive approach is commonly through *Gene Regulatory Networks* (GRNs), wherein each component represents some molecular aspect of the system that is involved in producing expression levels of mRNA, proteins and hormones [17].

Dynamics of GRNs are commonly understood both through analytical and computational means, where in the latter computer-generated models provide as a modern tool for better understanding the complex nature of many biological systems. Typically, reaction kinetics are modelled using various types of *Ordinary Differential Equations* (ODEs), although due to the vast amount of processing power available in the modern day, many recent studies also utilise traditionally more demanding resources such as *Stochastic Differential Equations* (SDEs). These also take into account the inherently random nature of molecular motions, interactions, and processes in order to capture dynamical features that deterministic versions cannot. For example, the ability of a cell to probabilistically change state depending on its environment, e.g. when committing to producing a certain protein or deciding to undergo apoptosis, can be the difference between life and death for the organism. Utilising the inherent noise of microscopic systems, cells have indeed been repeatedly shown to probabilistically tune their responses [23–25], showcasing the need of stochastic modelling for such types of phenomena.

In addition to stochasticity, the increase in computability has also allowed for the development of spatiotemporal modelling, where models are evaluated not only in a static context, but also in changing settings. This has allowed insights into how the distribution of gene expression varies over time, and how organisms orchestrate their developmental machinery such that organs and other types of tissue are developed at the right time [14].

Computer models in general have two separate aims: exploration and validation. In the former case, computer simulations can be the core for designing experimental experiments, where hypotheses from observed theoretical phenomena can be experimentally tested. In the latter, models are established in order to support hypotheses surrounding experimental findings, and can directly guide experiments through the identification possible mechanism behind an observed behaviour. The interplay between models and experiments, exploration and validation, causes the iterative process displayed in fig. 1.2, through which computational models and direct experiments work in conjunction to further common knowledge.

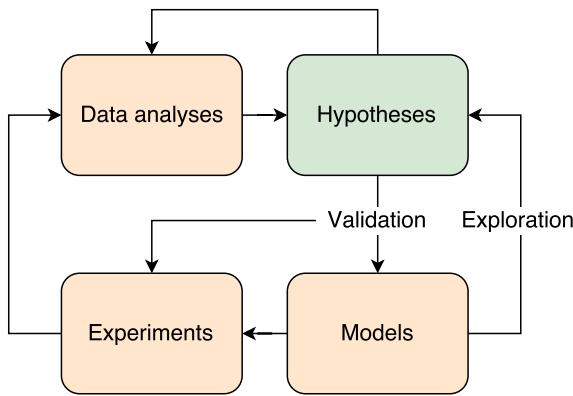


Fig. 1.2 Flowchart describing the work process in biological modelling. Investigations are driven by hypotheses, typically based on various types of data, or by the outcome of model simulations. Exploration or verification of these through either models or experiments in turn foster new hypotheses, creating an iterative process of continued research.

1.3 Regulatory Mechanics of Plant Stem Cells

1.3.1 Molecular tuning determines cell phenotype

The phenotype of a cell is largely determined by the underlying expressed genes and proteins, which in turn are regulated by the core GRN. Cells that have not yet undergone the differentiation process are those which are broadly described as stem cells. In addition to not having a specialised phenotype, stem cells continuously proliferate in order to give rise to new cells that can be used for development or repair [4]. Similarly to in animals, plant stem cells require an intricate network both specifying the multipotency, i.e. ability to assume different phenotypes, of the cell, and being able to maintain this both when the plant conformation or the environment changes. Effectively, this regulation causes the stem cell niches of the plant to be determined by various types of primarily molecular patterning, which also plays a role in phyllotaxis by specification of initiation zones for new primordia [27].

A viable and robust network maintaining patterning is thus important for the plant in order to undergo phyllotaxis in a functional manner. Molecules which rule this type of process are known as *morphogens*, and guide the initiation of organs and specialised cells by signalling processes, where cells are tuned to respond accordingly depending on their local configuration of molecular concentrations [21].

Morphogen patterns can consist of several types of spatiotemporal expression, including that of hormones, proteins and RNAs, although in extension to molecular interactions, also patterns of stress and strain have in recent studies been shown to play a role in determining both growth and cell identity [1, 11]. Typically, whenever gene expression is the focus of a study, it is often used as a proxy for protein expression, as fluorescent tagging and tracking of proteins sometimes interfere with the function or transport of the molecule.

1.3.2 Developmental Regulation in the Apical Meristem

The GRN in the SAM is determined mainly by two core genes – *WUSCHEL* (WUS) and *CLAVATA* (CLV). Their corresponding network consists of the homeodomain protein WUS and a ligand-receptor complex made up by CLV1 (receptor), CLV3 (ligand) and an assumed accessory protein CLV2. In particular the expression of CLV3, which is localised to few cells at the very apex CZ, correlates strongly with stem cell identity of the cells. In these cells, CLV3 encodes a small, secreted peptide which diffuses rapidly out of the cell and causes a gradient extending down to the OC where it interacts repressingly with WUS functionality [4]. Proper control of the CLV3 domain is essential for the development of the plant, with CLV3 induced plantlings terminate development prematurely [2].

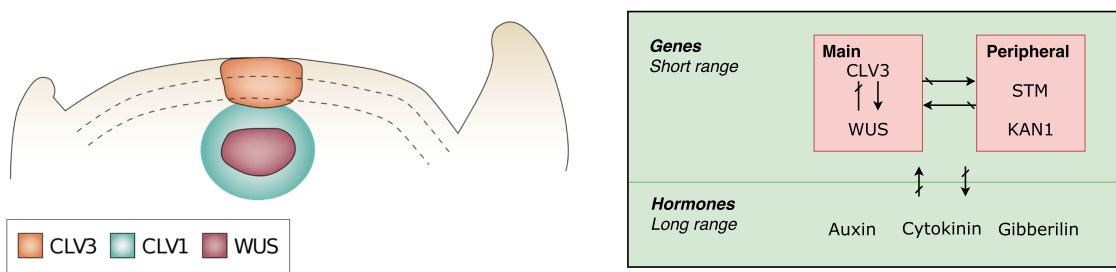


Fig. 1.3 Illustrative description of biochemical regulation in the shoot, where hormones are typically actively or passively transported to the shoot where they affect development. Genes instead perform short-range signalling, either through non-coding RNA or via their protein products. To the left is shown the regimes of a few key agents in the SAM. Stem cell identifier CLV3 is expressed at the very apex, with its corresponding receptor molecule CLV1 present in the OC. Together with WUS, CLV forms a dynamic feedback loop which is essential to maintaining a functional stem cell niche. Figure (left) partly adapted from Clark [4].

In contrast to CLV3, WUS agonistically activates the CLV pathway through diffusion of its homeobox protein. This activating interaction makes it necessary for maintaining an appropriate stem cell niche, and for repressing the differentiation process of the cells at the apex. This is particularly noticeable in WUS loss-of-function mutants, where the lack of the correct WUS gradient leads to defective shoots that terminate in aberrant flat structures [20].

Together, the CLV3-WUS feedback loop forms the core of the GRN regulating stem cell identity. Outside of the CZ, peripherally expressed genes such as KANADI (KAN1) are instead known to promote cell differentiation [16]. The core network itself is naturally also affected by the activity of other genes in extension, including hormonal intervention on WUS by the small and diffusive hormone cytokinin, which itself is activated by enzymes present in the meristem [13]. Also other homeobox encoding genes such as *Shoot Meristemless* (STM) are essential for correct development of the shoot by activation of CLV3 [29].

Aside of the system regulating the cellular identity, also other classes of molecules are key to the overall development. The plant hormone auxin in particular has been repeatedly shown to have an essential role in the coordination of growth, both in signalling initiation points of new primordia and elongation of the core stem. Because of this, auxin transport is key to asserting apical dominance in plants through the help of active transporters such as the PIN-FORMED (PIN) family, inhibition of this process leads to development of organless meristems. [19]. Also other hormones such as gibberellin are well-established signalling molecules important for SAM development [5].

On the whole, the molecular regulation of the development of the SAM consists of an intricate system that requires both tight regulation and precise coordination. This allows the plant to both counter and utilise noise that might be present due to volatile environments, or inherent molecular processes, so that functional and robust development can be ensured.

1.4 The role of plant stem cell modelling

An important question in developmental biology is how organisms can have robust development despite consisting of many independently variable parts [26, 3, 22]. At the same time, genotypically similar plants can nevertheless exhibit significant differences in phenotype, raising the questions of how, where and why noise impacts the development of the plant.

Understanding the true mechanics behind meristem development could have significant impact particularly on crop yield. Almost 80 % of the food supply of the modern world in some way derives from SAMs, in the form of various types of crops [7]. Elucidating the underlying functions and mechanics of the meristem is therefore of utmost relevance.

While plants, and AT in particular, have been studied thoroughly over the years, it is not until recently where significant advances in imaging has allowed for more fine-grained analyses both on 1) development of the plant *in vivo*, and 2) the extent and regulation of noise during development. The modern possibility of using 3D confocal microscopy to observe growth at the single cell level has in this spirit opened the door for quantified analyses on plant and cellular behaviour on the single cell level. Because of this, it is now possible to expand on this using timelapses of confocal images taken under a period to resolve not only the static image, but also the dynamic.

Prior research has set the ground for tracking cell lineages across multiple timepoints, although these have mainly focused on the mechanical aspects of division and morphogenesis [31]. In this thesis we instead use the approaches developed in previous studies and apply them for direct biochemical tracking of the stem cell niche on the single cell level – an endeavour which in itself is largely unprecedented.

Chapter 2

Methodology

2.1 Data consists of *in vivo* confocal timelapses

The majority of data used in this report is directly acquired from Willis et al. [31], and thus largely follows their approach for data acquisition and processing. An introductory description is nevertheless outlined below.

Six plants, labelled plant 1, 2, 4, 13, 15, and 18, were grown on a solution consisting of $10\mu M$ auxin transport inhibitor NPA to a depth of roughly 1 cm for 22-26 days. The inhibition of auxin prevents formation of new primordia, and this gives rise to a small and naked, organ-free meristem which is tractable for imaging.

The plantlets were marked with pUBQ10::acyl-YFP, which localises in the cell membrane, as well as with pCLV3::dsRED-N7, which was used as a nuclear tracker for CLV3 mRNA expression. Also pPin1::PIN1-GFP was tracked, but not quantified in this study. In addition, plant 1 did not express the nuclear marker for CLV3, whereas plant 18 did not undergo successful nuclear segmentation. Both are therefore completely excluded from the nuclear analysis.

Using confocal microscopy, the six plantlings were tracked in intervals of 4 hours up to 76 (plants 1, 2 and 4) or 84 hours (plants 13, 15 and 18), using a 63x/1.0 N.A. water immersion objective. Due to the high resolution of the images, the acquisition of each z-stack took \sim 10 minutes, which caused vertical stretching in the images due to stem elongation. Because of this, a second batch of z-stacks was acquired, using low-resolution imaging over \sim 10 seconds. The original images were then corrected, using this second batch as reference.

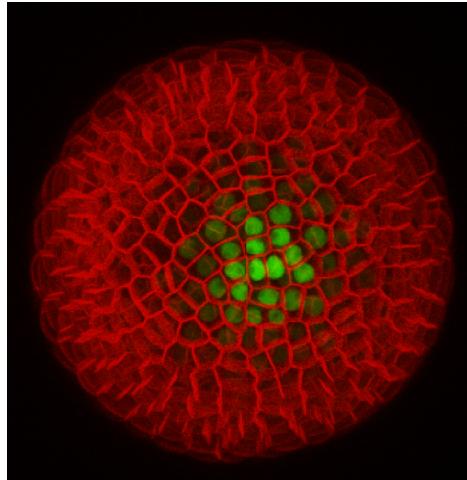


Fig. 2.1 Topview of raw data as produced by the confocal imaging, taken from plant 2 at timepoint 0 hours. In red can be seen the membrane channel, merged with the nuclear response in green. The nuclear vacuole is observable as shaded dots in the green channel. Note also how the L2 and L3 produce a perceived fuzziness in the peripheral regions.

2.2 Image pre-processing and segmentation

In order to eliminate segmentation errors, the ImageJ plugin StackReg was used to perform a translation transformation for each stack. Individual slices which contained horizontal shifts because of vibrations or other types of system disturbances were identified and replaced with the nearest slice that contained no such shift. The z-directional stretching due to stem elongation was corrected for by mapping the low-resolution stacks to the high-resolution ones in order to attain stretching factors that the images were thereafter corrected for.

For the membrane channel, noise removal was done by Gaussian and alternative-sequential filtering. The filtered z-stacks were then watershed in 3D using the algorithm implemented in the segmentation software MARS-ALT (see chapter C). Segmentation and tracking was thereafter performed using the same software. Cellular volumes were from this then calculated as the sum of voxel volumes belonging to the same cell. The tracking, also performed using MARS-ALT, was assessed for quality using an F1 score between the parent and corresponding daughter cell. For all analyses discussed in this report, a cutoff value of 0.30 was set for the tracking in order to account for incorrect mappings. These cells are included in the overall analysis, but excluded from all cell line related investigations. A longer outline of errors and exceptions in the segmentation is presented in section A.1.

The nuclear data were deconvolved to account for the microscope's point-spread function using the *PSF distiller* tool from Huygens software 15.05. As in the membrane case, the nuclear channels were adjusted with the corresponding stretching factors and thereafter segmented using segmentation tool Costanza [9] (see section C.1). Whenever we in this thesis mention the CLV3 expression, we refer to the mean fluorescence of the voxels belonging to the corresponding nucleus.

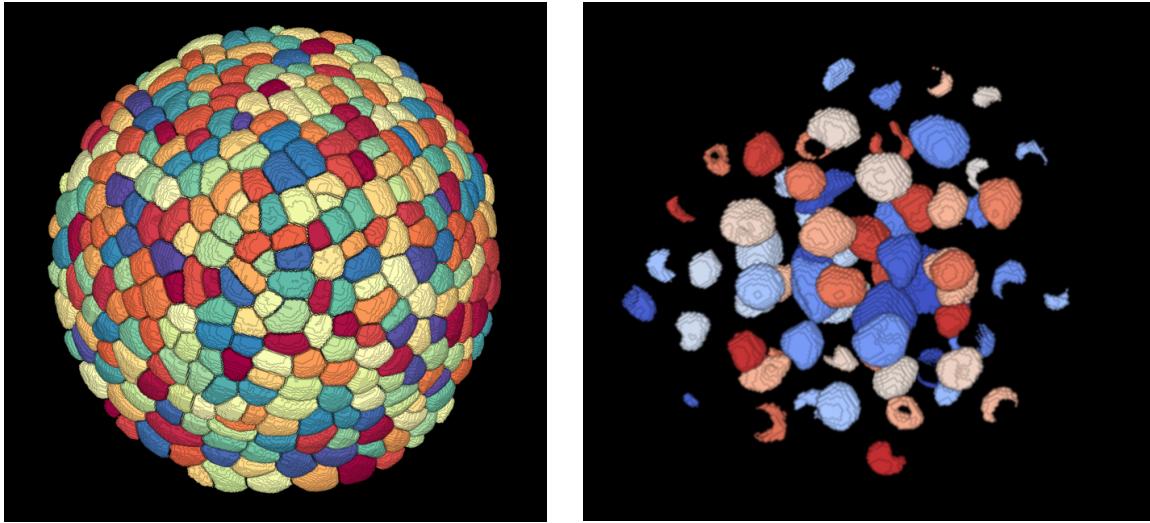


Fig. 2.2 Segmented membrane and nuclear channels as output by visualisation software TissueViewer after MARS-ALT and Costanza segmentation respectively (see chapter C). Note in the nuclear channel how the loss of signal due to the vacuole causes missegmentation in multiple cases. Figures taken from plant 18, timepoint 0 hours. Nuclear and membrane channels not to scale.

Due to the membrane and nuclear segmentations originating from different softwares, a separate mapping step was required in order to relate nuclei to their corresponding cell membranes. This was done by taking the centroid spatial coordinates, as defined by the basins of attraction acquired during segmentation, of all respective segmented items, and mapping them using a least-squares approach. Duplicately mapping nuclei were then consolidated as described in section 2.3.2.

Measures of distances to the top were done in multiple ways. The four methods herein considered consist of a definition of the top based on 1) the spatial coordinates, 2) the expression value, 3) expression-weighted spatial coordinates, and 4) a least-square fit of a paraboloid to raw meristem images. For the paraboloid fit, this is done by taking the top 75 images part of the confocal z-stack, in order to minimise the fitting impact of primordia that are more prominent in the deeper levels. In the case of spatial coordinates, the average $x - y$ coordinates of n nuclei were chosen, complemented with the highest z value registered in the corresponding timeframe. For the second case, the apex was defined as the mean spatial coordinates of the n highest expressing CLV3 nuclei. The weighted apices were similarly determined through transformation via $\bar{x}_w = \sum_i^n \bar{x}_i I_i / \sum_i I_i$, where n again denotes number of cells included, arranged by highest z -value, and I the CLV3 intensity for the corresponding cell. Lastly, the paraboloid fit to the meristem was used to define the apex by taking the coordinates of the region have a zero-valued derivative. For both the segmentation-dependent approaches, the data was set to exclude subepidermal layers in order to prevent biases.

To achieve a cell-resolution description of distances in the SAM, an auxilliary measure of cell distances was used in the form of a cell-wise grouping. In the cell value utilising definitions of the apex above, cells included in the definition was set to have a cell-wise distance of 0. The neighbours of these cells were in turn defined to have a distance value of 1, and so on recursively. Again, we

exclude the plants lacking nuclear data from measures utilising this. In addition, whenever cells deeper in the tissue than L2 are referenced, we refer to these as L3 and L3+ interchangeably.

2.3 Data processing

2.3.1 Development of a data analysis pipeline

Data was analysed predominantly using R through the development of a single modular pipeline, denoted *extractoR*. This was designed to build primarily on a sequential software design pattern, incorporating parallelisation and batch processing for treatment of the multiple timelapses. A more descriptive explanation is outlined in ??.

2.3.2 Data filtering

Due to thresholding effects for cell nuclei during the segmentation, some individual nuclei are occasionally identified as two or more. Nuclei were therefore mapped to the corresponding membranes using a minimum euclidian distance measure between the respective centroids. The nuclear quantified metrics were then corrected using the functions found in table 2.2. In addition to this, all mentions of numbers of nuclei are with respect to the number of cell membranes containing at least one nuclear volume identified within them.

For the data analysis section, data was excluded due to apparent segmentation errors. This was done for each plant in isolation, with the outline of the filtering described in table 2.1. The choice of allowed deviance was done based on the distribution shape, (TODO: Put these in appendix) with particular consideration taken to the nuclear and membrane volumes, where no lower boundary was set. The maximal neighbour distance was chosen due to the typical lack of data for cells more than 7 cell distances from the apex. Lastly, due to division events where loss of nuclear signal took place, we filter out expression values which are less than 70 % of the magnitude in the previous, as well as subsequent timepoint.

2.4 Models of Gene Regulatory Networks

2.4.1 Mathematical Formulation of Biochemical Reactions

Mass-action Kinetics

The formulation of processes in GRNs focus primarily on two aspects: synthesis and degradation of matter, which usually takes the form of molecular concentrations or absolute abundance. As outlined in section 1.2, we here work using an ODE or SDE description of our regulatory systems.

In this study, we represent our molecular reactions using mass-action interactions and Michaelis-Menten kinetics; both here relying on the naïve assumption that our primary reagents here act in

Parameter	Value
Maximal membrane volume	$\mu + 3\sigma$
Minimal membrane volume	0
Maximal nuclear volume	$\mu + 5\sigma$
Minimal nuclear volume	0
Maximal apical distance	$\mu + 3\sigma$
Maximal neighbour distance	7

Table 2.1 Filtering settings in order to account for outliers in the data. Typical causes of these are mis-segmentation such as the merging of multiple nuclear membranes of nuclei. Because the quality declines with the distance to the apex we here limit our analysis to cases where cells are within a distance of 7 cells from the defined apex.

Metric	Summary function
Coordinates (x, y, z)	mean
Nuclear volume	sum
Nuclear expression	mean

Table 2.2 Consolidation methods applied for cases where multiple nuclei were mapped as belonging to the same membrane cell. This happens mainly due to errors when one nuclei is identified as multiple in the binarisation step of the segmentation.

isolation of other possibly intervening molecules. As part of our formalism, we write



to express that some substrate S is turned into a product P by some given *forward affinity* k_f . Likewise, as the reaction is *reversible*, the product P is transformed back into S with the *backward affinity* k_b .

The *law of mass action* states that the rate of a reaction is proportional to its affinity, e.g. k_f , and the concentration of the reacting species, here S . The reaction rate of the production of P would thus be $r_f = k_f S$. However, the rate of change of reactant P also depends on the backward affinity, which would give the overall rate-of-change for P as

$$\Delta P = k_f S - k_b P. \quad (2.2)$$

In the infinitesimal limit, we analogously have

$$\frac{dP}{dt} = k_f S - k_b P, \quad (2.3)$$

i.e. on the form of a differential equation, which will be the baseline for our formulations. Similar to the formulation of a rate-of-change of P , we can do the same for species S , and our system is then fully represented as a system of differential equations.

Expanding on this, we can easily solve for the steady-state concentrations of the system by assuming that all rates average to zero. In our example above, this gives us

$$\frac{k_f}{k_b} = \frac{P}{S}, \quad (2.4)$$

which holds in general, regardless of the number of reacting species.

Michaelis-Menten Kinetics

A conceptual drawback of the mass-kinetics formulation is the possibility to have infinite reaction rates, whereas the molecular reactions in nature typically are restricted by some means. One way to account for this is through *Michaelis-Menten kinetics*, which describes enzymatic chemical reactions. In these, the trivial example introduced in eq. (2.1) is expanded to include an enzymatic agent such that



In other words, an enzyme-like molecule binds to the substrate S such that the complex ES is formed. This complex is thereafter transformed into the product molecule P and again the enzyme E . Assuming a total enzyme concentration of E_{tot} and the assumption that the enzyme-substrate binding process is in equilibrium, the rate-of-change of the product can be rephrased to be on the form

$$\frac{dP}{dt} = V_{max} \frac{S}{K + S}, \quad (2.6)$$

where $K = K_f/K_b$ and $V_{max} = k_p E_{tot}$. This expression is said to be on *Michaelis-Menten* form, where V_{max} is the maximal activation rate of the protein, and K can be thought of as a saturation coefficient.

Extrapolating on this type of reaction, introducing n enzymatically acting molecules instead gives the standard *Hill equation* form, namely

$$\frac{dP}{dt} = V_{max} \frac{S^n}{K^n + S^n}, \quad \text{and} \quad (2.7)$$

$$\frac{dP}{dt} = V_{max} \frac{K^n}{K^n + S^n} \quad (2.8)$$

for an activating and repressing reaction respectively.

2.4.2 Numerically solving stochastic systems

Gillespie Algorithm

The Gillespie algorithm is a discrete approach for simulating stochastic molecular dynamics. It first appeared in print by Dan Gillespie in 1977, and has since been widely used for stochastic simulations in multiple fields.

While being computationally expensive, the Gillespie algorithm compensates for its lack in tractability by producing a statistically exact trace of the molecular dynamics of a system.

The algorithm originates in the formulation of the *chemical master equation*, which specifies the

rate of change of the transition probability between states in the form of

$$\frac{\partial P(x, t|x_0, t_0)}{\partial t} = \sum_{j=1}^M [a_j(x - v_j) P(x - v_j, t|x_0, t_0) - a_j(x) P(x, t|x_0, t_0)] \quad (2.9)$$

where a defines the reaction probability, or propensity, for each type of reaction, and v the stoichiometry, i.e. information of how the molecular species are changed due to the reaction. $P(x, t|x_0, t_0)$ on its own denotes the probability of $X(t) = x$, given that the initial value is x_0 . Solving the master equation analytically is usually complicated, so simulating a complex biological system using the Gillespie approach can often be far more tractable.

Procedurally, the algorithm can be formulated in four steps:

Initialisation Generation of number of molecules and reaction parameters.

Randomisation Generation of random numbers to determine 1) next interaction, and 2) the time increment.

System update Time and molecular numbers are updated correspondingly to the determined event in step 2.

Repetition Step 2-4 are repeated until some stop condition is met.

In principle, the Gillespie algorithm is interested in two fundamental questions: 1) When does the next reaction happen? 2) Which is the next reaction? The time until the next reaction at time t is denoted τ and can be shown to be an exponential distribution centered at $1/\sum_{j=1}^M a_j(x)$ for some molecular concentration x , i.e.

$$p(\tau = t') = \sum_{j=1}^M a_j(x) e^{-\sum_{j=1}^M a_j(x)t'} \quad (2.10)$$

with the reaction probability instead being described by the normalised propensity. Historically, due to the limitation of random number generators, the time update has been described as being drawn from

$$\tau = \frac{1}{\sum_{j=1}^M a_j(x)} \ln \frac{1}{r_1} \quad (2.11)$$

with r_1 denoting a uniformly distribution random number in the interval $(0, 1]$.

Milstein's Method

If there is no requirement for exactness, less computationally intense alternatives to Gillespie's algorithm exists. One such example is the *Langevin* formulation of chemical systems, which utilises SDEs to attain an approximate solution to the system trajectory, and is particularly useful when the number of molecular reagents is high.

The Langevin formulation, like Gillespie's, utilises the chemical master equation to compute the behaviour of the system. In principle, the Langevin formulation can be said to reformulate a deterministic increment of the form

$$X_i(t + dt) = X_i(t) + \sum_{j=1}^M v_{ji} a_j(X(t)) dt \quad (2.12)$$

to the stochastic form

$$X_i(t + dt) = X_i(t) + \sum_{j=1}^M v_{ji} a_j(X(t)) dt + \sum_{j=1}^M v_{ji} a_j^{1/2} N_j(t) dt^{1/2} \quad (2.13)$$

where X denotes the molecular number, v the stoichiometric coefficient of the equation in question, and N_j are temporally uncorrelated and statistically independent, Gaussian random numbers with mean 0. From this stage, the equation is then easily extended to its multivariate form, namely

$$X_i(t + dt) = \sum_{j=1}^M v_{ji} a_j(\bar{x}) dt + \sum_{j=1}^M v_{ji} a_j^{1/2}(\bar{x}) N_j(t) (dt)^{1/2} \quad (2.14)$$

Milstein's approach to solving this equation numerically utilises eq. (2.13) on the differential form

$$dX_t = a(X_t) + b(X_t) dW_t \quad (2.15)$$

where W_t is a continuous-time stochastic process. The simulation interval $[t_0, T]$ is then partitioned into parts of size $\Delta t = T/N$, where N is the number of partitions. We thereafter define the update

$$Y_{n+1} = Y_n + a(Y_n) \Delta t + b(Y_n) \Delta W_n + \frac{1}{2} b(Y_n) b'(Y_n) \left((\Delta W_n)^2 - \Delta t \right) \quad (2.16)$$

$$Y_{n+1} = Y_n + a(Y_n) \Delta t + b(Y_n) \Delta W_n + \frac{1}{2} b(Y_n) b'(Y_n) (\Delta W_n)^2 \quad (2.17)$$

on Itō and Stratonovich form respectively. Here b' denotes the spatial derivative of b , whereas $\Delta W_n = W_{\tau_{n+1}} - W_{\tau_n}$. The difference between the Itō and Stratonovich form in turn is the interpretation of the integral of dW_t [18]. In many cases, it is preferable to express the numerical update on a derivative-free form, which can be done through a Runge-Kutta like approach [8] in order to state the final, multivariate expression as

$$Y_{i,n+1} = Y_{i,n} + a_i(Y_{i,n}) \Delta t + b_{ii}(Y_{i,n}) \sqrt{\Delta t} N_i + \frac{1}{2\sqrt{\Delta t}} [b_{ii}(\bar{x}, n) - b_{ii}] \Delta t (N_i)^2 \quad (2.18)$$

where a supporting predictory step is calculated in the form of

$$\bar{x}_i = x_i + a_i(Y_n) \Delta t + b_{ii} \sqrt{\Delta t}. \quad (2.19)$$

Algorithmically, the Milstein approach is of strong order of convergence $\mathcal{O}(\sqrt{\Delta t})$ and weak order $\mathcal{O}(\Delta t)$ [8]. In this thesis, we utilise the Milstein approach under the Stratonovich interpretation.

2.5 Network modelling approach

In this study, we model regulation of gene expression using hill equations for production and exponential decay for degradation. Gene products are instead set to undergo linear production with respect to the activating gene, as well as diffusion. Like for gene expression, proteins are modelled using exponential decay. In total, the equations governing the regulation for a gene X and its protein x can then be formulated as

$$\frac{dX}{dt} = \prod_{a=1} V_{max} \frac{X_a^{n_a}}{K_a^{n_a} + X_a^{n_a}} \prod_{r=1} \frac{K_r^{n_r}}{K_r^{n_r} + X_r^{n_r}} - dX, \quad \text{and} \quad (2.20)$$

$$\frac{dx}{dt} = pX + D\Delta x - dx, \quad (2.21)$$

where Δ is the Laplace operator. In a deterministic framework, the steady state of the system can then be solved algebraically if the parameters of the system are known. In this thesis, whenever we simulate our stochastic networks, we first compute the steady state of the system using this approach.

Chapter 3

Results

3.1 Regulation of the CLV3 domain induces periodicity when perturbed

When tracking the number of CLV3 nuclei identified by Costanza, the results observable in fig. 3.1 shows the number of observed CLV3 nuclei increasing for plants 2, 4, 13, and 15, corresponding to a visually observable enlargement of the CLV3 domain analogous to the one seen in fig. 3.2. The fluctuations do not significantly correlate with the mean intensity level except for plant 15 ($p = 0.99, 0.36, 0.76, 0.01$), and thus generally correspond to a pulsating enlargement of the domain as opposed to fluctuations in the CLV3 intensity.

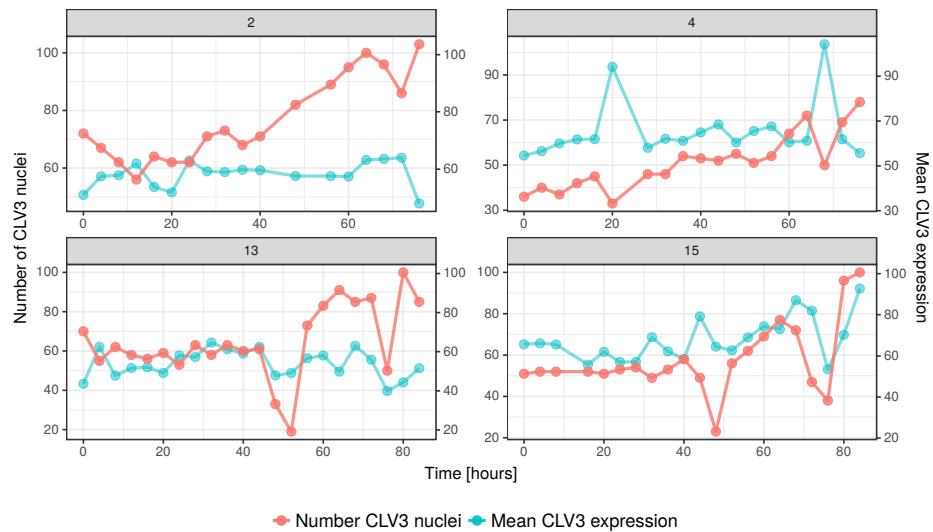


Fig. 3.1 Development of number of CLV3 nuclei over time, as mapped next to the mean nuclear intensity. All plants exhibit an increase in the number of nuclei over time, and appear to dip in a systematic manner. The mean intensity occasionally shifts in accordance with the number of CLV3 nuclei, but in a non-systematic manner.

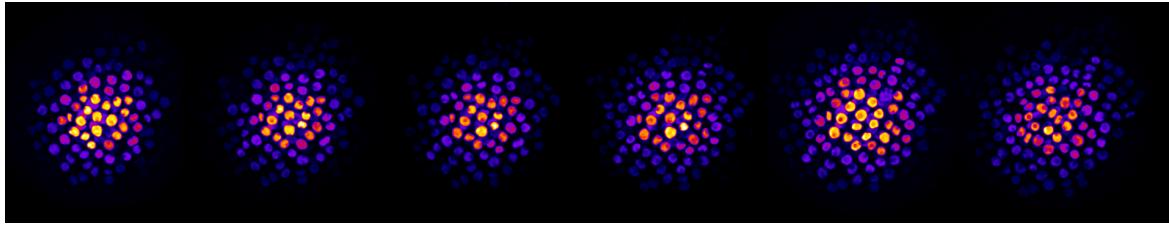


Fig. 3.2 Visually observable fluctuations in the CLV3 domain. Images are taken from plant 4 in a linear fashion between 56 and 76 hours.

In order to assess the extent of the fluctuations the lines were detrended using a second order Loess fit to each curve, producing the curves found in fig. B.3. When thereafter performing a continuous time Fourier transform to extract amplified modes, we see that the plants are biased towards the fourth mode, i.e. a period of ~ 16 hours, in each respective transformation, as depicted in fig. 3.3.

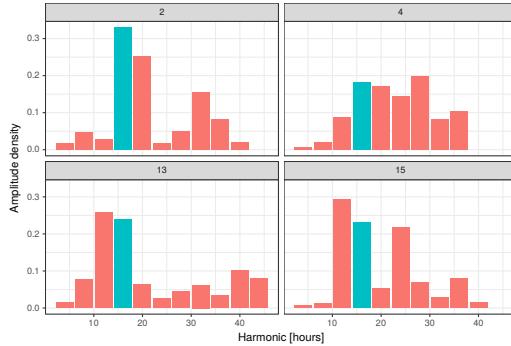


Fig. 3.3 Plant-wise periodogram for detrended number of nuclei trajectories. All plants show an emphasised fourth harmonic (blue), denoting periodicity with a 16-hour period.

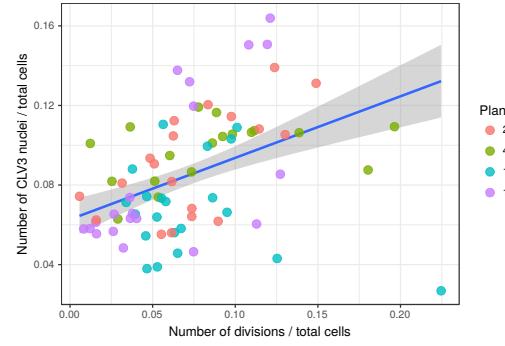


Fig. 3.4 The number of division events correlates linearly with the number of CLV3 nuclei, signifying a functional variation due to the number of CLV3 nuclei.

In addition to the apparent periodicity, the number of division events correlates strongly with the number of active CLV3 nuclei when normalised by the total number of cells observed, showing that the number of division scales with the size of the CLV3 domain.

3.2 CLV3 expression is stable for apical cells

Variance in the expression of CLV3 is suppressed in the apical cells, as can be seen in fig. 3.5. Generally, the CLV3 distribution assumes a shape with low variance close to the apex, larger variance at intermediate distances, and again low variance at larger distances. This appears true for all plants, although plant 13 in particular shows tendencies of having a few lowly expressing apical cells at several timepoints. It should however be noted that it is possible that errors in the tracking produces this type of data.

Closer investigations in the distribution of expression values at distinct distances shows non-bimodal tendencies. Instead, the data suggests an apparent CLV3 gradient, proportional to the distance to the apex more in radial terms than in bimodal ones.

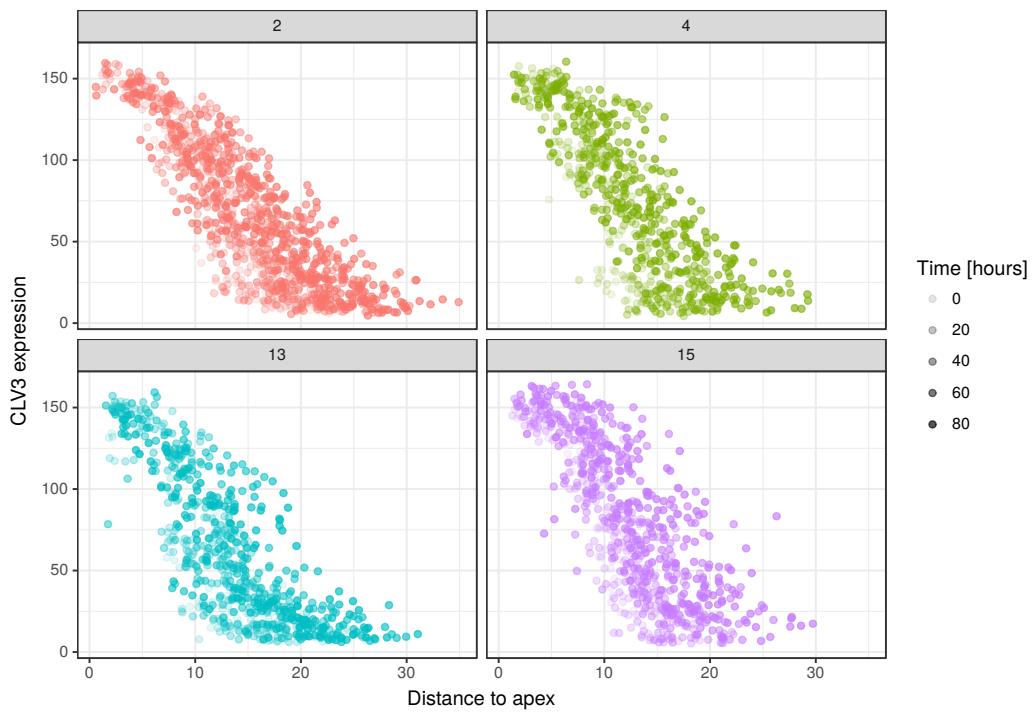


Fig. 3.5 CLV3 expression distributions for the L1, with the apex of the meristem defined as the mean coordinates of the 4 topmost expressing cells at each given timepoint. The distributions all assume a hysteresis-like appearance, with minor variance at the apex, a highly variable body, and a yet again lowly variable tail.

Using a simple model mimicking the epidermis, we are possible to replicate the shape of the distribution assuming enzymatic CLV3 activation by a WUS gradient (fig. 3.6). The enzymatic activation is in this formulation necessary in order to attain the sigmoidal shape of the curve.

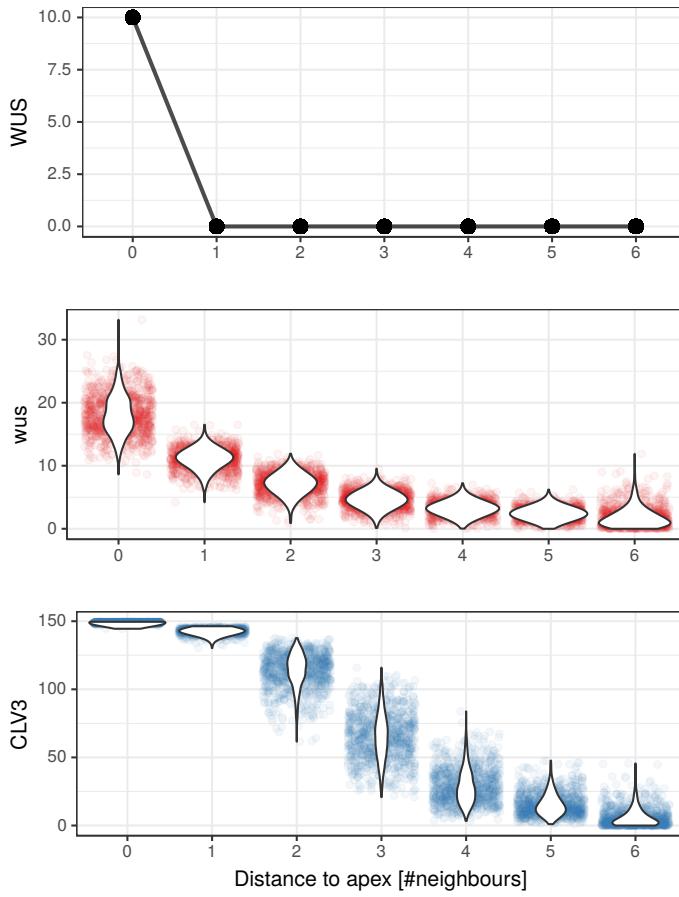


Fig. 3.6 Epidermis model recreating the results observed in fig. 3.5. Here, a gradient of WUS induces sufficient activating mechanics for producing the correct CLV3 gradient.

Aside of the CLV3 distribution in the L1, we observe a layer-wise separation of the CLV3 expression in the comparison between L1 and L2, as visualised in fig. 3.7. In L1, the correlation between CLV3 expression and nuclear volume appears sigmoidal, although consideration should be taken to the fluorescence ceiling due to the laser. In contrast, both L2 and L3 have linear relationships, indicating the possibility of epidermal activation alternatively subepidermal repression altering the distributions. Also in the distributions of CLV3 expression in relation to the distance to the apex this effect is observable.

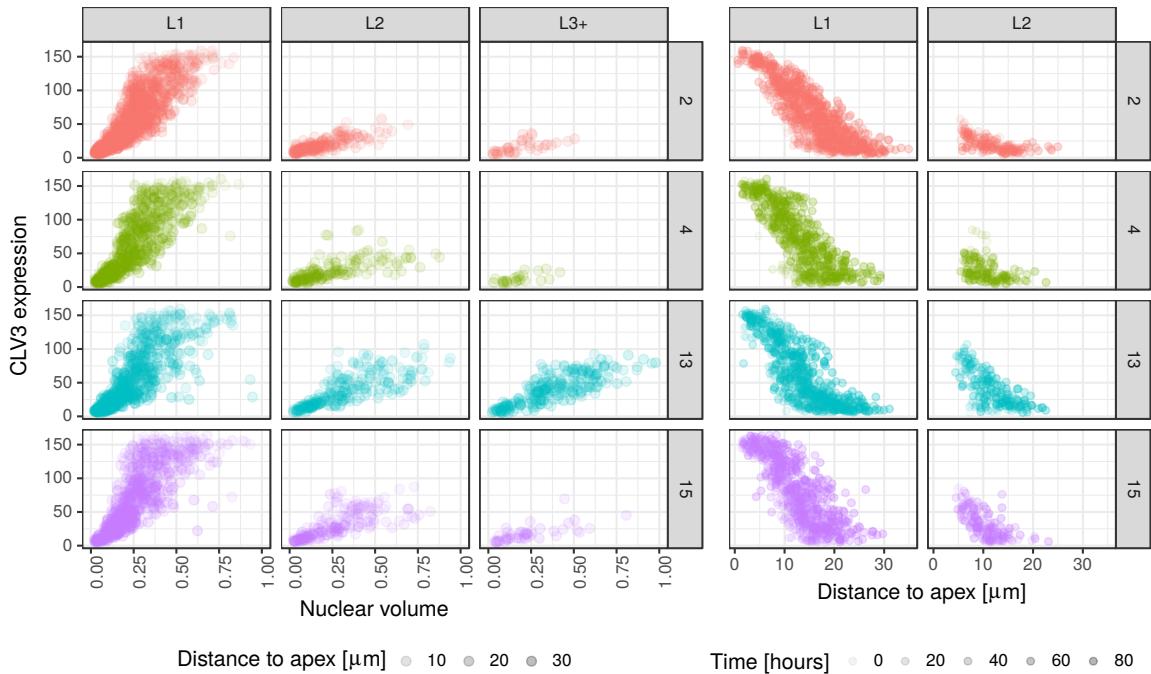


Fig. 3.7 Distribution data suggest differing regulation between L1 and the subepidermal layers. In penetrating deeper into the tissue, the CLV3 expression relative the nuclear volume changes from a sigmoidal to linear relationship (left). Similarly, the CLV3 expression relative the distance to the apex appears to have similar shape in the L2 as in the L1, but shifted.

3.3 Functional clustering implies dsRED technical artefacts

The apical cells appear to be clustered functionally, suggesting that a significant portion of the CLV3 expression observed outside of these topmost cells may in fact be reporter remnants, stemming from the decay time of the dsRED reporter.

Figure 3.8 shows a tendency of the apical cells to assume a differing average membrane volume between cells at the apex (distance 0), and the ones neighbouring ($p = 5e - 12, 1e - 2, 7e - 4, 7e - 5$). This contrasts with the observed nuclear volume, where instead an overall decreasing trend in the response to CLV3 appears, as fig. B.2 depicts.)

In addition, the topmost cells undergo significantly fewer divisions per cell in that region as opposed to the other cells in the L1 (fig. 3.9). Most division events instead appear to be focused in the first to third cell layer away from the apex, and thereafter decline. Under the assumption that the number of cells scale with the area such that $N \propto r^2$, we should expect to see the number of division events increase linearly with r ; here we instead see the linear tendency only for cell distances 1-3, where the topmost cells behave differently. Generally, the observations can be observed in three points:

1. Topmost cells divide less frequently than cells in layer 1-3

2. Cells in layers 1-3 divide at a similar frequency per cell

3. At cell distances > 3 , the number of divisions per cell decreases steadily

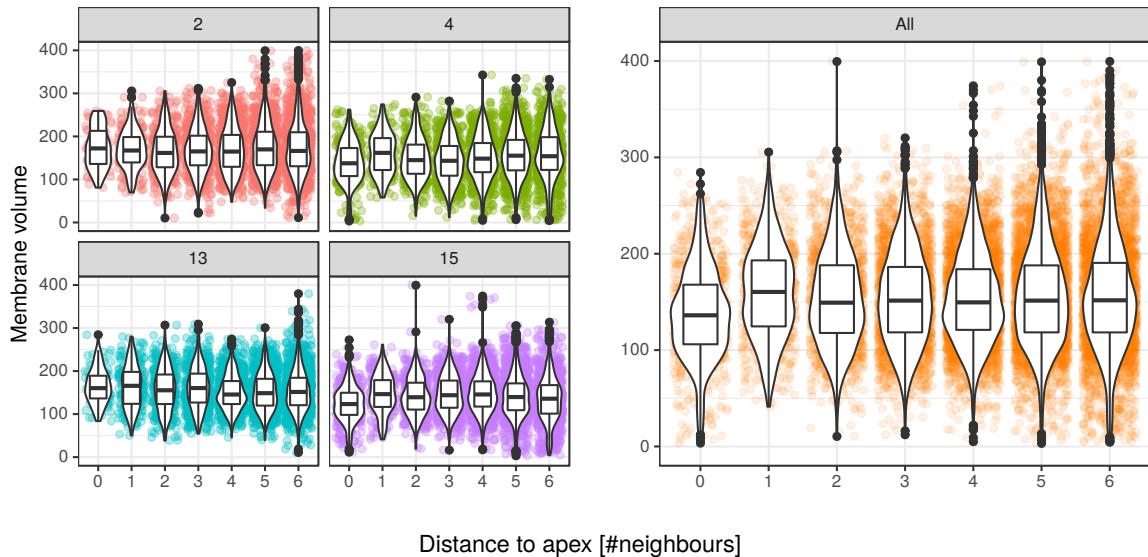


Fig. 3.8 Membrane volumes in the L1 clustered at different distances from the apex in number of cells. Cells as part of the apex here are defined as the 4 topmost expressing cells; their neighbours in turn correspond to the cells at distance 1, and so on. Here all examples but plant 2 have a statistically significant separation of average membrane sizes between bin 0 and bin 1.

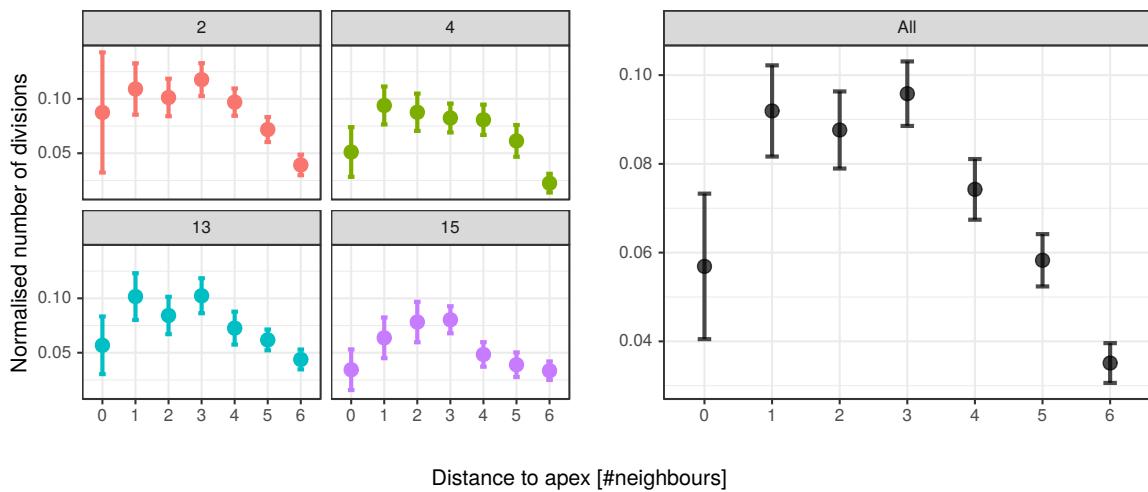


Fig. 3.9 Division analogue to fig. 3.8, with the average number of divisions normalised by the number of cells at that corresponding distance, along with the standard error of the mean. As in the membrane volume case, plant 2 shows a different trend than the other plants and in particular exhibits a larger variance, showing noise in the definition of the topmost cells.

Separating out the lineages which are present throughout the full time course reveals an apparent contained maintenance of ca four cells where decline in the CLV3 expression can not be noted

(fig. 3.10; lineages 7, 16, 17). Overall, cells appear to be in either one of two state with respect to the rate of change of their CLV3 expression, in a few select trajectories, where further analysis indicates that cells assume a declining state when pushed out of the very apex. Cells that in contrast remain in this region maintain their high expression profiles, signifying the possibility of only the very apical cells effectively expressing CLV3.

It is unclear whether the resolution of the trajectories is significant enough to observe the exponential decay in dsRED, or whether a linear approximation is sufficient. Figure 3.11 shows these two alternatives where the interpretation of the half-life differs with a factor of two depending on the interpretation, where the exponential one corresponds to a half-life of roughly 41.7 hours, and the linear correspondingly 83.3 hours, assuming a Taylor expansion to the first order.

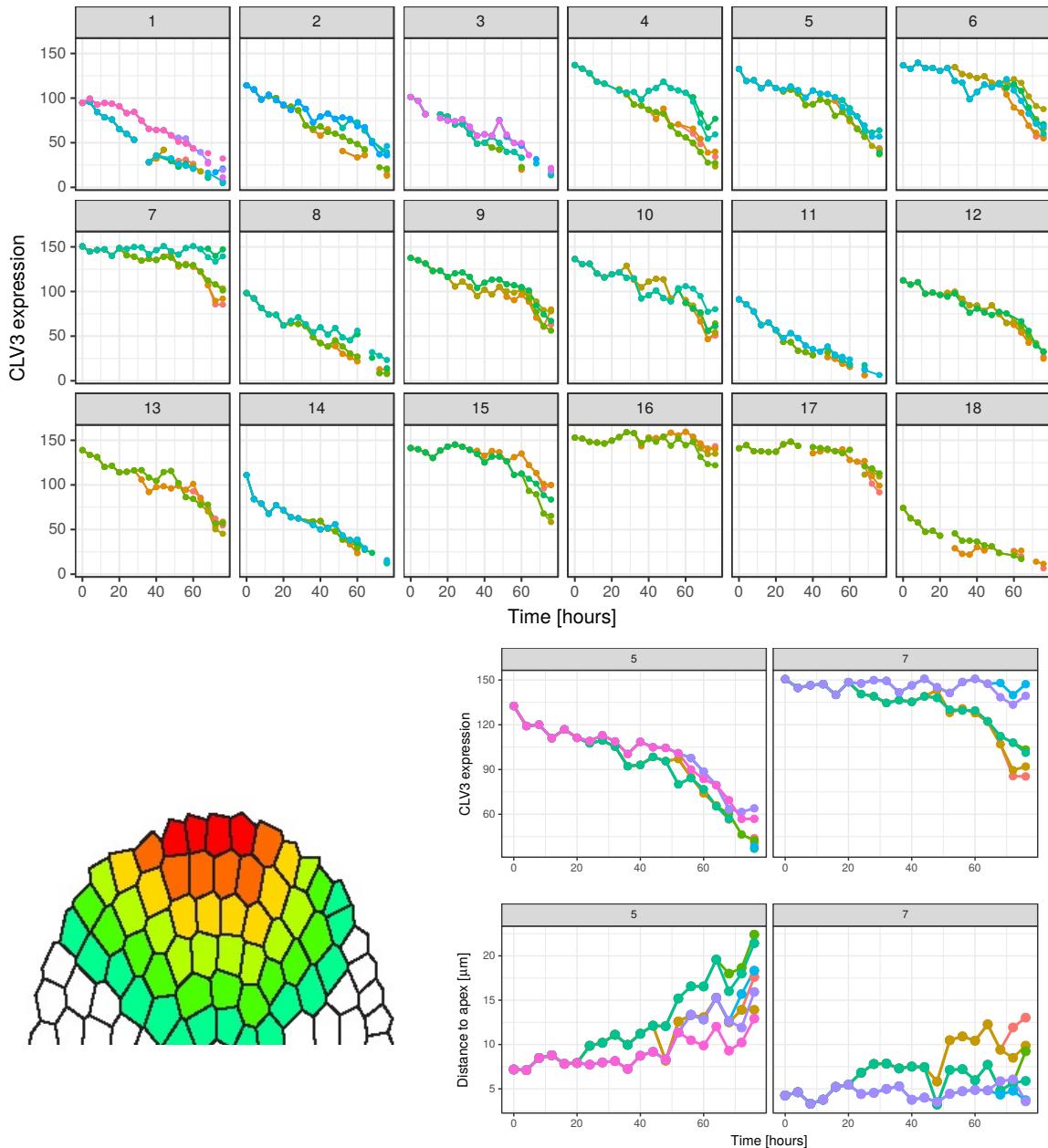


Fig. 3.10 Cell line trajectories for lineages in plant 2 with cells present at both the initial and final timepoint in the series. Cells generally have a declining or maintained behaviour corresponding to the two cases (cell line 5 and 7) emphasised in the bottom right. This corresponds in to the behaviour illustrated in the bottom left, with cell lineages at the very top having preserved expression levels, and with decreasing signal outside outside of this region. Lacking timepoints in trajectories in most cases correspond to data filtered out due to division events (see section 2.3.2).

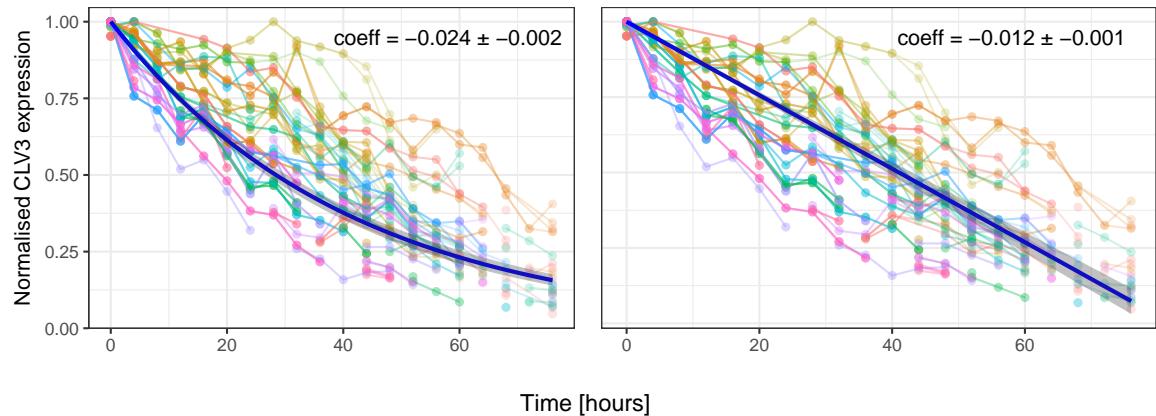


Fig. 3.11 Exponential and linear approximation of dsRED decay respectively. Data shows plant 2, with all sublines with at least 12 data points, downshifted to timepoint 0 hours for comparison, and divided by the timeseries maximum.

Fitted lines are taken as the mean coefficients of singular fits to each individual line, presented with the 95 % confidence interval. Also plants 4, 13 and 15 produce results within the margin of errors for both approximations.

Assuming a situation where dsRED is not produced, we can hypothesise of the formulation of at what rate CLV3 decays with the distance to the apex. Given that cells are on average pushed out of the CZ at a rate $\dot{R} = k_R$, the rate of change for the CLV3 expression can be written on the form

$$\begin{aligned}\dot{C} &= \frac{dC}{dt} = \frac{dC}{dR} \frac{dR}{dt} \Rightarrow \\ \frac{dC}{dR} &= \frac{1}{k_R} \frac{dC}{dt}\end{aligned}$$

where R denotes the radial distance. Under linear CLV3 decay, e.g. $\dot{C} = -k_C$, CLV3 should decay linearly with the radius as $\frac{dC}{dR} = \frac{k_C}{k_R}$, which appears possible given the distribution in fig. 3.5.

3.4 Long-lived cells are amassed centrally in the deep tissue

A clear cluster of cells are clustered at an age corresponding to roughly $3\times$ the normal cell cycle in the deeper layers. Notably, this effect is not accompanied by a similar cluster at $2\times$ the normal cell length, which would be expected in the event of missed divisions in the tracking. The effect is the most noticeable in plants 2, 4, and 15, where all of these show an amassment of division events at the 60 hour mark, most protruding in the L3.

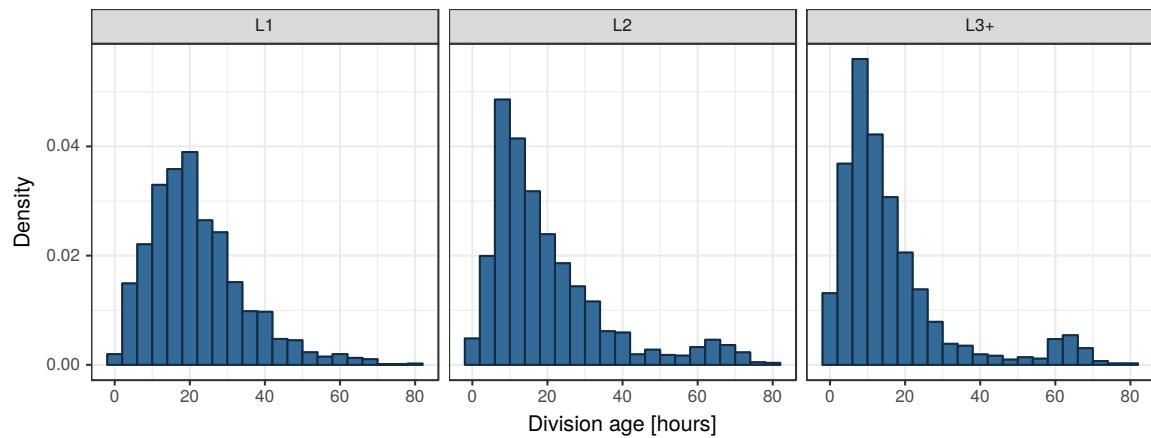


Fig. 3.12 Distribution of ages before division. The subepidermal layers have an amassment of cells dividing around the 60 hour mark, i.e. at ca $3\times$ the normal cell cycle.

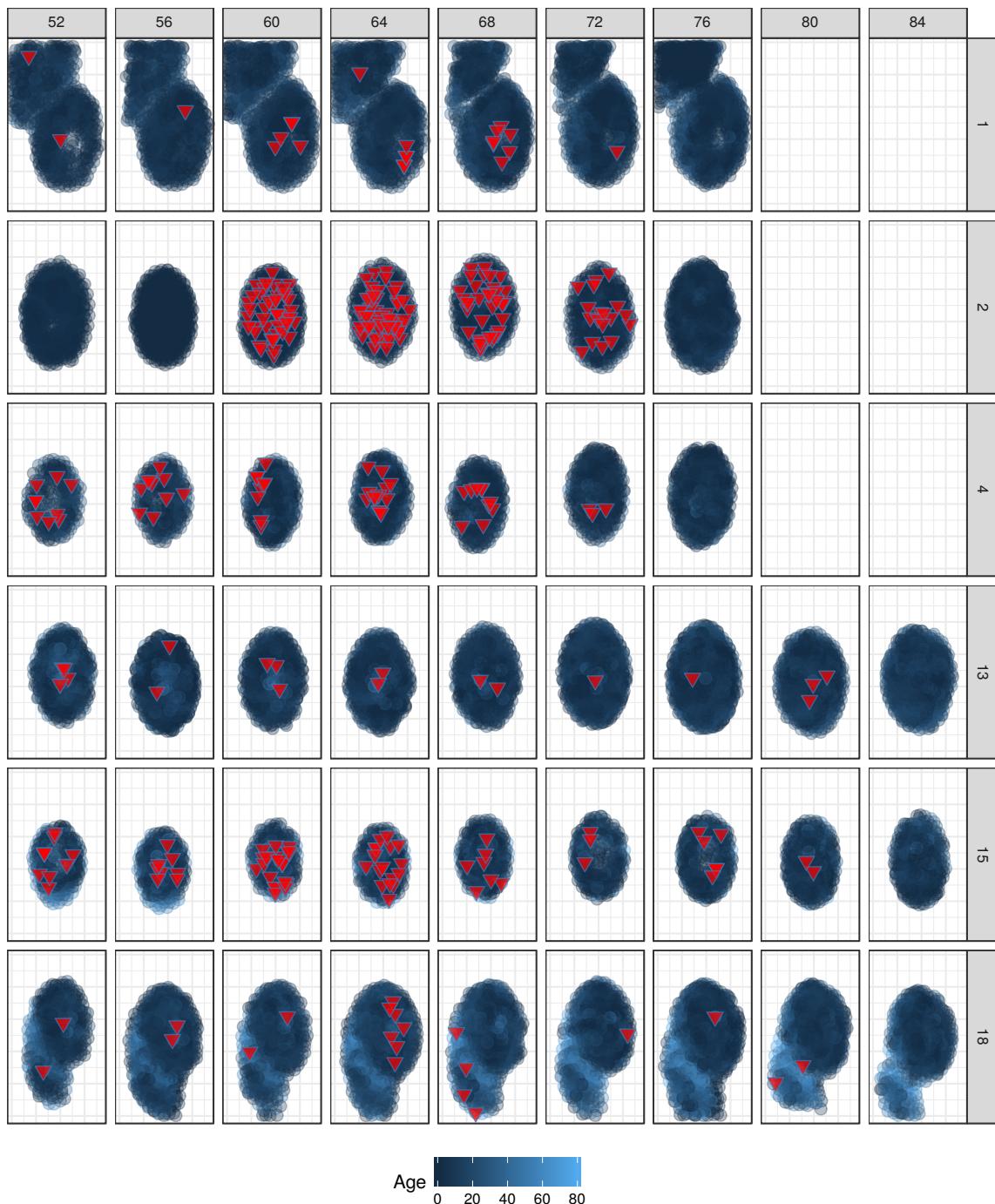


Fig. 3.13 Cell coordinates grouped by plant and time for cells not in L1, coloured by age, with division events for cells with age > 48 denoted as red triangles. There appears to be clustering in the center of the meristem in several cases. Note also how primordia are appearing in plants 1 and 18, and how cells in the boundary between primordia tend to be affiliated with higher ages.

3.5 The CLV3 apex does not coincide with the geometric

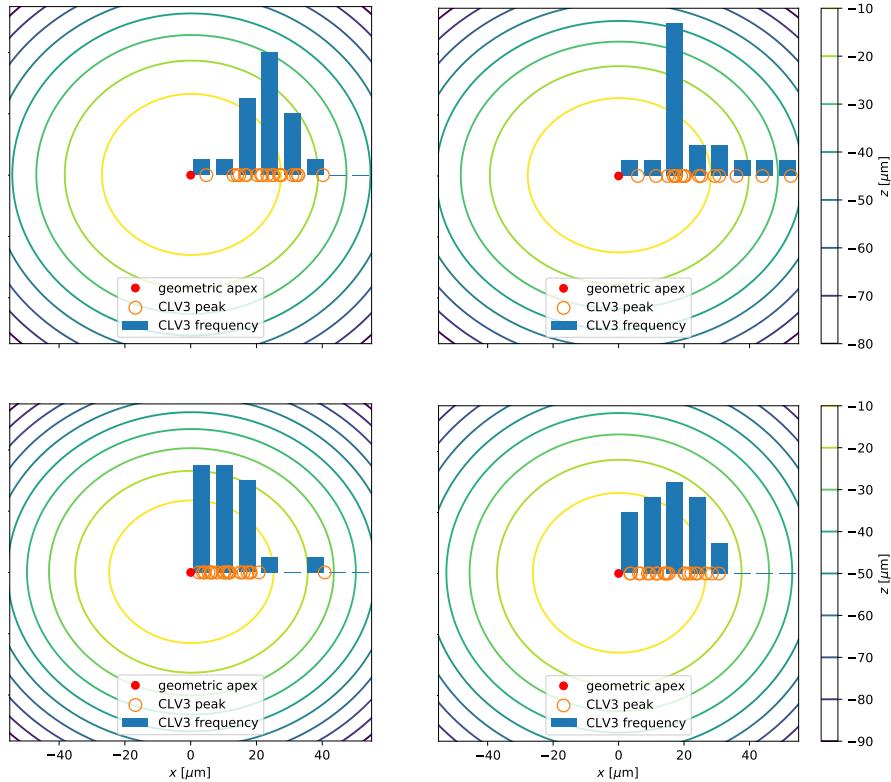


Fig. 3.14 Distributions of radial distances from the geometric apex, as fitted with a paraboloid shape, for plants 2, 4, 13 and 15. CLV3 apices are defined as the mean coordinates of the four highest expressing CLV3 cells. Notably, all plants have distributions peaking at a distance of roughly 10-20 μm .

Notably, the CLV3 peak and the geometric apex does typically not coincide, with the CLV3-based apex typically at a distance of 10 – 20 μm , corresponding to 2-4 cells in distance. While the quality of the fit can be affected by the presence of primordia, the evolution of paraboloid parameters over time is relatively constant, as can be see in fig. 3.15.

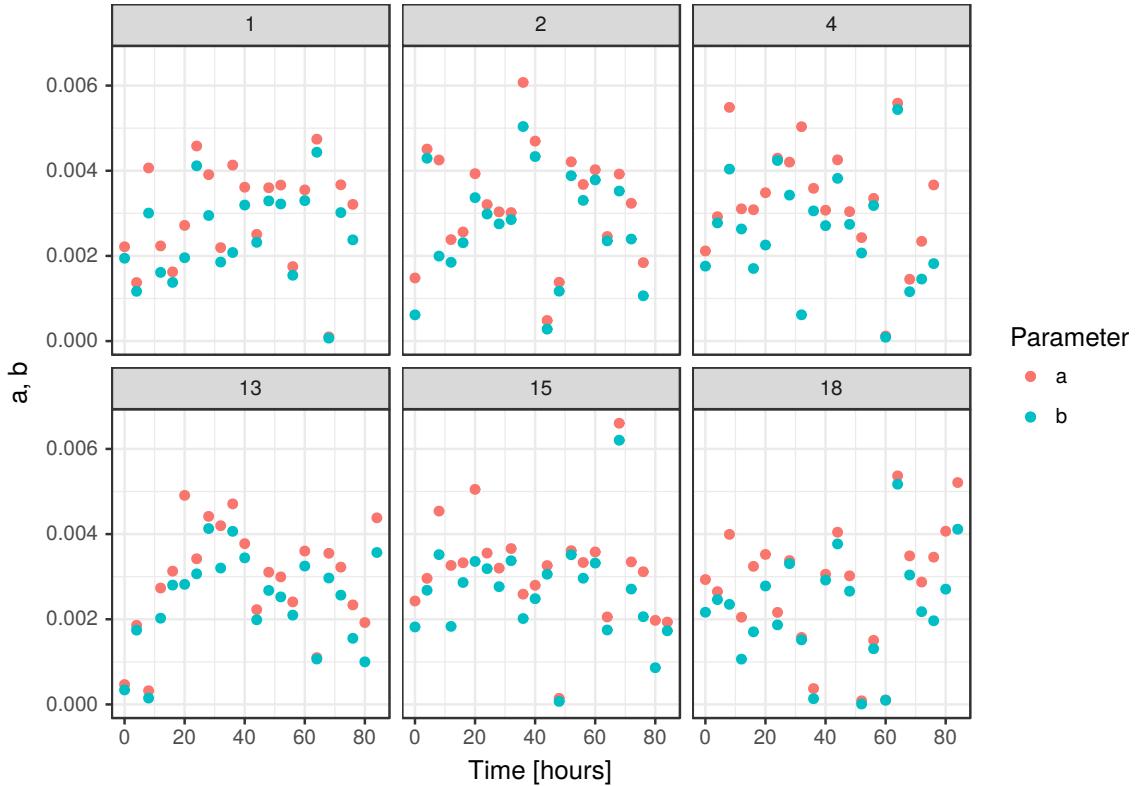


Fig. 3.15 Evolution of paraboloid parameters, where the paraboloid has been fit according to the equation $ax^2 + by^2 = z$. We have here defined b as determining the axis of largest curvature.

For emphasis, we have included an example of the evolution of the largest axis of curvature for plant 15 (fig. 3.16). In general, the paraboloid shape does not vary greatly between timepoints, demonstrating the reliability in the definition of the geometric apex.

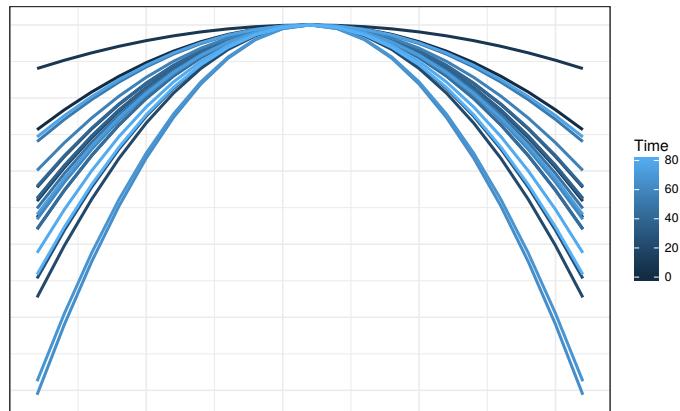


Fig. 3.16 Example depiction of evolution of paraboloid parameters, taken from plant 15, where we have here visualised the axes of curvature according to the equation $y = -bx^2$, where b thus has the same significance as in fig. 3.15. One significant outlier can be seen, corresponding to $t = 48$ hours; other parabolas have curvatures similar in magnitude.

Chapter 4

Discussion

4.1 NPA dilution induces a periodic response

It is likely that the state transition that appears in fig. 3.1 are caused by the slight dilution of NPA that occurs during imaging. Hypothetically, was NPA is diluted, auxin transport would incrementally be restored in the plant and cause initiation of primordia. In support of this, the fact that the observed cyclicity appears to happen in a \sim 16 hour interval suggests that the fluctuations are not primarily driven by the circadian clock, which could otherwise be a cause of the periodicity. Since it is well-known that the network regulating the niche at the SAM consists of multiple feedback loops, it is possible that these are what causes the oscillatory tendencies as the plants trends towards the new steady state induced by the NPA dilution.

While some of the variations between plants can be attributed to biological differences, the overall tendency of the four plants in their behaviour over time indicate a biochemical response to the NPA dilution in a periodic manner. A possible direct pathway for this type of effect could be through the direct WUS or CLV1/CLV3 interaction of NPA, although more likely is downstream factors being influenced by the initialisation of auxin patterns. Nevertheless, a feedback motif responding to the system perturbations would be precisely the type of mechanism which would be expected to drive a periodic response.

That the fluctuations have an increasing amplitude as time progresses can be explained by the iterative dilution of NPA, and a resulting avalanche effect, where more auxin present in the cells effectively reduce the electric repulsion induced by the former. That auxin transport is indeed being restored is emphasised by multiple plants beginning to form primordia throughout the timecourse, as exemplified in fig. B.1.

As the number of division events correlates clearly with the number of nuclei identified, the perturbations in the system seem structural in the sense that the stem cell niche grows not only in size, but also in what number of cells actively proliferate. As the plant grows, we are therefore observing not only a larger number of cells in general, but also a larger number of cells that are part of the effective stem cell niche.

4.2 Distributional cues hints at possible regulatory mechanisms

The regulation of noise at the SAM is explained by our simple activating model, where the upregulation of CLV3 by WUS is sufficient to produce the hysteresis-like features of the distribution. The reason for this is explained by production saturation at the apex when CLV3 levels are already high, and the highest activating concentration being present in this region, such that when CLV3 experiences a minor stochastic degradation event, the activating signal quickly overcomes this downgrading and resets the high expression in the apical cells. Technically, with our configurations used herein, our GRN creates a bistability in the system, where the saturation of the CLV3 activating agent causes high amounts of noise at intermediate distances from the apex. Because of this, the enzymatic activation is sufficient to produce the CLV3 distribution. In the peripheral tissue, WUS is simply not abundant enough for producing a high CLV3 concentration, which also results in a lower tail variance.

While our simple model is sufficient to explain the type of distribution in the epidermis, the apparent scaling in values due to deeper penetration in the tissue suggests a mechanism for repressing CLV3 expression in this region. While both direct repression and repression of an activator such as WUS are possible regulatory causes, previous research has suggested epidermal signalling as a possible mechanism to scaling the stem cell niche. In addition, the distribution differences between the epidermis and the subepidermal layers with respect to the relationship between nuclear volume and CLV3 expression indicate differing regulation between the layers, as we otherwise would expect to see the same types of distributions, albeit scaled. Lastly, as the number of divisions seem to quickly decline in cells beyond the fourth cell from the apex, it is possible that cells at that point are entering a regime which is under the influence of differentiation inducing signals, making them less prone to proliferate. In support of this, the average membrane size per layer does not vary significantly, suggesting that molecular signals, rather than growth and size related ones, are what gives rise to the observed effect.

4.3 *In vivo* tracking sheds light on steady regulation of the true stem cell niche

The behavioural discrepancy between the topmost cells and their neighbours suggests a phenotypic separation between the roughly four highest expressing cells and their neighbours. The fact that plant 2 disagrees might be due to the absolute definition of which cells are the topmost ones, i.e. four in our configuration. It is possible that plant 2 simply maintains a differently sized stem cell niche, or is generally more variable in its expression levels, which would introduce significantly more noise in the measure and explain the lack of observed clusters.

The *in vivo* tracking of lineages in the L1 reveals insight into the maintenance of CLV3 expression at the top. A select few cells are clearly subject to highly regulated CLV3 expression values, whereas the cells outside of this regime appear to be steadily degraded. The repeated signs of the apical cells being special with respect to their phenotype does however suggest that a significant portion of the

observed CLV3 expression in fact stems from the long half-life of dsRED, which is known to be able to last for extended periods on the order of magnitude of several days. Our analysis allows us to better approximate the extent of the factual CLV3 domain, as we through our analysis can account for the degradation time of dsRED. While our approximation of the reporter half-life is approximate and hampered by the lack of complete data, it shows promise for extended analyses of how dsRED is degraded within the living tissue.

4.3.1 Longevity could be induced by the WUS domain

The observed cluster of cells surviving for extended periods of time could most easily be explained by tracking errors and failure to identify a division event, which would cause an amassment of cells living for multiples of the average cell cycle length. However, this type of effect would in principle cause an peak in the division age density also at roughly the 40 hour mark, which we do not see in fig. 3.12. The only clusters seen are instead emphasised around the 20 and 60 hour mark, with a slight downwards shift for *L2* and *L3+* due to cells being lost in the tracking reappearing as newly cells in the analysis.

In addition, the longevity cells appear to be clustered to regions close to the center of the meristem, as depicted in fig. 3.14. A causation of this could be an influencing factor in the OZ, e.g. WUS, which significantly represses proliferation.

A possible factor affecting the results could be the penetretion depth of the laser in the deeper tissue, which due to the paraboloid-like shape of the meristem could cause errors in tracking. Nevertheless, this would not explain the lack of peaks at the ~40 hour mark.

4.3.2 Non-centric CLV3 apex for directed development

That the CLV3 peaks typically tend to be shifted slightly from the geometric apex could have a role in directing growth in phyllotaxis. The fact that we can note a seemingly normal distribution of distances for the CLV3 peaks might indicate that the CLV3 peak rotates slightly around the geometric center over time, in order to specify the direction of growth for organ primordia, which in AT are initiated sequentially in a spiral-like structure. It is also possible that we see under correct angles would be able to see a slight wiggling over the apex over time, although the lack of control of the angle in which the plants were imaged makess this analysis difficult.

Chapter 5

Outlook

5.1 Significance of results

Our analysis presented herein has given extensive novel insights into the dynamics of the stem cell niche of AT. In particular, we have showed how the SAM appears to maintain a few single cells that are truly expressing CLV3, and thus also set the grounds for further investigations into the variance in and extent of this small pool of cells; for example, questions we have not addressed herein are the variability in the size of the true CLV3 domain in relationship to the overall size and eccentricity of the SAM itself.

While previous research has primarily focused on the static view of the stem cell niche, we have here provided parts also of the dynamic perspective – something which has previously been severely lacking. Because of this, we are able to better support the hypothesis of epidermal regulation in the CZ, and also pointed at cues of the extent of differentiation driving peripheral signals. That is, our observed differences in functional behaviour for cells at various distances from the CZ suggests a possible domain for where peripheral signalling is either present, or simply begins to affect cell functionality. Using this information can be of use in particular for modelling, where understanding the extent of unknown regulatory mechanisms can aid in unraveling their nature.

We have also shown how the CLV3 apex does not coincide with the geometrical apex, which might have an impact on anisotropic growth and phyllotaxis. This is a completely unprecedented result, which might provide future insights into how the interplay is orchestrated between the CZ and the peripheral regions which undergo primordia initiation.

5.2 Expansion of modelling framework

Our models herein have taken an abstract approach, which while sufficient for our purposes does not capture the complexity and intricacies of the real tissue. Modern software tools are now able to generate contact maps from confocal images, thus creating the means for accurately comparing model

simulations to the real tissue. This is a possible extension to the work herein, which would also be a natural future validation step.

Establishing models for the $> 1D$ case would in addition allow for validating models relating to the combined aspect of our results. For example, is epidermal regulation sufficient for producing both our distribution for the CLV3 intensities with respect to the distance to the apex, as well as the apparent distributional shift between the L1 and L2? A simple model in e.g. 2D could likely verify the potentiality of such an hypothesis.

Also a spatial model relating to growth rates could be established to test if there are mechanical constraints limiting the growth rate of the apical cells in comparison of their neighbours, and possible biochemical signals regulating this. The combined mechanical and biochemical framework is likely also something which will be able to provide a more complete picture of meristem growth, relating to both scaling of the stem cell niche itself, but also how that relates to cell features such as growth and division rates depending on their spatial localisation.

5.3 Data investigation

A natural and very feasible next step in quantifying the dynamics of the genes present in the SAM would be to also include the PIN1 reporter and observe how this behaves in relation to the CLV3 signal. Possibly this aid in elucidating the extent of both technical noise and what observed fluctuations are true biological events. It also allows insight into the actual auxin transport perspective of the setting, clearly showing that auxin transport is indeed resumed, as well as a glimpse into how this affects the periodicity we observe for the stem cell niche.

Like for PIN1, there are multiple pieces of the data analysis not fully investigated in this thesis. This includes more thorough research into differences between the different layers of the SAM, and how the spatial localisation relates to e.g. cell growth. In particular the CZ of the SAM, steady maintenance between the highest expressing cells and their corresponding wall surface areas facing the CZ could prove to gain a better understanding of the extent and regulation of the CLV3 expressing cells.

The possible clustering of long-lived cells in the deeper, central regions tending towards the OC is a fascinating suggestive discovery, which is somewhat hampered by the difficulty to accurately identify cells within the tissue. One possibility in this case would be to develop methods able to determine cellular age from static images, where the plant can then be cut in the coronal plane in order to improve imaging. This would be enabled by information about, among other measures, relative cell wall angles between neighbouring cells. An additional approach could be to tag cells with cell cycle identifying markers, although also in this case it is likely to be hampered by the penetration depth of imaging tools.

As the data in addition includes multiple segmentation errors of various types, a reevaluation of these could help in better narrowing down the behaviour of especially individual cell lines – something which in our analysis is slightly hampered due to lesser tracking quality outside of the CZ. This holds

especially true for the individual timepoints in the tracking which have been manually corrected for, and therefore do not include cells outside of the CZ; a simple re-tracking of these timepoints would provide both more data for the data set overall, but would in particular enable insight into the cell line dynamics of the subepidermis, and how they correspond to the ones in the L1. Performing an accurate segmentation step also for plant 18 would on top of this provide a larger data set to base the extended analysis on.

5.4 New experiments

Because of the difficulty in unravelling which parts of a GRN is essential for giving it its correct functionality, it is also difficult to know which molecules to track for in *in vivo* time-lapse assays. In the analysis of the regulation of the stem cell niche, having access to tracker information also for the WUS would improve upon the identification of core network dynamics significantly, although this type of data has not been available until recently. Extending the network further, to be able to also track also the patterning activity of auxin on top of the PIN1 data give useful information not only of when and how much auxin transport is activated, but also where this localisation happens in the tissue. Being able to relate this information in particular to the behaviour of cells in particular lineages would provide information of how the CLV3 and cell dynamics depend on auxin localisation.

As the CLV3 level of expression does not appear to provide cues for when a cell is to undergo a division event, identifying the driving genes causing this event to happen is a future endeavour. In the event that no such gene exists, and that it is predominantly mechanical signalling that determines proliferation, the interplay between the mechanical and the chemical is perhaps the more relevant. The question in that case would then primarily be to identify growth-inducing agents, rather than direct division event markers.

A drawback with the analysis presented herein in the reliance on meristems grown on NPA for the analysis which is better suited for the steady state. However, as physically removing organs and primordia for enhancing imaging is majorly perturbing the system in itself, doing so at this point is not feasible. The future might hold tools for performing this type of research, which would greatly enhance to quality and robustness by which the analysis of e.g. the CLV3 distributions and behaviour over time is done.

Lastly, the reason for the use of a promotor tagging molecule in dsRED in our experimental data is the size of the CLV3 peptide. As attempts to track the protein itself leads to non-functional plant phenotypes, understanding the specific protein patterning is as of yet practically unfeasible in a dynamic setting. The future might hold advances in methodology that enables direct feedback of peptide localisation, which undoubtedly would provide a major stepping stone towards truly understanding plant development.

References

- [1] Bozorg, B. (2016). *Models of Mechanics and Growth in Developmental Biology: A Computational Morphodinamics [sic.] approach*. PhD thesis, Lund University, Sweden.
- [2] Brand, U., Grünwald, M., Hobe, M., and Simon, R. (2002). Regulation of clv3 expression by two homeobox genes in arabidopsis. *Plant physiology*, 129(2):565–575.
- [3] Ciliberti, S., Martin, O. C., and Wagner, A. (2007). Robustness can evolve gradually in complex regulatory gene networks with varying topology. *PLoS computational biology*, 3(2):e15.
- [4] Clark, S. E. (2001). Cell signalling at the shoot meristem. *Nature reviews. Molecular cell biology*, 2(4):276.
- [5] Debeaujon, I. and Koornneef, M. (2000). Gibberellin requirement for arabidopsis seed germination is determined both by testa characteristics and embryonic abscisic acid. *Plant physiology*, 122(2):415–424.
- [6] Fernandez, R., Das, P., Mirabet, V., Moscardi, E., Traas, J., Verdeil, J.-L., Malandain, G., and Godin, C. (2010). Imaging plant growth in 4d: robust tissue reconstruction and lineageing at cell resolution. *Nature methods*, 7(7):547–553.
- [7] Food and of the United Nations, A. O. (2017). Dimensions of need: an atlas of food and agriculture 2017.
- [8] Garcia-Alvarez, D. (2011). A comparison of a few numerical schemes for the integration of stochastic differential equations in the stratonovich interpretation. *arXiv preprint arXiv:1102.4401*.
- [9] Green, M., Krupinski, P., Melke, P., Sahlin, P., and Jönsson, H. (2017). Costanza - confocal stack analyzer application.
- [10] Gruel, J., Landrein, B., Tarr, P., Schuster, C., Refahi, Y., Sampathkumar, A., Hamant, O., Meyerowitz, E. M., and Jönsson, H. (2016). An epidermis-driven mechanism positions and scales stem cell niches in plants. *Science advances*, 2(1):e1500989.
- [11] Hamant, O., Heisler, M. G., Jönsson, H., Krupinski, P., Uyttewaal, M., Bokov, P., Corson, F., Sahlin, P., Boudaoud, A., Meyerowitz, E. M., et al. (2008). Developmental patterning by mechanical signals in arabidopsis. *science*, 322(5908):1650–1655.
- [12] Heidstra, R. and Sabatini, S. (2014). Plant and animal stem cells: similar yet different. *Nature reviews. Molecular cell biology*, 15(5):301.

- [13] Hutchison, C. E. and Kieber, J. J. (2002). Cytokinin signaling in arabidopsis. *The Plant Cell*, 14(suppl 1):S47–S59.
- [14] Ietswaart, J. H. R. (2015). *Spatiotemporal modelling in biology: from transcriptional regulation to plasmid positioning*. PhD thesis, University of East Anglia.
- [15] Jönsson, H. and Levchenko, A. (2005). An explicit spatial model of yeast microcolony growth. *Multiscale Modeling & Simulation*, 3(2):346–361.
- [16] Kerstetter, R. A., Bollman, K., Taylor, R. A., Bomblies, K., and Poethig, R. S. (2001). Kanadi regulates organ polarity in arabidopsis. *Nature*, 411(6838):706.
- [17] Kitano, H. (2002). Systems biology: a brief overview. *Science*, 295(5560):1662–1664.
- [18] Klimontovich, Y. L. (1990). Ito, stratonovich and kinetic forms of stochastic equations. *Physica A: Statistical Mechanics and its Applications*, 163(2):515–532.
- [19] Kvrevcek, P., Skuupa, P., Libus, J., Naramoto, S., Tejos, R., Friml, J., and Zavimalova, E. (2009). The pin-formed (pin) protein family of auxin transporters. *Genome biology*, 10(12):249.
- [20] Laux, T., Mayer, K., Berger, J., and Jurgens, G. (1996). The wuschel gene is required for shoot and floral meristem integrity in arabidopsis. *Development*, 122(1):87–96.
- [21] Lawrence, P. A. and Struhl, G. (1996). Morphogens, compartments, and pattern: lessons from drosophila? *Cell*, 85(7):951–961.
- [22] Lempe, J., Lachowiec, J., Sullivan, A. M., and Queitsch, C. (2013). Molecular mechanisms of robustness in plants. *Current opinion in plant biology*, 16(1):62–69.
- [23] Locke, J. C., Young, J. W., Fontes, M., Jiménez, M. J. H., and Elowitz, M. B. (2011). Stochastic pulse regulation in bacterial stress response. *science*, 334(6054):366–369.
- [24] Losick, R. and Desplan, C. (2008). Stochasticity and cell fate. *science*, 320(5872):65–68.
- [25] Menn, D. J. and Wang, X. (2001). Stochastic and deterministic decision in cell fate. *eLS*.
- [26] Mirabet, V., Besnard, F., Vernoux, T., and Boudaoud, A. (2012). Noise and robustness in phyllotaxis. *PLoS Computational Biology*, 8(2):e1002389.
- [27] Reinhardt, D., Pesce, E.-R., Stieger, P., Mandel, T., et al. (2003). Regulation of phyllotaxis by polar auxin transport. *Nature*, 426(6964):255.
- [28] Scheres, B. (2007). Stem-cell niches: nursery rhymes across kingdoms. *Nature Reviews Molecular Cell Biology*, 8(5):345–354.
- [29] Scofield, S., Dewitte, W., Nieuwland, J., and Murray, J. A. (2013). The arabidopsis homeobox gene shoot meristemless has cellular and meristem-organisational roles with differential requirements for cytokinin and cycd3 activity. *The Plant Journal*, 75(1):53–66.
- [30] Wickham, H. et al. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40(1):1–29.
- [31] Willis, L., Refahi, Y., Wightman, R., Landrein, B., Teles, J., Huang, K. C., Meyerowitz, E. M., and Jönsson, H. (2016). Cell size and growth regulation in the arabidopsis thaliana apical stem cell niche. *Proceedings of the National Academy of Sciences*, 113(51):E8238–E8246.

Appendix A

Data quality and errors

A.1 Tracking quality

The data tracking quality is defined as the F1 score, also known as the Dice-Sørensen score, between timepoints. The F1 score gives an appreciation of the overall accuracy of a test on the form of

$$F_1 = 2 \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}, \quad (\text{A.1})$$

i.e. as the harmonic mean between precision and recall. It can also be expressed on set form as

$$F_1 = 2 \frac{X_t \cap X_{t+1}}{X_t \cup X_{t+1}} \quad (\text{A.2})$$

where X_t denotes the set in question in timepoint t . In other words, the more cells are contained between timepoints, the higher the F1 score.

As fig. A.1 shows, the quality for the tracking is overall relatively high, with only a few timepoints having bigger clusters with low F1 scores. These cells are however typically positioned at the periphery of the SAM, at the very edges of the segmentation, and are therefore both of lesser interest and importance for the dynamics of the stem cell niche. As described in section 2.3.2, we have in all of our analyses excluded the tracking of cells whose F1 scores are less than 0.30. The errors in tracking occasionally causes one cell to be recognized as more than two in the next timeframe. In these cases, we choose the two daughter cells by order of tracking quality.

That a timepoints 40 and 48 hours in plant 2, as well as timepoint 20 hours in plant 4 have perfect F1 scores are due to the tracking procedure failing at these points. The tracking has therefore been manually corrected for the cells within 30 μm of the highest CLV3 expressing cell.

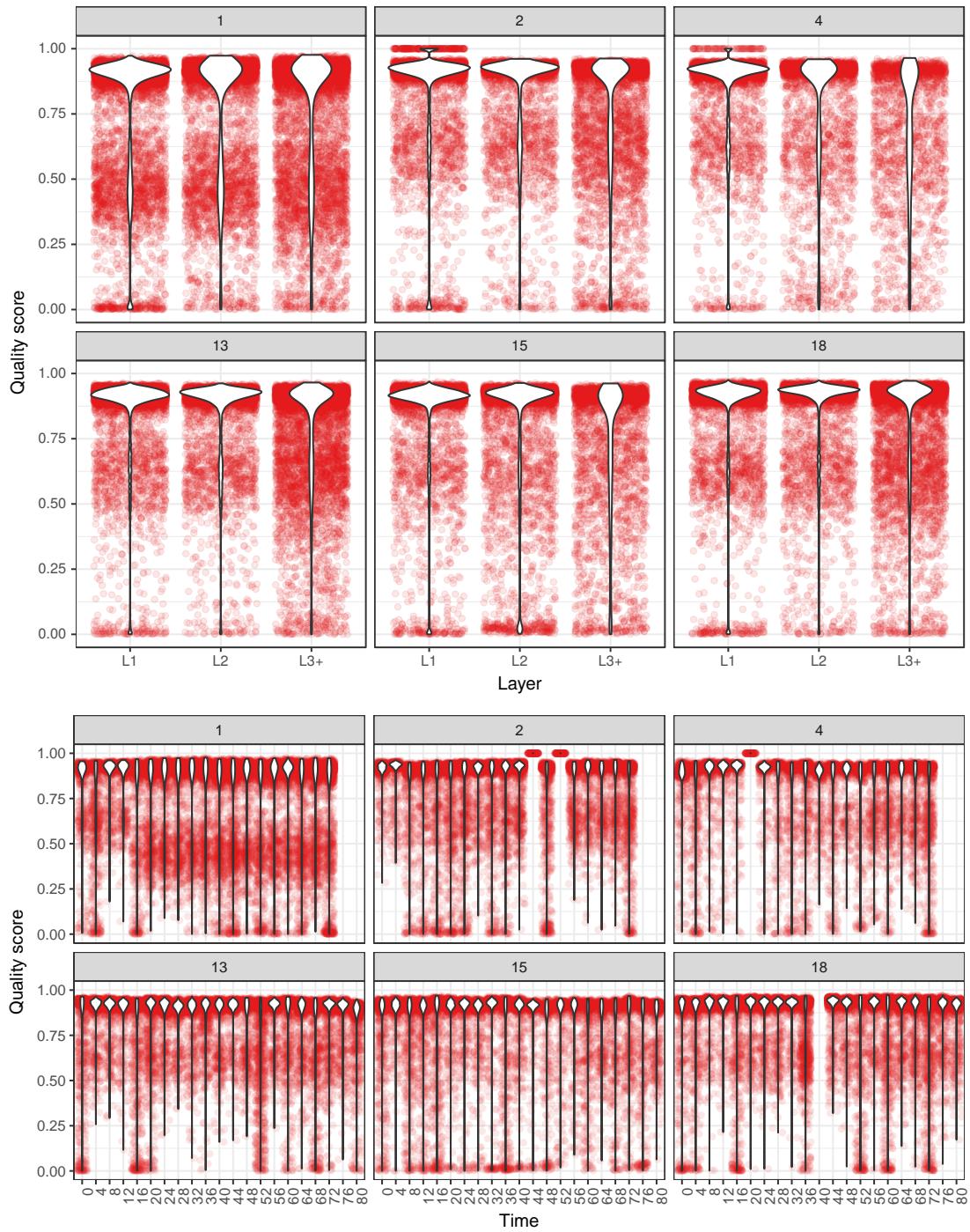


Fig. A.1 Quality distributions per plant, layer, and timepoint for all plants. Some timepoints have distributions scoring in the ~ 1 regime, which corresponds to situations where the automatic tracking failed. These timepoints have been manually tracked for cells in the L1 within $30 \mu\text{m}$ of the topmost CLV3 expressing cell.

A.2 Segmentation quality

The data is afflicted with several types of segmentation errors. For both the nuclear and the membrane segmentation, this typically takes the form of basins of attraction identified as either one when there in fact are multiple, or multiple when there is in fact only one. In particular, this causes errors with respect to the appreciation of membrane and nuclear volumes. It also indirectly affects the tracking quality, as clumped cells in one timepoint can lead to an erroneously identified division event in the next, in case the segmentation becomes correct.

Segmentation errors are however the most prevalent in the peripheral regions of the SAM, and with deeper penetration into the tissue. Generally, the membrane segmentation holds higher quality than the nuclear one, where the latter degrades quickly after deeper penetration than the L2. The overall quality is also decreasing with the distance from the apex, although to a minor extent in the L1.

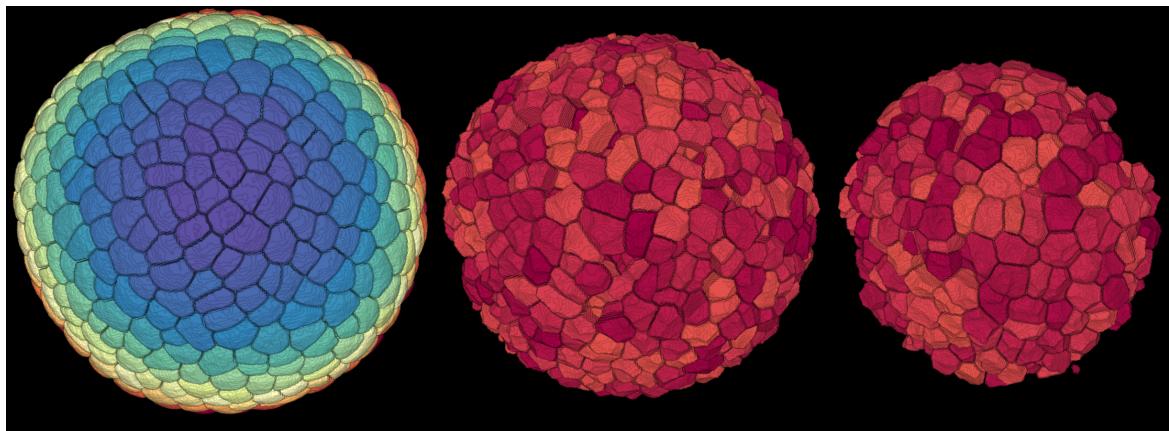


Fig. A.2 Membrane segmentation quality separated into L1, L2 and L3 respectively. L1 shows a smooth upper surface with well-segmented cells. Also L2 and L3 show high segmentation quality, but with a rougher surface to the cells being restricted by the cell layer on top, causing a ruggedness in the cell surface. Note in particular how the quality in the periphery of the SAM generally are of lesser quality, with occasional cell remnants being identified as actual cells, e.g. in the bottom right.

Generally, the overall quality for the membrane segmentation is high in all three layers, as shown in fig. A.2, although the number of errors increase with deeper tissue penetration. For the nuclear segmentation, mostly cells from the L1 and L2 are sufficiently reliable, with declining quality with larger distance to the CZ. The errors with multiply matching nuclei are occurring mostly in the subepidermal cells, and in those cases largely in the peripheral regions.

A.3 Missing timepoints

The dataset contains several missing timepoints in the nuclear and membrane data. As previously noted, plant 1 and 18 lacks all timepoints for the nuclear data, whereas plant 4 and 15 lack nuclear data for timepoints 24 and 12 hours respectively. Plant 18 also lacks data for timepoint 40 hours.

Appendix B

Supplemental figures

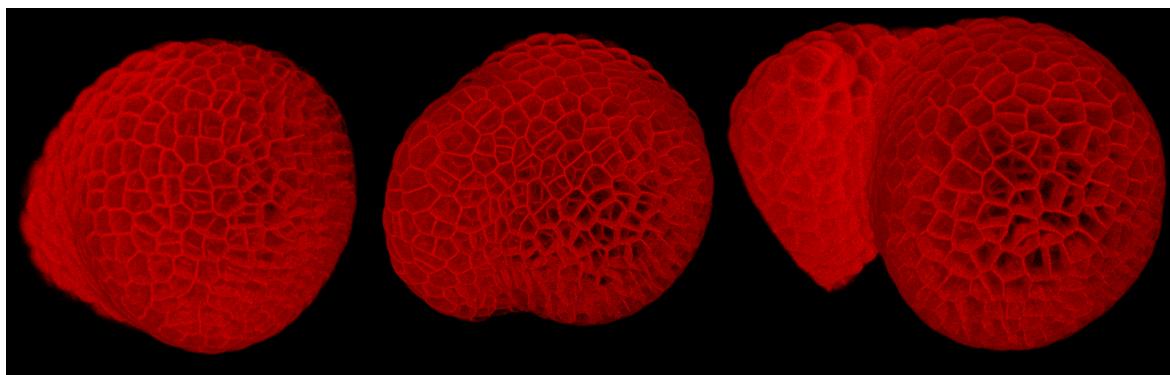


Fig. B.1 Membrane channel for plant 18 at timepoints 0, 36 and 84 hours respectively. In the upper left, formation of a primordium throughout the course of the timelapse is clearly seen, indicating NPA dilution and reactivation of auxin transport.

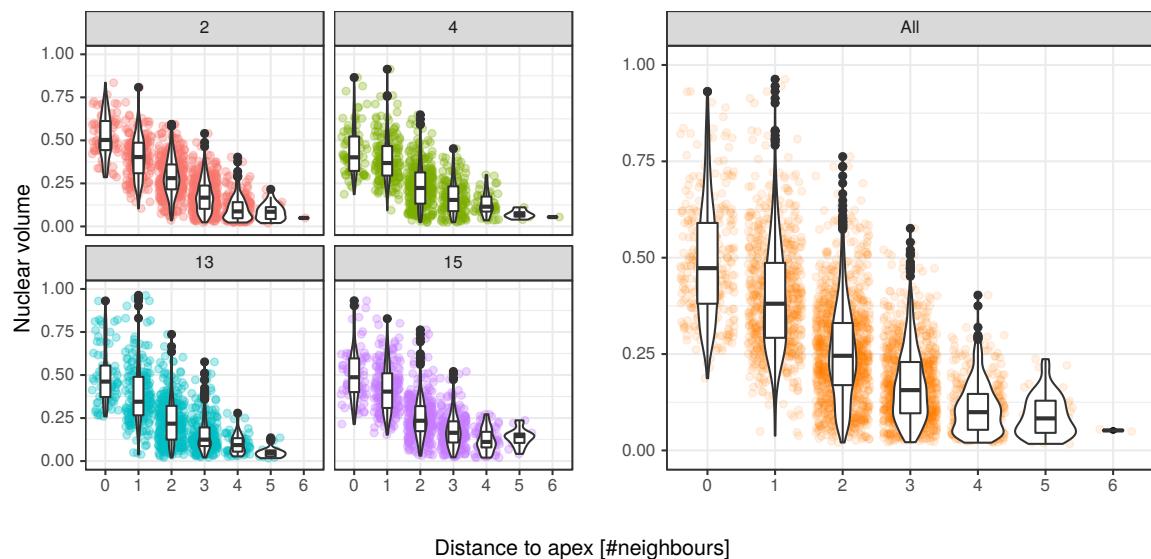
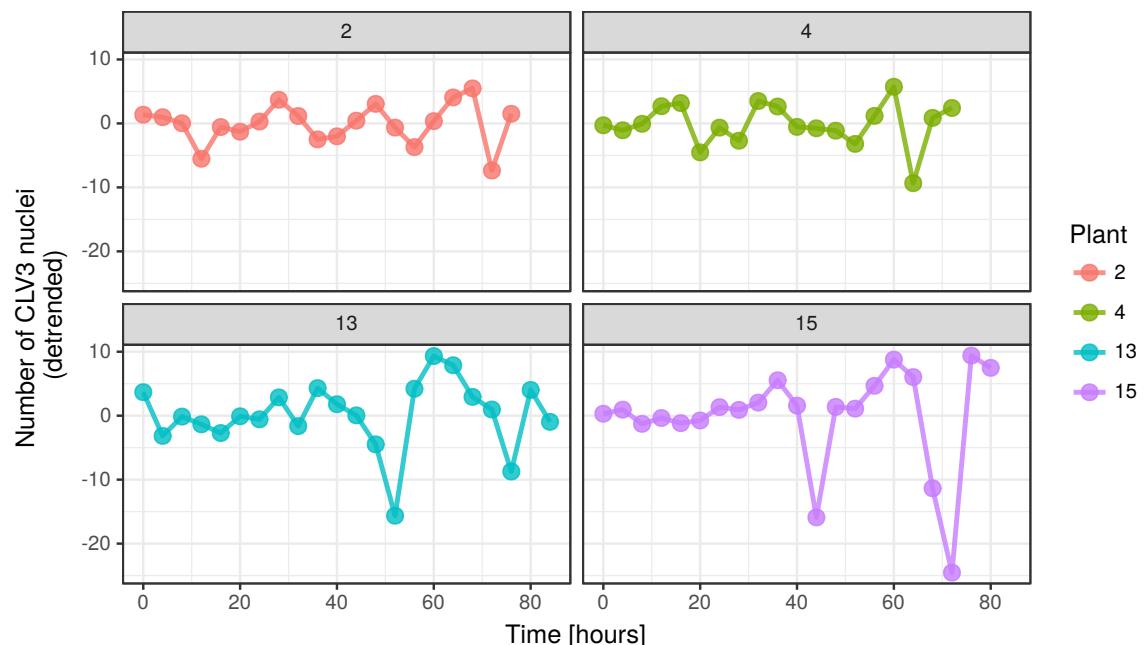


Fig. B.2 Distribution of CLV3 nuclear volumes at apex, in the epidermis. Lack of data in the periphery is due to CLV3 loss-of-signal with increasing distance from the CZ.



. The detrending is done by subtracting the value curve attained by fitting a second order Loess curve to each individual nuclear trace.

Fig. B.3 Detrended view of trajectories visualised in ???

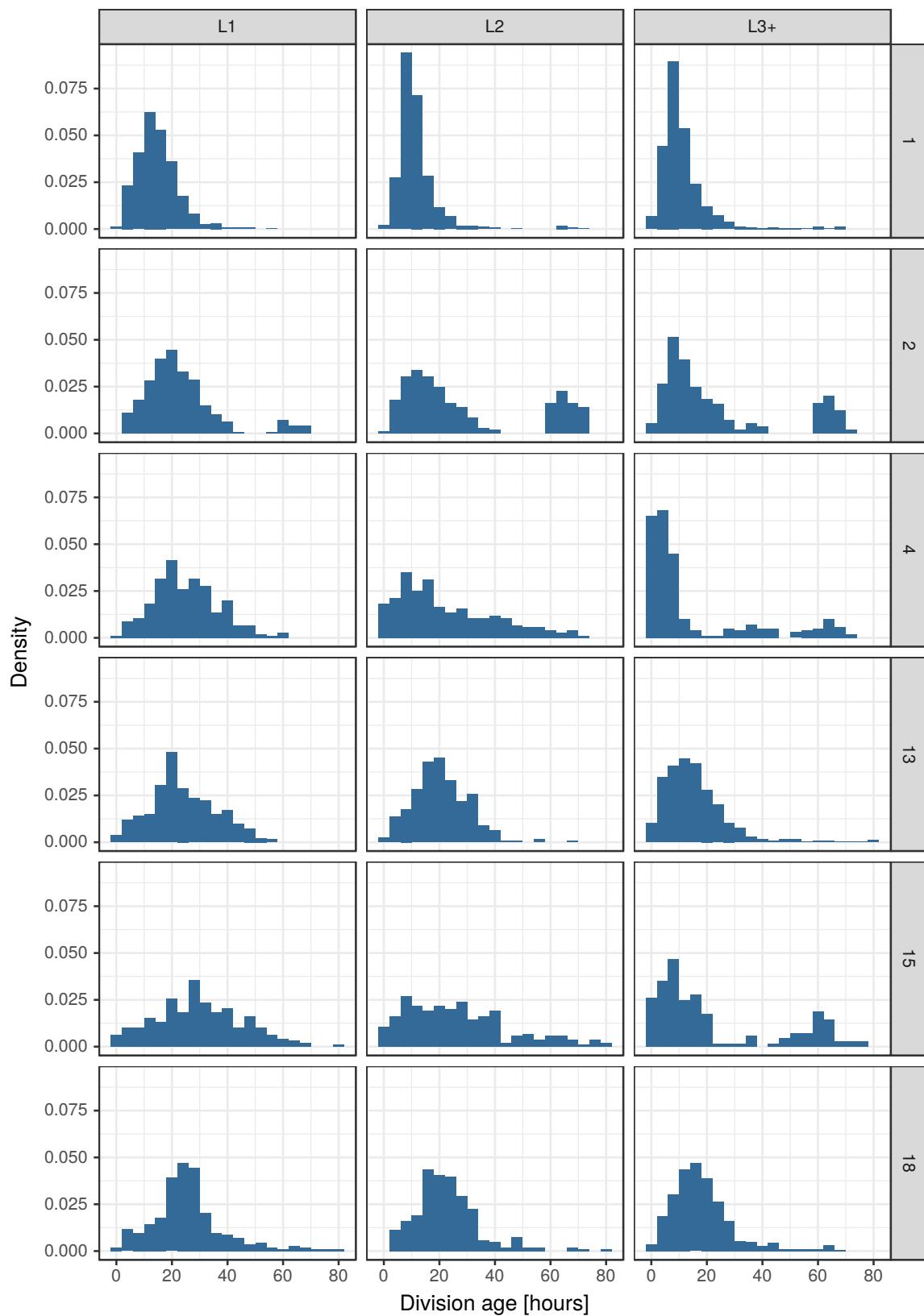


Fig. B.4 Age distribution for all plants, showing the general tendency of more cells of higher typically being present in the L2 and L3.

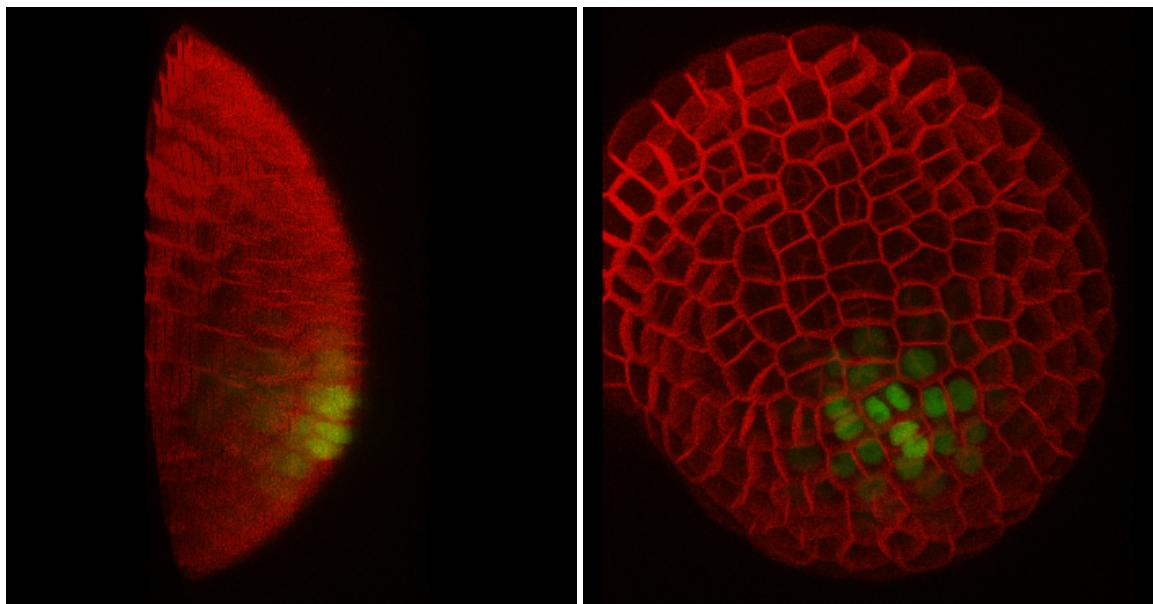


Fig. B.5 Raw data illustration in sideview (left) and topview (right). The CLV3 peak expression, defining the CLV3 apex, can be seen in green, clearly not coinciding with the geometric apex. Example taken plant 4, timepoint 48 hours.

Appendix C

Software Descriptions

C.1 Costanza

Costanza (COnfocal STack ANalyZer Application) [9] is an ImageJ plugin for segmenting compartments in the form of cells and extract quantitative data, including intensities. Primarily, Costanza is used to segment nuclei marked cells in three dimensions and effectively extract information relating to the intensity of the used GFP markers.

The software utilises a steepest gradient ascent approach for segmentation, which initiates at each voxel in the stack and attempts to find local intensity maxima by ascension in the neighbourhood of the voxel. All paths leading to the maximum are then grouped and recorded as a Basin Of Attraction (BOA), i.e. cell. In the event that multiple neighbourhood voxels have a higher intensity than the current one, the path with the highest ratio of intensity difference over spatial difference is chosen, i.e. the path returning $\max\left(\frac{\Delta I}{\Delta \vec{r}}\right)$.

Costanza performs preprocessing in the form of intensity inversion, background extraction and applied denoising filters. For postprocessing, it allows for Basin-Of-Attraction (BOA) removal and merging, in order to exclude misidentified cells in the background and avoid faulty separation of individual cells.

C.2 MARS-ALT

MARS-ALT is software developed for spatiotemporal tissue reconstruction and lineageing at cell resolution. The software consist of the two parts MARS (Multi-angle image Acquisition, 3D Reconstruction and cell Segmentation), and ALT (Automated Linage Tracking). It works by importing fluorescently stained confocal images, and (optionally) correcting these using low-resolution reference stacks.

For cell segmentation, MARS denoises the image using an alternate sequential filter in order to increase the signal / noise ratio. Seeds are then extracted by computing voxel minima, and merging those that have a lower largest valley between them than some defined value. The background is

then extracted from the largest connected component found in the image after thresholding. Cells are thereafter watershed using the seeds as references, with cell volumes less than some specified value are filtered. Markers are subsequently removed from the seeds, and the watershed algorithm is repeated until convergence [6].

C.3 Organism

Organism is C++ software for simulating biological systems, in particular with multiple compartments primarily in the form of cells. It includes both biochemical and mechanical rules, including rules for proliferation, for numerically simulating the system. Organism relies on a representation of compartments such that all of them can be described by a fixed number of variables and parameters, such as spheres, cylinders and similar. It can also be used for simulations of natural tissues, extracted from confocal microscopy images. As part of the Organism toolset, software for visually inspecting simulated tissue from output data is accessible (Newman), as well tools for performing parameter optimisation of GRNs [15, 10].

For this thesis, an R wrapper for interacting with parts of the Organism software was constructed, allowing for R-based solver, model, parameter, and initialisation file setup. Code relating to this is hosted at the *Sainsbury Laboratory at the University of Cambridge* GitLab portal, and can be accessed at https://gitlab.com/slcu/teamHJ/henrik_aahl/organism_wrappeR.

C.4 extractoR

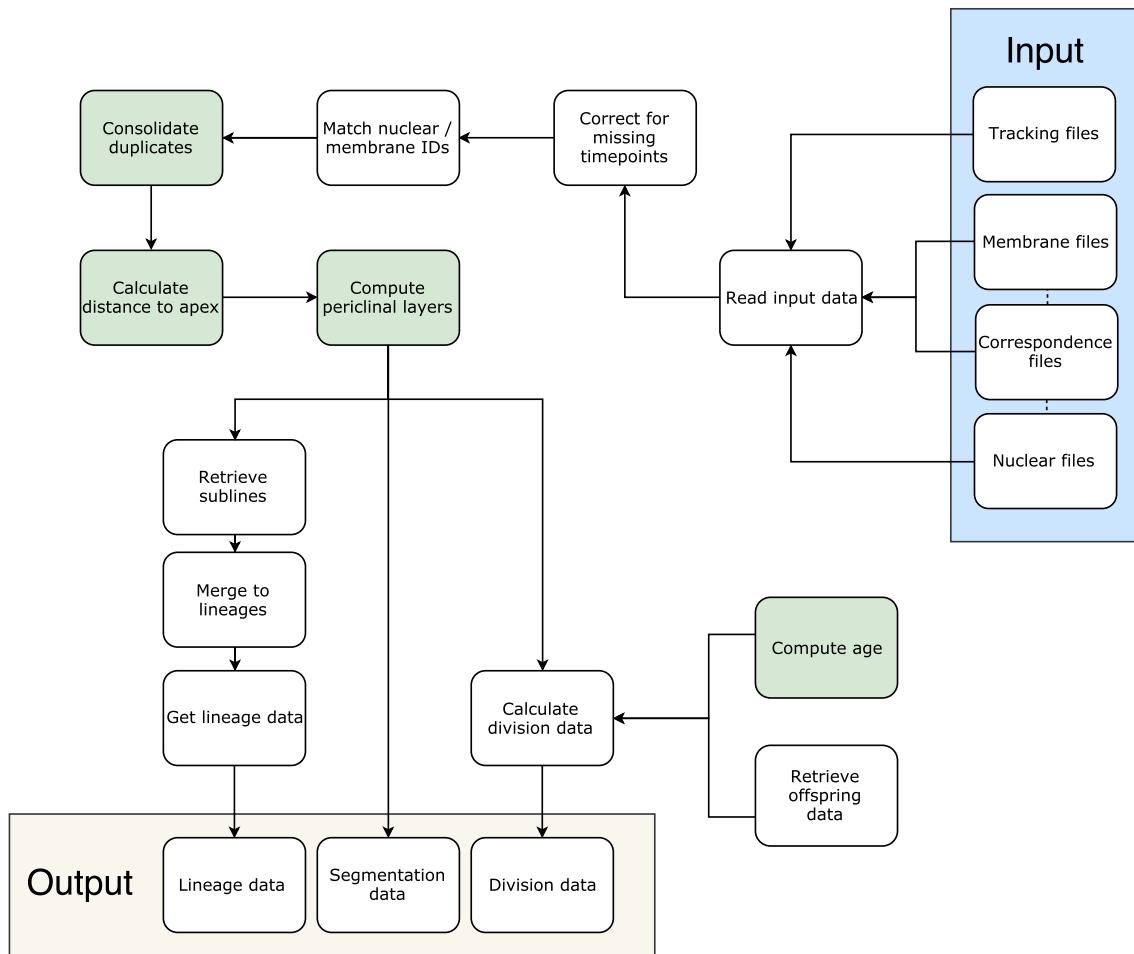


Fig. C.1 Flowchart description of the extractoR software used within this study.

extractoR is convenience software developed for the data analysis part of this thesis, with the general workflow described in fig. C.1. The software is developed in *R*, with minor data import scripts written in Python. It builds on the *tidyverse*, or *split-apply-combine*, principle [30] with data separated primarily into tabular format. For convenience, extractoR outputs the three tabular data objects

1. Segmentation data
2. Division data
3. Cell line data

all consisting of identifiers for the corresponding plant, timepoint, cell ID, and other related metrics. The software utilises minor parallelisation in several processing steps for optimisation purposes.

Code and related files are as of the date of publication accessible on the SLCU GitLab page via https://gitlab.com/slcu/teamHJ/henrik_aahl/extractoR.

C.5 TissueViewer

TissueViewer is in-house software primarily developed by Yassin Refahi at SLCU. The software provides a user-friendly visualisation tool for segmented cellular images, and is equipped with a toolbox able to separate visualisation to select cells, e.g. layer-wise.

TissueViewer also includes the plugin Meristem Phenotyper 3D, developed by Max Brambach at SLCU, which allows for fitting of geometric structures to raw meristem images. In this thesis, we use this to fit paraboloid shapes to our meristem images in order to extract information about the geometric apex, for comparison to the one given by the gene expression.