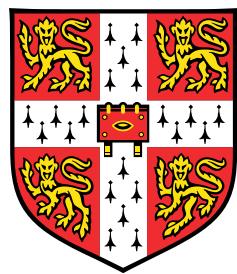


Interpreting noise in stem cell regulation in the shoot apical meristem of *Arabidopsis thaliana*



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

Insert pretentious quote here

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 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Henrik Åhl
August 2017

Acknowledgements

José, Henrik, Jérémie, Max, Julia, Yassin, Klara + Jönsson group

Abstract

What-Why-How-Results-Significance

Table of contents

List of figures	xiii
List of tables	xv
Nomenclature	xvii
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	3
1.3.1 Molecular tuning determines cell phenotype	3
1.3.2 Developmental Regulation in the Apical Meristem	3
1.4 Background (etc.)	4
2 Methodology	5
2.1 Raw data	5
2.2 Processed Image data	5
2.3 Data Filtering	6
2.4 Models of Gene Regulatory Networks	7
2.4.1 Mathematical Formulation	7
2.4.2 Stochastic Simulations	9
2.5 My models (come up with title)	11
3 Results	13
3.1 Regulation of the CLV3 domain induces periodicity when perturbed	13
3.2 CLV3 noise is kept at bay at apex	15
3.3 Distribution shifting suggests epidermal regulation	16
3.3.1 CLV3 is induced in the epidermis	16
3.4 Functional clustering implies dsRED technical artefacts	17
3.5 Deep-tissue cells divide at a slower rate	19

4 Discussion	21
4.1 NPA dilution induces oscillations	21
4.2 <i>In vivo</i> tracking of apical stem cells agrees with hypothesis	21
4.3 Distributional cues give hints at regulatory scheme	21
4.3.1 Activation of CLV3	21
4.3.2 Longevity (Don't know what this shows yet)	21
4.3.3 Definition of center	21
5 Outlook	23
5.1 Model expansion	23
5.2 Data investigation	23
5.3 New experiments	23
References	25
Appendix A Addition Information on Data Acquisition	27
A.1 Experimental procedure	27
A.2 Segmentation	28
A.2.1 Settings and approach	28
A.3 Errors in data	28
A.4 Layer Quality Assessment	28
Appendix B Extended data analysis	29
B.1 NPA dilution is revealed by primordia initiation	29
B.2 Just cram in the rest of the data	29
Appendix C Software Descriptions	31
C.1 Costanza	31
C.2 MARS-ALT	31
C.3 Organism	32
C.3.1 Main software	32
C.4 extractoR	32

List of figures

3.1	Add caption	13
3.2	Add caption	14
3.3	Add caption	14
3.4	Add caption	15
3.5	Add caption	16
3.6	Add caption	17
3.7	Add caption	17
3.8	Add caption	18
3.9	Add caption.	18
3.10	Add caption	19
B.1	Add caption	29

List of tables

2.1	Consolidation methods applied for duplicate nuclei.	7
2.2	Filtering settings	7
3.1	TODO	14

Nomenclature

Acronyms / Abbreviations

AT *Arabidopsis thaliana*

CLV3 CLAVATA-3

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

CZ Central zone

GRN Gene Regulatory Network

KAN1 KANADI-1

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

RAM Root Apical Meristem

SAM Shoot Apical Meristem

SLCU Sainsbury Laboratory at the University of Cambridge

WUS WUSCHEL

Chapter 1

Introduction

1.1 The Shoot Apical Meristem of *Arabidopsis thaliana*

Plant stem cells are governed 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, which 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.

ADD CITATIONAs 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.

The stem cells at the SAM contribute to the construction of new organs and general tissue by dividing frequently at the top and subsequently being mechanically pushed out of the center in order to differentiate. 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. **ADD CITATION**

In a simple outline of the SAM, it can be said to consist of three core regions: 1. The *central zone* (CZ), which harbors the aerial *stem cell niche* of the plant; 2. The *Rib Meristem*, which is located beneath the CZ and consists of the cells constructing the stem of the plant; 3. The *Peripheral Zone* (PZ), where cells form organs and new tissue through differentiation. 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 one, 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 L1 or L2 correspondingly. In contrast, this is not the case for L3, where cells can divide in all directions. In addition, it has been shown that the epidermis is involved in both promoting and restricting shoot development, 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

Due to the interaction of molecules in various ways, e.g. the activation or repression of transcription by certain proteins, organismal development can be considered in the framework of being a *complex system*. 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. In a molecular setting, the typical descriptive approach is through *Gene Regulatory Networks* (GRNs), where each component represents some molecular aspect of the system that is involved in producing expression levels of mRNA, proteins and hormones. **ADD CITATION**

GRNs are commonly understood both through analytical and computational means, where in the latter computer-generated models provides as a modern tool for better understand the complex nature of many biological systems. Typically, reaction kinetics are modelled using various types of *Ordinary Differential Equations* (ODEs). However, due to the large supplies of computer power available in the modern day, many recent studies also utilise more computationally demanding resources such as *Stochastic Differential Equations* (SDEs), where also the inherently random nature of molecular motions, interactions, and processes are accounted for. In particular stochastic modelling of biological systems able to capture dynamical features that deterministic versions cannot. For example, cells often require in various ways to be able to switch between active and inactive states, e.g. when committing to producing a certain protein or not. Utilising the inherent noise of microscopic systems, cells have been shown to probabilistically tune their responses; **ADD CITATION** modelling these types of phenomena using stochastic approaches has gained much insight into the decision-making process of cells both specifically and in general. **ADD CITATION**

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 a changing setting. A straightforward example taken directly out of the context of plant development is how the distribution of gene expression varies during plant growth and organ formation, both spatially and over time. Typical modelling aspects at such a problem case involves the formulation of which genes and molecules are important for the problem of interest, as well as how the discretisation and representation of spatial elements is done. **ADD CITATION**

Computer models in general have two separate aims: exploration and verification. In the former case, computer simulations can be the core for designing experimental experiments, where observed theoretical phenomena can be experimentally tested. An example of this is the classic example of the *Repressilator* **ADD CITATION**, where researchers set up a model framework for how oscillations could occur due to cyclic repressive interactions between three genes. This was then verified to occur by synthetically implementing the system in a bacterium, showing how gene expression profiles could oscillate due to the system motif constructed.

In the latter case, which is the more prevalent in modern computational biology, computer models are established in order to verify or support the potential of an hypothesis due to experimental

observations. **ADD CITATION****REWRITE**

1.3 Regulatory Mechanics of Plant Stem Cells

1.3.1 Molecular tuning determines cell phenotype

The ultimate phenotype of a cell is to a large degree 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. **ADD CITATION**Similar to in animals, stem cells require an intricate network both specifying the pluripotency to 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 patterning, which also plays a role in specifying zones of initiation of primordia.

A viable and robust network maintaining patterning is thus important for the plant in order undergo phyllotaxis in a functional manner, and to know when and how to commit to more structural changes. Substances 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 its local configuration of molecular concentration.

Morphogen patterns can consist of several types of spatiotemporal expression, including that of hormones, proteins and RNA localisation, 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. **ADD CITATION**(x2) 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 CLV3, which is expressed in a 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 quickly out of the cell. The clv3 gradient extends down to the OZ where it acts repressively on the WUS gene.

ADD FIGURE

In contrast, 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

Add information on CLV

Add flowchart of shoot n

WUS loss-of-function mutants, where the lack of the correct WUS gradient leads to defective shoots that terminate in aberrant flat structures. **ADD CITATION**(Laux 1996)

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 KAN1 **ADD CITATION**are known to promote cell differentiation. 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. Also other homeobox encoding genes such as *Shoot Meristemless* (STM) are essential for correct development of the shoot.

Aside of the system regulating the cellular identity, even further substances 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. **ADD CITATION**

As a 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 Background (etc.)

An important question in developmental biology is how organisms can have robust development despite consisting of many independently variable parts. **ADD CITATION**(x3)

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.

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.

Quantification of the SAM tissue at the single cell level

Stable regulation (and regulation *per se*)

CLV3 tracking and anisotropic growth

Chapter 2

Methodology

2.1 Data consists of *in vivo* confocal timelapses

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 **ADD CITATION**, 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.

Using confocal microscopy, the six plantlings were tracked in intervals of 4 hours up to 76 (*plants 1, 2, 4*) or 84 hours (*plants 13, 15, 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 induced 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. **ADD CITATION**

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

Add figure of segmented

segmentation software *MARS-ALT*. Segmentation and tracking was thereafter done 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 likely incorrect mappings. These cells are included in the overall analysis, but excluded from all cell line related investigations.

A longer outline of this is presented in section A.3.

The nuclear data were deconvolved to account for the microscope's point-spread function using the *PSF distiller* tool from Huygens software 15.05 **ADD CITATION**. As in the membrane case, the nuclear channels were adjusted with the corresponding stretching factors and thereafter segmented using segmentation tool *Costanza* **ADD CITATION**.

In order to link nuclei to membranes, the coordinates of each variable were fitted using a least square approach. Duplicately mapping nuclei were then consolidated as described in section 2.3. Note that whenever spatial coordinates are referenced in relation to membrane or nuclei, they are done so as the centroid coordinates of the basin of attraction found in the segmentation process.

Measures of distances to the top were done in multiple ways. The three methods herein considered consist of a definition of the top based on 1) the spatial coordinates, 2) the expression value, and 3) a least-square fit of a paraboloid to raw meristem images. 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 n average spatial coordinates of the highest expressing CLV3 nuclei. 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.

In order 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.

2.3 Data Filtering

Due to thresholding effects for cell nuclei during the segmentation, some individual nuclei are occasionally identified as two or more. In order to account for this, nuclei were 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 2.1. 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 below in table 2.2. The

Table 2.1 Consolidation methods applied for duplicate nuclei.

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

choice of allowed deviance was done based on the distribution shape, 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.

figure of these?

For reference, a quality assessment of the data can be found in section A.3. In addition, descriptions of the software used in segmentation and tracking are briefly introduced in ??.

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 implied in ?? (modelling biological systems), we here work using an ODE or SDE description of our regulatory systems.

We in this thesis 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

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 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.

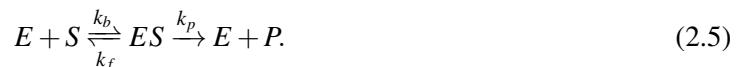
Expanding on this, we can easily solve for the steady-state concentrations of the system by assuming that all rates average to zero. In our example above, this gives us

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

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.

Update of system Time and molecular numbers are update 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

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

although modern high-level programming languages do in some cases perform better using the inherent random number generator for a specific type of distribution. **ADD CITATION**

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 . **ADD CITATION** 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 **ADD CITATION** and gives 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 $\mathcal{O}(\Delta t)$. In this thesis, we utilise the Milstein approach under the Stratonovich interpretation.

2.5 My models (come up with title)

In this thesis, 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 ∇ is the *Laplace operator*.

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 in [ADD FIGURE](#)(timelapse of example plant over 4 timepoints?). Similarly, fluctuations in this number can be seen not to correlate with the mean expression as described in table 3.1.

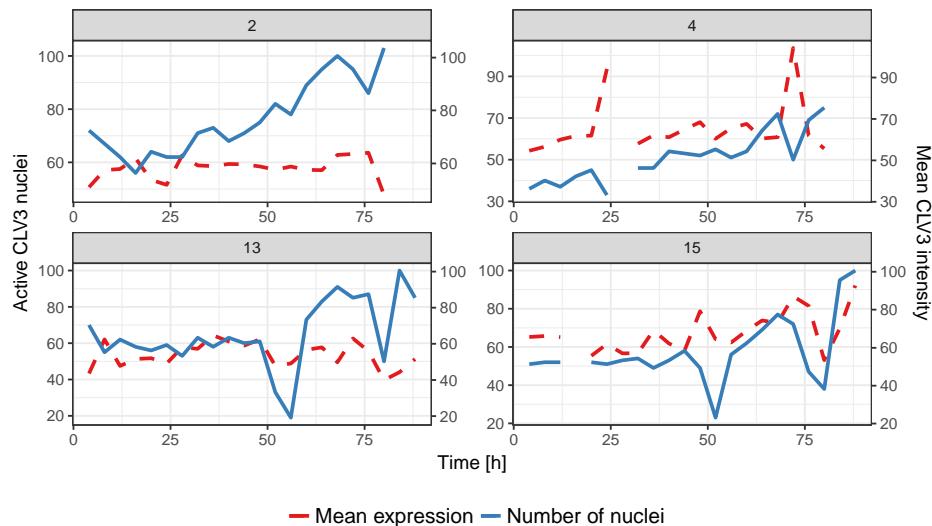


Fig. 3.1 Add caption

In order to assess the extent of the fluctuations the lines were detrended using a second order Loess fit to each curve. When performing a continuous time Fourier transform in order to extract amplified modes, the plants have are biased towards the fourth mode in each respective transformation, as depicted in ??, which corresponds to periodicity of ~ 16 hours.

Table 3.1 TODO

Plant	p-value
2	0.48
4	0.84
13	0.91
15	0.11
All	0.50

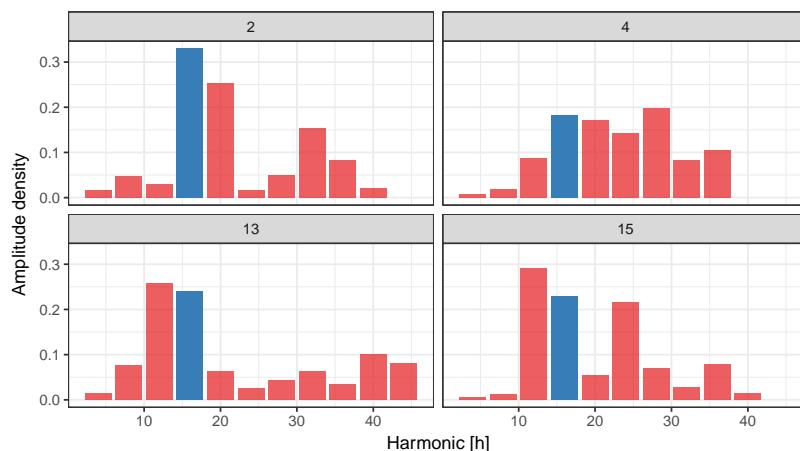


Fig. 3.2 Add caption

In addition to this periodicity, the number of nuclei identified correlates positively with the number of division events observed in each timepoint for three out of the four plants (fig. 3.8). Plant 13, like in fig. 3.1 is the only anomalously behaving specimen.

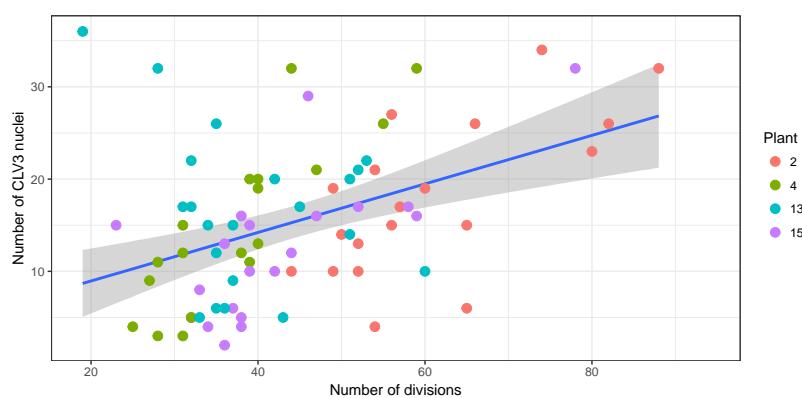


Fig. 3.3 Add caption

3.2 CLV3 noise is kept at bay at apex

Variance in the expression of CLV3 is suppressed in the apical cells, as can be seen in fig. 3.5. This appears true for all plants, although plant 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. Accounting for this possibility, all plants have hysteresis-like shapes. Nevertheless, the distribution of expression values at distinct distances does not appear to be bimodal, as would indicate bistability in stem cell regulation. Instead, the data suggests a CLV3 gradient.

At lower longer distances, the expression again assumes a low variance.

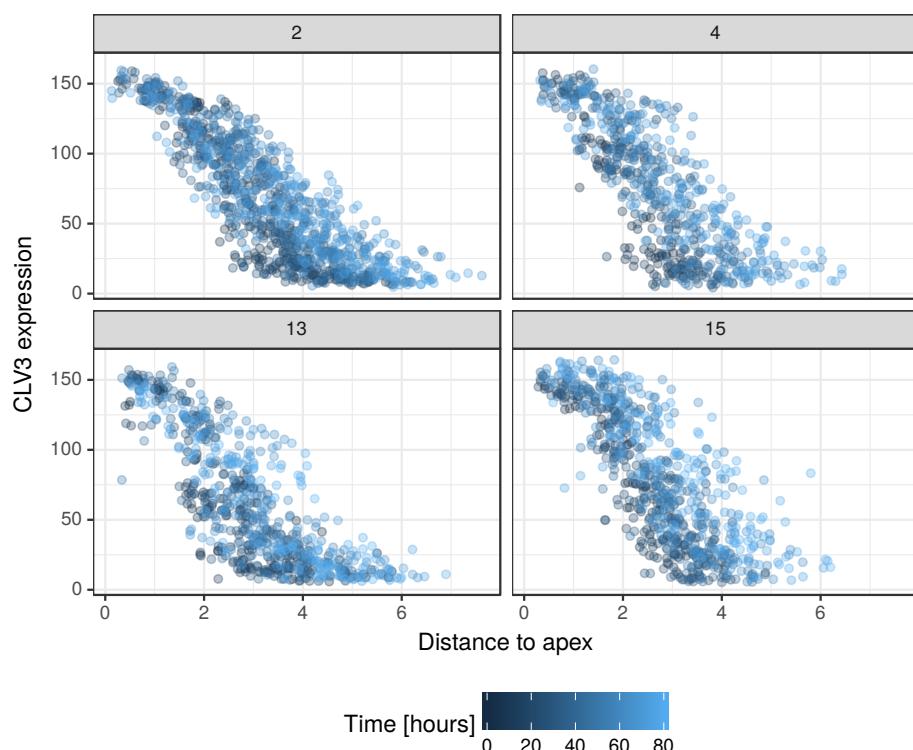


Fig. 3.4 Add caption

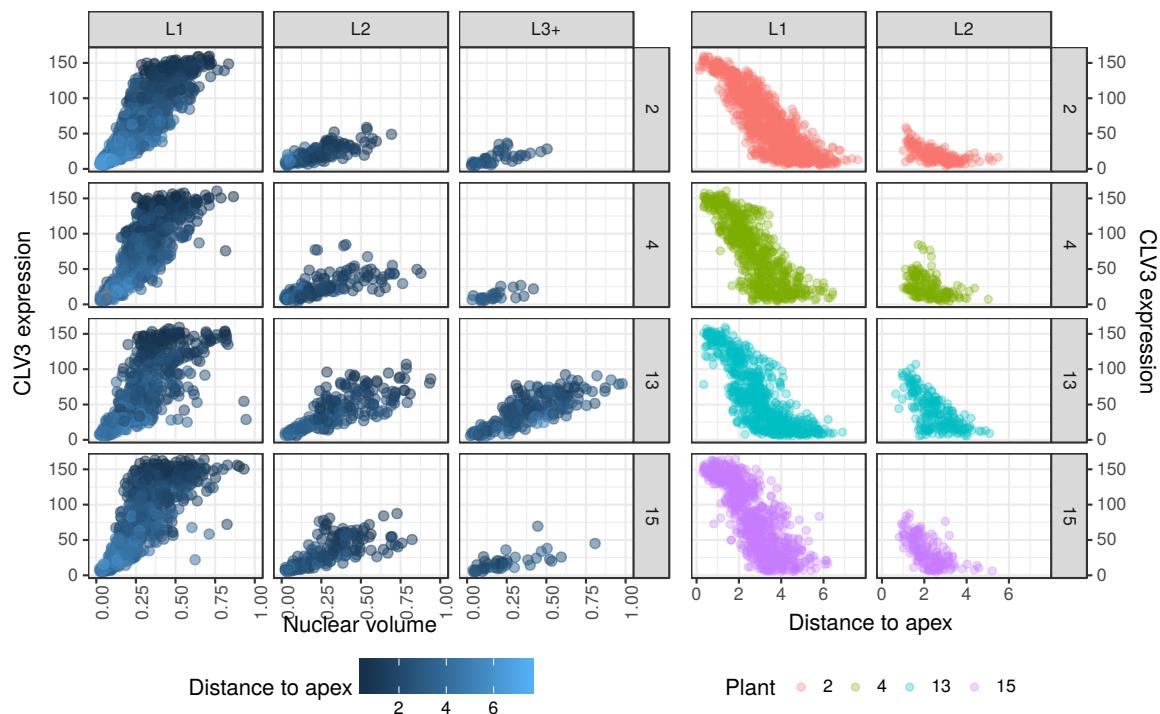


Fig. 3.5 Add caption

3.3 Distribution shifting suggests epidermal regulation

3.3.1 CLV3 is induced in the epidermis

A layer-wise separation of the CLV3 expression can be seen in particular between L1 and L2, as visualised in ??, where L3 has been excluded to the low quality in the nuclear segmentation (see ??).

3.4 Functional clustering implies dsRED technical artefacts

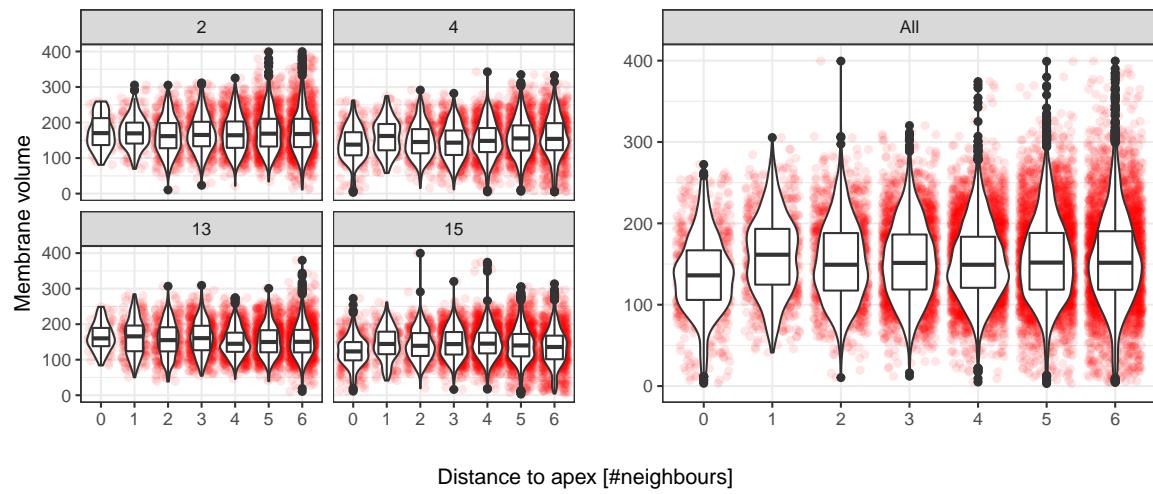


Fig. 3.6 Add caption

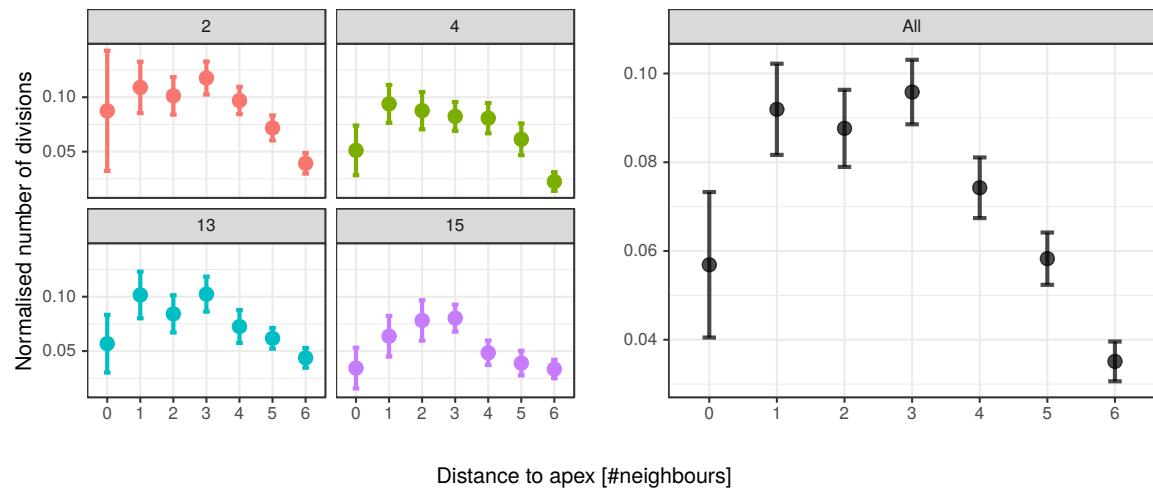


Fig. 3.7 Add caption

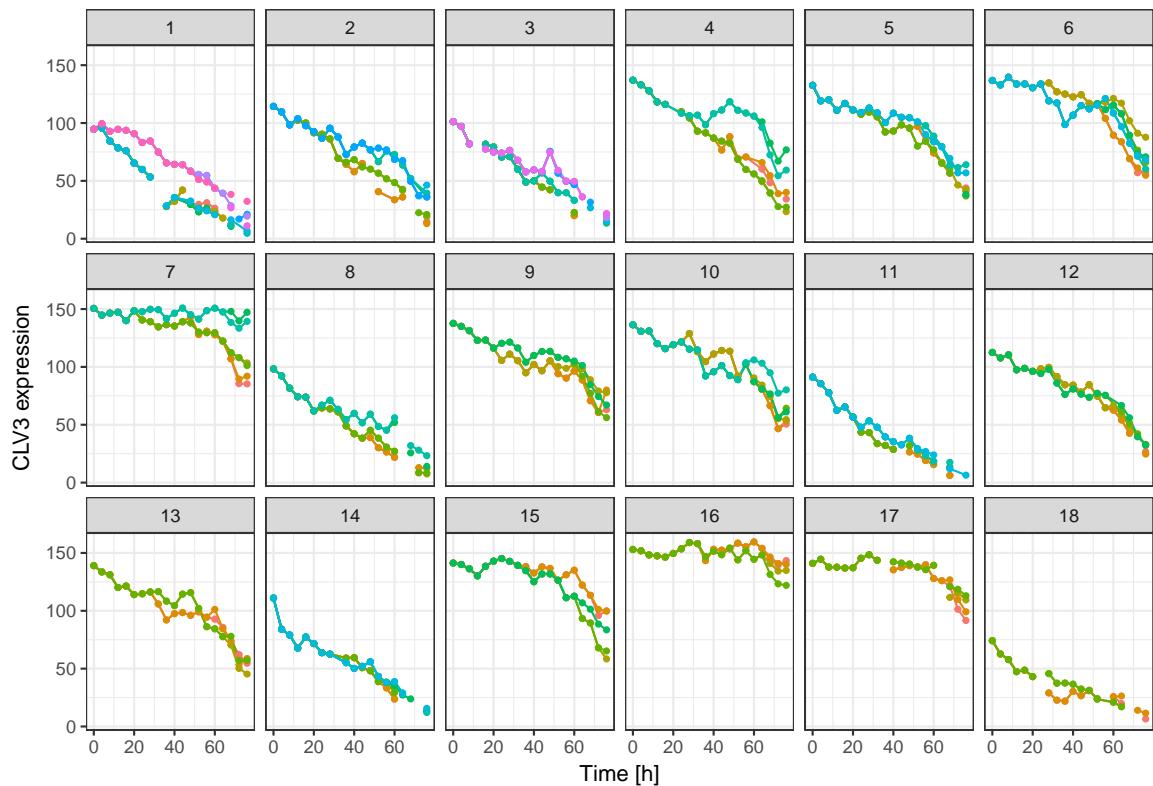


Fig. 3.8 Add caption

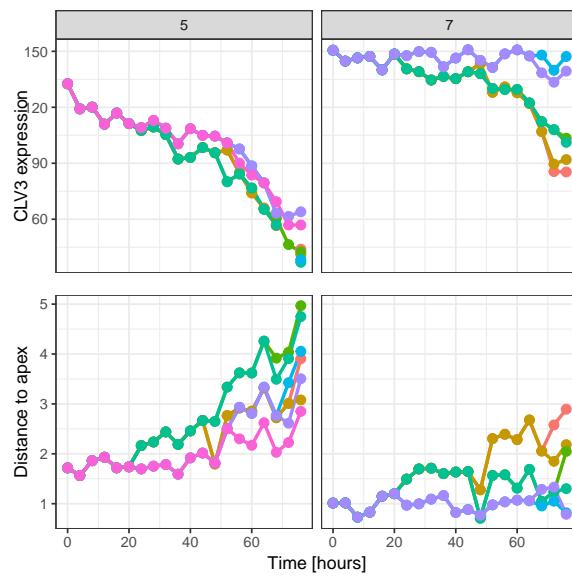


Fig. 3.9 Add caption.

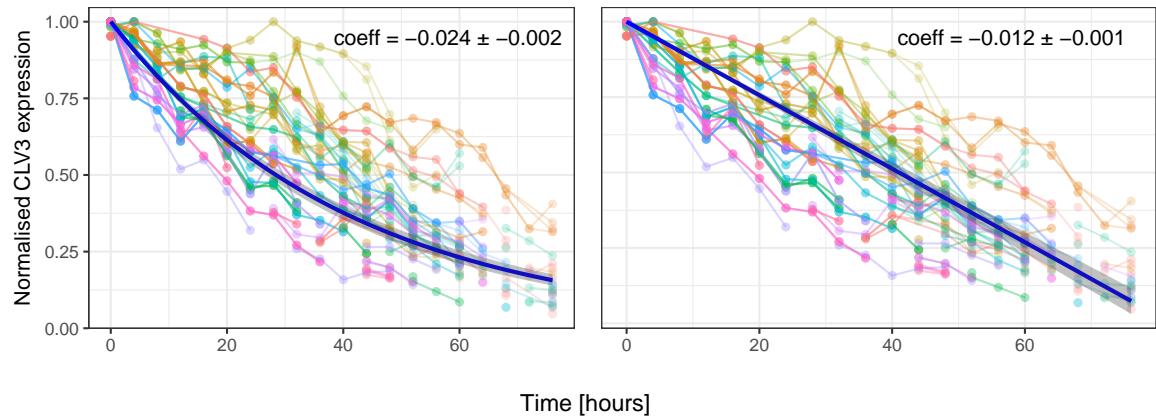


Fig. 3.10 Add caption

TODO: Justification for decay. Assume movement out of CZ proportional to division rate k_{div} , such that $\dot{R} \propto k_{div}$. We thus have $R = kt$. Assume linear degradation of CLV3 such that $\dot{C} = -k_{deg} = \frac{dC}{dt} = \frac{dC}{dR} \frac{dR}{dt}$. We therefore have $\frac{dC}{dR} = \frac{dC}{dt} = -\frac{k_{deg}}{k_{div}}$. Justify decay rate observed in trajectories or smth like that?

3.5 Deep-tissue cells divide at a slower rate

Still unknown.

Chapter 4

Discussion

4.1 NPA dilution induces oscillations

Henrik makes it sound like we actually see oscillations? Might be better to be a bit conservative here.

4.2 *In vivo* tracking of apical stem cells agrees with hypothesis

4.3 Distributional cues give hints at regulatory scheme

4.3.1 Activation of CLV3

Might be a good place to discuss errors and why we don't see stem cells in L2.

4.3.2 Longevity (Don't know what this shows yet)

Spatial cues from division data

4.3.3 Definition of center

Peak intensity in CLV3 follows general direction of growth? Reasons to doubt this?

Chapter 5

Outlook

5.1 Model expansion

Epidermal regulation
Cell size at CZ + division rate
Longevity stuff

5.2 Data investigation

Quality reassessment
PIN1 data
Control for feedback – wildtype?
dsRED decay time. Might be time to do some simple analysis?

5.3 New experiments

CLV3-WUS

References

Appendix A

Addition Information on Data Acquisition

A.1 Experimental procedure

The Yellow Fluorescent Protein (YFP) marker for the plasma membrane was amplified using PCR with primers attb1-mYfwd (5'-AGAAAGCTGGTTACTTGTACAGCTCGTCCATGCCGAGAGTG) and attb2-YFPrev (5'-AGAAAGCTGGTTACTTGTACAGCTCGTCCATGCCGAGAGTG), with the forward primer sequence containing a motif known to acetylate in plant cells CITEHERE. 50 μ L solution was amplified in 96 °C for 1 minute, followed by 25 cycles of 96 °C for 30 seconds, and a final elongation for 30 seconds. 5 μ L of the result was then used in a second reaction consisting of 40 μ L solution in total, with primers B1 adapt (5'-GGGGACAAGTTGTACAAAAAAGCAGGCT) and B2 adapt (5'-GGGGACCACTTGTACAAGAAAGCTGGGT) included. Similar to the first solution, the second mixture was amplified by PCR in 95 °C for 2 minutes, followed by 94 °C for 30 seconds, 48 °C for 30 seconds, and 72 °C for 1 minute, 20 cycles of 94 °C for 30 seconds, 55 °C for 30 seconds, and 72 °C for 1 minute. Finally, elongation took place under 72 °C for 1 minute.

Membrane

Nuclei

Unnused data (PIN)

A.2 Segmentation

A.2.1 Settings and approach

A.3 Errors in data

Segmentation Errors

Tracking errors

Manual tracking in some timepoints...

A.4 Layer Quality Assessment

Appendix B

Extended data analysis

B.1 NPA dilution is revealed by primordia initiation

B.2 Just cram in the rest of the data

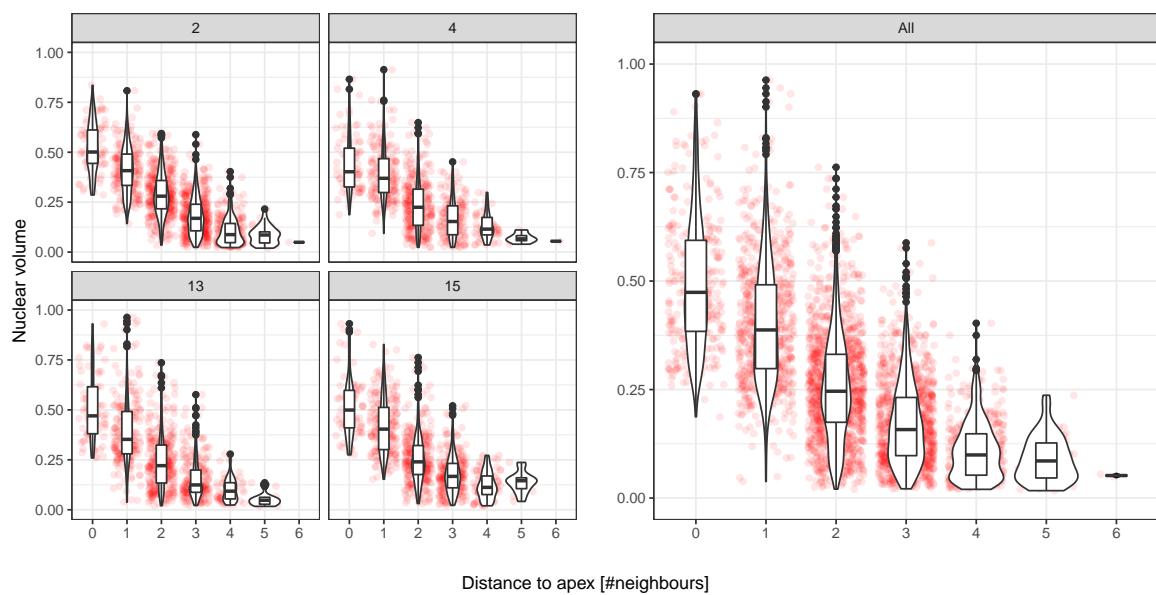


Fig. B.1 Add caption

Appendix C

Software Descriptions

C.1 Costanza

Costanza (COnfocal STack ANalyZer Application) **ADD CITATION** 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. **ADD CITATION**

C.3 Organism

C.3.1 Main software

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. **ADD CITATION**

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

Add flowchart

Something about Max's software, or include that in TissueViewer?