Imperial College London

Department of Life Sciences

Understanding the core gene regulatory network underlying epidermal stem cell patterning in *Caenorhabditis elegans*

Alicja Brożek

Submitted in fulfilment of the requirements

for the degree of Doctor of Philosophy at

Imperial College London, April 2024

I hereby declare that this thesis and the work reported herein was composed by and originated entirely from me. Information derived from the published and unpublished work of others has been acknowledged in the text and references are given in the list of sources.

Alicja Brozek (2024)



The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International Licence (CC BY-NC-ND).

Under this licence, you may copy and redistribute the material in any medium or format on the condition that; you credit the author, do not use it for commercial purposes and do not distribute modified versions of the work.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

Despite the breadth of knowledge on *C. elegans* and its various cells noone has yet described a Gene Regulatory Network (GRN) underlying the patterning of *C. elegans*'s hypodermal stem cells, the seam cells. In this thesis we aim to produce a seam cell GRN with several different approaches. We start by focusing on three core transcription factors: *elt-1*, *egl-18*, and *ceh-16*. We construct an initial network at the late L1 stage, a time point where no cell is actively dividing. To do this we combine mathematical and biological approaches to construct a minimal GRN. We test our hypotheses informed by this model on a double mutant of two of the two core genes *ceh-16(bp323);egl-18(ga97)* and in the process discover a highly robust, but erroneous seam cell phenotype at the end of larval development. We further evaluate the network over different time points and delve into the minutiae of the differences in expression patterns between the 16 seam cells at various time points. We test their sensitivities and uncover differences between the seam cells. We identify key positional patterns in the sensitivities of the seam cells to changes in their core network and link these sensitivities to Wnt signalling. Finally, we dissect the *ceh-16(bp323);egl-18(ga97)* double mutant to further elucidate the mechanisms behind its highly robust loss of self-renewal divisions. Overall in this thesis we incorporate mathematical approaches into the standard genetics based approaches for GRN inference. We also demonstrate the importance of maintaining a spatial and temporal resolution when studying the epidermal stem cells of *C. elegans*.

Acknowledgements

My PhD has been an exceptionally rewarding experience and has taught me to push past my comfort zone, to persevere, and to have fun with science. I don't think it would have been possible without my two exceptional supervisors. They have given me the independence I needed to explore new avenues while making sure to support me every time things didn't work or got difficult. Michalis taught me to express my ideas but also to filter them for those that will be relevant to the bigger questions and those that won't be, and I am a much better scientist thanks to him. Vahid introduced me to computational biology and helped me gain confidence in more theoretical approaches. They have both been an immense support both scientifically and emotionally throughout this PhD and I couldn't ask for better supervisors.

I am also particularly grateful to post doc Mark Hintze for teaching me most of my lab skills and that there are no stupid questions but also for helping me believe in myself even when I struggled to.

I'd like to say a huge thank you to my present and past coworkers: Manish, Domenica, Franzi, Ken, Ming, Jon, Florence, Michael, Andreas, Dimitris V. and particularly the members of the seam cell group Dimitris K. and Amelia, with a special thanks to Mar who was with me every step of the way, from the first look down the microscope to the last smFISH, and who taught me the beauty of good organisation. I swear I'll be there for any Nikon problems you will ever have.

I'd like to thank undergraduate and masters students Xiaotong, Vincent, Jingyi, and particularly Arianna for their help with my research and for scientific discussions.

I'd also like to express my immense gratitude the QuBi team at AstraZeneca with whom I worked during my PIPS. Thank you all for making me feel like an important part of the team despite only working there for 3 months, and for allowing me to learn from all of you. In particular I'd like to thank Sergio and Ben who were my supervisors and who taught me so much about drug development research, but also about team building and cross-disciplinary work. It was truly wonderful to work with you.

I'm so grateful for all my friends who supported me through these four years. Bella and Lili for being my rocks through any storm and being there even when all the worms died and none of the results were making sense but also for reminding me to celebrate the wins and the good moments. I'd like to thank Carlotta for complimenting late lab nights with dinner, tv and good conversation and for making sure my head is always screwed on tight, but also for

housing me for 3 months during a personal housing crisis. Going to miss our 6am coffee chats and your amazing soups. I'm grateful for Livia for the long mood-lifting walks and deep conversations by the Thames. I thank Eleonora for always being there with hot tea, a strong opinion and a sharp outfit and Will for always being there to help lift anything, from boxes to spirits. A huge thank you to Alec for being there for me since the first year of uni and for always providing a comedic take on my life. I'd also like to thank Alex and Emily for feeding me meals that were more delicious than most restaurants I've been to and Eloise and Max for providing a home away from home in Manchester.

I'd like to thank John for reminding me that cookies are not a meal and that coffee is not a replacement for water and for writing down the seam cell counts for north of 2000 worms while sitting in a dark room with me on weekends.

I'd also like to thank my grandma who I could always call to talk about anything under the sun and who was always an inspiration to me and my aunt who always kept me grounded in reality. I'd like to thank my siblings, in particular Marius and Annette who let me run away from my problems to Norway, and Mati who never let me take myself seriously.

I couldn't have done this without my parents, who kept me focused and were always there for me. They taught me to aim high and to know that if I fell they'd be there to catch me. Their unconditional love and belief that I was a scientist were one of the main reasons I started this journey.

Let's open up this bag of worms

List of Abbreviations

gene name - mRNA

GENE NAME - protein

GRN - **G**ene **R**egulatory **N**etwork

TF - **T**ranscription **F**actor

RNAi - **R**ibo **N**ucleic **A**cid **i**nterference

LB broth - **L**uria **B**ertani broth

smFISH - **s**ingle **m**olecule **F**luorescent **I**n **S**itu **H**ybridisation

mRNA - **m**essenger **R**ibo **N**ucleic **A**cid

CRISPR - **C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats

sgRNA - **s**ingle **g**uide **R**ibo**N**ucleic **A**cid

PCR - **P**olymerase **C**hain **R**eaction

MRA - **M**odular **R**esponse **A**nalysis

ODE - **O**rdinary **D**ifferential **E**quation

MCMC - **M**arkov **C**hain **M**onte **C**arlo

SLING - **S**parse **L**ikelihood-free **I**nference usi**n**g **G**ibbs

SCM - **S**eam **C**ell **M**arker

GFP - **G**reen **F**luorescent **P**rotein

dsRNA - **d**ouble **s**tranded **R**ibo**N**ucleic **Acid**

NGM - **N**ematode **G**rowth **M**edium

IPTG - **I**so**P**ropyl β-D-1-**T**hio**G**alactopyranoside

WBA - **W**nt β-catenin **A**symmetry

Contents

Abstract **4**

Acknowledgements **5**

1 General Introduction **19**

1.1 Introduction to Gene Regulatory Networks (GRNs) 19

1.1.1 Boolean models 22

1.1.2 Modular response analysis (MRA) 24

1.1.3 Ordinary Differential Equations (ODEs) 25

1.1.4 Bayesian inference 26

1.1.5 GRNs in stem cells 28

1.2 Introduction to *C. elegans* 31

1.2.1 *C. elegans* development 31

1.2.2 *C. elegans* epidermis 33

1.2.3 *C. elegans* and stem cells 35

1.2.4 Gene regulatory mechanisms of seam cell patterning 39

1.2.5 *Wnt* regulatory system in C. elegans 44

1.3 Aims 46

2 Methods **49**

2.1 *C.elegans* specific methods 49

2.1.1 Maintenance 49

2.1.2 Freezing 49

2.1.3 RNA interference (RNAi) by feeding 51

2.1.4 Genetics 52

2.1.5 Temperature treatment of animals . . . 52
2.2 Microscopy and Image analysis . . . 52
2.2.1 Seam cell counts . . . 52
2.2.2 Lineaging seam cells . . . 53
2.2.3 Single molecule mRNA fluorescent in situ hybridisation (smFISH) . . . 54
2.2.4 CRISPR-Cas9 . . . 54
2.2.5 Worm lysis and PCR . . . 57
2.3 Mathematical methods . . . 57
2.3.1 Statistics . . . 57
2.3.2 MRA . . . 59
2.3.3 ODE . . . 59
2.3.4 SLING . . . 60

3 A quantitative Gene Regulatory Network underlying seam cell patterning 61
3.1 Introduction . . . 61
3.2 Results . . . 63
3.2.1 Gene expression analysis using smFISH data identifies new interactions between core seam cell TFs . . . 63
3.2.2 Constructing a gene regulatory network using Modular Response Analysis 64
3.2.3 Using the MRA results to infer a set of ODEs to explain expression patterns of the core seam cell genes . . . 66
3.2.4 Mathematical testing of the newly developed GRN . . . 71
3.2.5 Biological validation of new model using the *ceh-16(bp323);egl-18(ga97)* mutant . . . 73
3.2.6 Evaluating our model in a protein context . . . 77
3.3 Discussion . . . 80

4 Variations in the seam cell GRN across time and space 84
4.1 Introduction . . . 84
4.2 Results . . . 86
4.2.1 Constructing the seam cell gene network in the early L1 stage . . . 86
4.2.2 Constructing the seam cell gene network at the L2 symmetric seam cell division . . . 91

4.2.3 Constructing the seam cell gene network at the L2 asymmetric seam cell division . . . 98
4.2.4 The core seam cell genes follow an L2 centric expression pattern over time . . . 104
4.2.5 Significant differences exist between gene expression in different seam cells at various time-points . . . 106
4.2.6 Understanding how expression differences in the core genes impact different seam cell sensitivity . . . 116
4.3 Discussion . . . 126
4.3.1 The layout of the core seam cell GRN changes over time . . . 126
4.3.2 The frequencies of seam cell losses and gains are correlated to core gene expression patterns . . . 130
4.3.3 The positional effect of *elt-1* is modulated by Wnt . . . 131

5 Understanding the robust seam cell number phenotype caused by the combined loss of *ceh-16* and *egl-18* **133**
5.1 Introduction . . . 133
5.2 Results . . . 134
5.2.1 Generating and characterising a seam cell specific *ceh-16* mutant . . . 134
5.2.2 The *ceh-16(bp232);egl-18(ga97)* is more phenotypically resistant to both environmental and genetic perturbations than *ceh-16*$_{LoxP}$ . . . 136
5.2.3 The double mutant converts the L2 symmetric division to an asymmetric division . . . 144
5.2.4 Investigating possible mechanisms of symmetry loss through mutations in both *ceh-16* and *egl-18* . . . 151
5.3 Discussion . . . 158
5.3.1 Robustness of the double mutant . . . 158
5.3.2 Seam cells lose capacity to divide symmetrically at L2 in *ceh-16(bp323);egl-18(ga97)* and *ceh-16*$_{LoxP}$ . . . 161
5.3.3 *ceh-16* and *egl-18* both influence and are influenced by genes known to be involved in division symmetry . . . 161

6 General Discussion **166**

6.1 The new quantitative gene regulatory network underlying core seam cell genes at early L1 166
6.2 The seam cell GRN changes over time and position 168
6.3 The complete loss of *egl-18* coupled with a partial loss of *ceh-16* causes a robust loss of symmetry 170
6.4 The core seam cell gene network regulates essential genes involved with asymmetry 172

7 Appendix **200**

List of Tables

2.1 **List of all strains used** 50
2.2 **List of all RNAi used** 51
2.3 **List of all smFISH probes used** 55
2.4 **List of all primers, sgRNA and repair templates used** 56

7.1 **Additional description of the various ODE models used: A)** Parameter names and definition. **B)** Description of the AND/OR logic of different models tested based on the initial MRA. 202

List of Figures

1.1 **Diagram showing a model of a GRN** 20
1.2 **A brief summary of the principles of Modular Response Analysis** 24
1.3 **The stem cell GRN in human stem cells** 30
1.4 **The comparison between the structure of the *C. elegans* epidermis and the mammalian epidermis** 34
1.5 **Illustration of the epidermal syncytia of *C. elegans*** 35
1.6 **Diagram showing the division patterns of the seam cells** 36
1.7 **Diagram showing canonical Wnt signalling components** 44

2.1 **Description of how seam cell-specific abnormalities are identified** 53
2.2 **Diagram showing the layout of the LoxP insertions in *ceh-16*** 58

3.1 **Phenotypic and gene expression analysis in wild type and TF mutant backgrounds** 65
3.2 **Using MRA to construct a seam cell gene network** 67
3.3 **Model fitting using SLING** 72
3.4 **Testing additional conditions for SLING** 74
3.5 **Validating model predictions using the *ceh-16(bp323);egl-18(ga97)* double mutant** 78
3.6 **Core seam cell genes' protein fusions' relative fluorescence in RNAi treated animals** 79

4.1 ***elt-1* expression 10 hours after bleaching in wild type, *ceh-16*, *egl-18* and *elt-1* mutants** 87
4.2 ***egl-18* expression 10 hours after bleaching in *ceh-16*, *egl-18* and *elt-1* mutants** 88

4.3 ***ceh-16* expression 10 hours after bleaching in wild type, *ceh-16*, *egl-18* and *elt-1* mutants** . . . 89
4.4 **Strengths of interactions between core genes in early L1 calculated using MRA** . . . 90
4.5 ***elt-1* expression 26 hours after bleaching, in the symmetric daughters in wild type, *ceh-16*, *egl-18* and *elt-1* mutants** . . . 93
4.6 ***egl-18* expression 26 hours after bleaching, in the symmetric daughters in wild type, *ceh-16*, *egl-18* and *elt-1* mutants** . . . 94
4.7 ***ceh-16* expression 26 hours after bleaching, in the symmetric daughters in wild type, *ceh-16*, *egl-18* and *elt-1* mutants** . . . 96
4.8 **Strengths of interactions between core genes in the symmetric L2 daughters calculated using MRA** . . . 99
4.9 ***elt-1* expression 26 hours after bleaching, in the asymmetric daughters in wild type, *ceh-16*, *egl-18* and *elt-1* mutants** . . . 101
4.10 ***egl-18* expression 26 hours after bleaching, in the asymmetric daughters in wild type, *ceh-16*, *egl-18* and *elt-1* mutants** . . . 102
4.11 ***ceh-16* expression 26 hours after bleaching, in the asymmetric daughters in wild type, *ceh-16*, *egl-18* and *elt-1* mutants** . . . 103
4.12 **Strengths of interactions between core genes in L2 asymmetric daughters calculated using MRA** . . . 105
4.13 **Core gene expression throughout early larval development** . . . 107
4.14 **At early L1 seam cell core gene expression shows similarities and differences between groups of seam cells** . . . 108
4.15 **At early L2 seam cell core gene expression shows similarities and differences between groups of seam cells** . . . 111
4.16 **At late L2 seam cell core gene expression shows similarities and differences between groups of seam cells** . . . 115
4.17 **Changes in seam cell sensitivities caused by different dilutions of *elt-1* RNAi bacteria** . . . 120
4.18 **Changes in seam cell sensitivities caused by different dilutions of *egl-18* RNAi bacteria** . . . 123

4.19 **Seam cell counts in *Hox* and *Wnt* mutants upon treatment with 50% *elt-1* RNAi** 124
4.20 **The *elt-1*-specific seam cell sensitivities are disrupted in *Wnt* but not *Hox* mutants** 125
4.21 **A diagram illustrating quantitative GRNs for the core seam cell genes in early larval stages** 127

5.1 **Impact of temperature on *ceh-16(bp323);egl-18(ga97)* and *ceh-16*$_{LoxP}$** . . 136
5.2 **The *ceh-16(bp323);egl-18(ga97)* mutant and the *ceh-16*$_{LoxP}$ show frequent losses in anterior cells, and gains in tail cells** 138
5.3 **The core gene RNAi on *egl-18* and *ceh-16* mutants showcases the robustness of *ceh-16(bp323);egl-18(ga97)* in the face of genetic perturbations** 141
5.4 ***eff-1* distribuition in the double mutant and *ceh-16*$_{LoxP}$ 26 hours after bleaching** 146
5.5 **Impact of symmetry regulators on the loss of symmetry in the double mutant and parental strains** 149
5.6 **Survival of *lit-1* RNAi** 150
5.7 **Change in *pop-1* and *rnt-1* levels between wild type and *ceh-16(bp323);egl-18(ga97)* double mutant and *ceh-16*$_{LoxP}$ at the L2 symmetric division** . . 152
5.8 **Change in *efl-3*, *lit-1* and *apr-1* levels between wild type, *ceh-16(bp323);egl-18(ga97)* double mutant and *ceh-16*$_{LoxP}$ at the L2 symmetric division** . . 154
5.9 **Change in *ceh-16* and *egl-18* levels between wild type and *apr-1(bp298)* at the L2 symmetric division** 157

7.1 **P-values of the comparisons between the late L1 seam cells' gene expression patterns** 201
7.2 **P-values of the comparisons between the L2 asymmetric daughter seam cells' gene expression patterns including the differentiating cells** 203
7.3 **P-values of the comparison of each cell's frequency of abnormal counts with each other cell's frequency of abnormal counts at increasing concentration of *elt-1* and *egl-18* RNAi** 204
7.4 **P-values comparing the mean number of cells in wild type and Hox and Wnt mutant strains, with and without 50% *elt-1* RNAi treatment** 205

7.5 **P-values of the comparison of each cell's frequency of abnormal counts with each other cell's frequency of abnormal counts in wild type and Hox mutants *(egl-5(n945), lin-39(n1880) and lin-39(n1760);mab-5(e1239);egl-5(n945))* treated with the empty RNAi vector (ht115) and 50% *elt-1* RNAi.** 206

7.6 **P-values of the comparison of each cell's frequency of abnormal counts with each other cell's frequency of abnormal counts in wild type and Wnt mutants *(bar-1(ga80), egl-20(hu105) and lin-17(n671))* treated with the empty RNAi vector (ht115) and 50% *elt-1* RNAi** 207

7.7 **P-values between cells of *wild type*, *ceh-16(bp323)*, *egl-18(ga97)*, *ceh-16(bp323);egl-18(ga97)* and *ceh-16*$_{LoxP}$ at 15℃, 20℃ and 25℃** 208

Chapter 1

General Introduction

1.1 Introduction to Gene Regulatory Networks (GRNs)

All genes within any species' genome are interconnected through a web known as a gene regulatory network (GRN) (Alon, 2019; Goodwin, 1963). These networks are immensely complex, posing challenges for scientists across disciplines, from biology to physics. A GRN comprises molecular players and their interactions, governing gene expression in terms of timing, location, and quantity (Davidson and Levin, 2005). Precise spatiotemporal patterns of gene expression are crucial for correct organismal development (Peter and Davidson, 2011). As such GRNs must be able to act with extreme precision. There are many regulatory genetic modules that integrate thousands of signals, ranging from internal signals from regulatory proteins (activators and repressors) as well as small chemical molecules, and external signals such as temperature and access to food (Davidson, 2010). The output of these modules is precise transcriptional control. Downstream of these modules, there are other regulatory modules (Figure 1.1A), which contribute to the creation of the complex regulatory system that allows the control of the majority of biological functions (Alon, 2019) (Figure 1.1A-B). As experimental methods progressed, it became possible to acquire more and more large-scale data on the links between genes in the network. Nevertheless, the scale of the problem meant that manual analysis was often insufficient in the complete understanding of the mechanisms and as such computational methods had to be integrated in the study of GRNs.

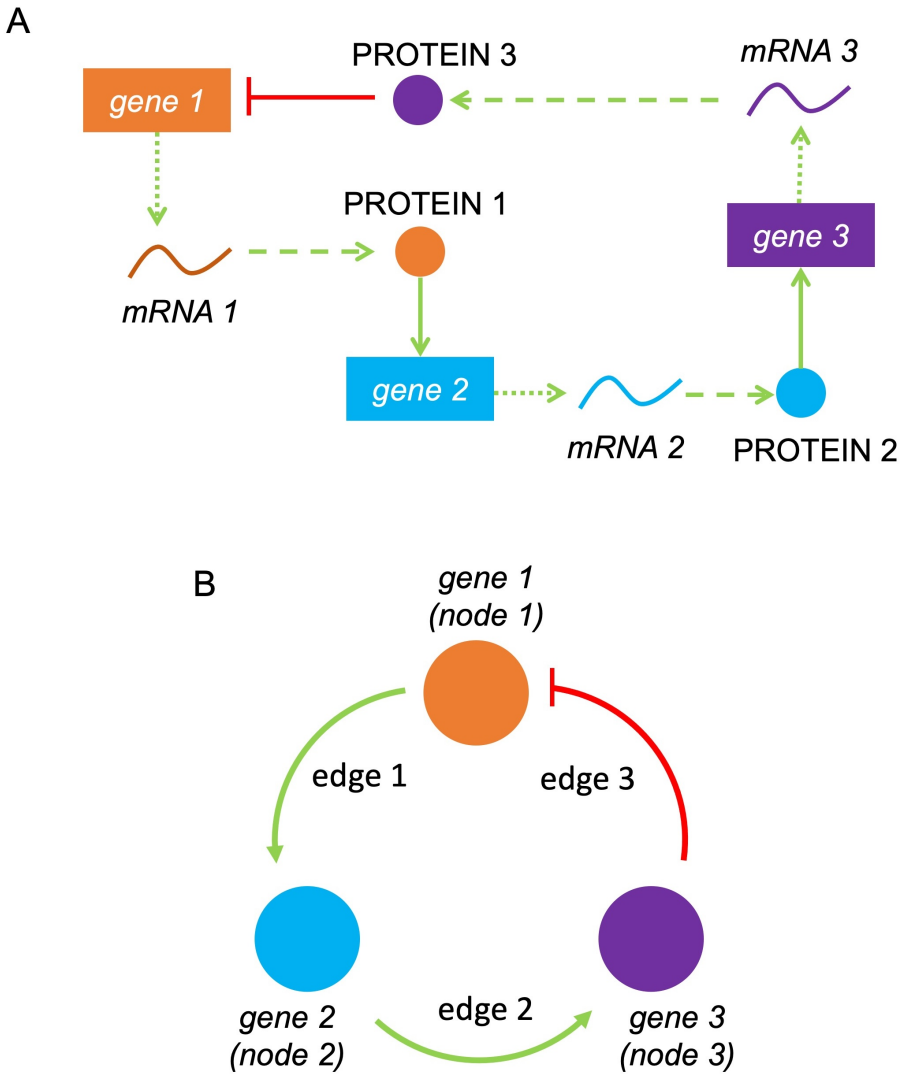


Figure 1.1: **Diagram showing a model of a GRN: A)** A diagram of a GRN model showing the regulatory mechanisms between three TFs. Dotted arrows denote transcription, dashed arrows translation and block line arrows show activation or repression of the transcription of the gene. Activation is shown using a green arrow and repression using a red blunt end line. **B)** The same diagram as shown in A but simplified to a format commonly used in GRN representation.

GRNs can be inferred at various levels of detail, focusing on either the identification or a more detailed description of links, often referred to as edges, between specific genes, known as nodes (Figure 1.1B). Edges denote both direct and indirect interactions between nodes, which encompass genes as well as coding and non-coding regulatory elements, such as enhancers or promoters. The most fundamental methods of gene regulatory network inference rely on correlating gene expressions. Experimental techniques like microarrays and RNA-seq, are used to inform these network models and jump-start large-scale network inference (Montenegro, 2022; Jupiter et al., 2009). These methods facilitate the high-throughput inference of links between individual genes within the network. Recent advancements have even enabled the use of these correlation based techniques to infer link directionality (Khosravi et al., 2015). Despite this, correlation data does not allow for the inference of causality (Schäfer and Strimmer, 2005). Causality in this context means a direct relationship between one gene and another where the expression of one has a direct influence on the other. In order to look at the causality of a certain gene's changes in expression one must perturb the network (Hughes et al., 2000). These perturbations normally involve physically disrupting the function and/or expression of one of the genes in the network and then observing the impact of this change on other genes in the system. Experimental perturbations and measurements of their effect on a specific gene allow the classification of interactions as positive or negative as well as identifying feedback loops and even giving a quantitative prediction of the changes in gene product based on the changes in another gene. It allows the determination of the architecture of a GRN with detailed directed edges (Gardner and Faith, 2005). However, there are situations where introducing perturbations can be challenging or impossible. As such it is useful to understand the tools available for GRN inference from purely correlation based data to more complex methods (Faith et al., 2007).

The simplest inference methods are correlation based methods (Delgado and Gómez-Vela, 2019), which predominantly rely on co-expression data to establish gene-gene interactions based on expression dependence levels and do not aim to determine the causality of the interactions. The choice of thresholds for establishing dependence changes the sparsity of the network. There are many measures to establish dependence levels, most commonly

correlation coefficients like Kendall or Pearson, however mutual information (the amount of information obtained about one variable by observing another variable) or euclidean distances (the smallest distance between the pairs of points within datasets) can also be utilised. Due to the increased ease and decreases in costs of carrying out RNA-seq, the data available for correlation-based networks has expanded, providing insights into the transcriptomic behavior of organisms under many different conditions and enabling the evaluation of millions of genes simultaneously. These networks also boast a low computational costs, leading to the development of many methods, including REVEAL (Liang et al., 1998), ARCANE (Margolin et al., 2006), and MRNET (Olsen et al., 2008) tailored for constructing GRNs using correlation and by assessing dependence levels. Some methods rely on the directed information metric (DTI) (Rao et al., 2007) which has been applied to infer causality from scRNA-seq data (Qiu et al., 2020). Nevertheless inferring causality, using this method, required time variable data, posing experimental challenges, particularly for the collection of single-cell data. Furthermore Gaussian graphical models (GGMs) have been employed for the identification of conditional links within GRNs and distinguishing between direct and indirect links (Xie et al., 2016). Another computational method that has managed to bypass the need for perturbations to extract causal relationships is the MIIC algorithm (Verny et al., 2017). However this algorithm relied on single cell approaches that targeted multiple genes in the same cell.

Overall correlation-based networks have low complexity and are based on large-scale data, much of which is readily available. These methods are particularly valuable for constructing large, multi-node networks. However, they require further data to infer causality and are not inherently suited for detailed quantitative network development or dynamic system predictions (Margolin et al., 2006). Once causality is established within a connection, one can advance to modelling at a higher complexity level by incorporating dynamic information (Davidson and Levin, 2005; Gardner et al., 2003).

1.1.1 Boolean models

Boolean models are one of the simplest approaches for capturing dynamic information within (GRNs). In standard Boolean gene network models, genes exist in two states: Active (1) or

Inactive (0). The transition of gene states at time point $(t+1)$ can be either stable or unstable and depend on the activity of other members within the GRN at time point (t). Stuart Kauffman first introduced this model, drawing an analogy between gene behavior and switches to model cell cycle and replication across various organisms (Kauffman, 1969, 1970). In these models, continuous parameters such as gene expression levels are discretised through thresholding. This means that a 0 value doesn't necessarily indicate a complete absence of expression but rather low expression. The conversion of expression data to binary values is the most complex step as it may lead to the loss of crucial nuances as genes often exhibit a continuum of expression levels rather than binary states. Despite this simplification, meaningful biological predictions can be derived from Boolean models (Shmulevich et al., 2002; Tabus et al., 2002). The advantage of Boolean models is their focus on general gene behavior rather than detailed quantitative biochemical relationships. This simplifies data collection processes and reduces the risk of over-fitting. While Boolean networks can extrapolate links from correlation-based data, they can also be adapted for more complex datasets containing causal information (Liang et al., 1998). Numerous algorithms utilize Boolean systems, including those presented by Maki et al. (2000), and Davidich and Bornholdt (2008), along with the Akutsu system (Akutsu et al., 1999) and the REVEAL algorithm (Liang et al., 1998), which despite being an information theory based model incorporates Boolean modelling in its mechanic. Boolean models rely on seven logical statements IF, AND, OR, NOT, NAND, NOR and XOR and are often illustrated in the same way as electrical systems. Boolean network dynamics are modelled through Markov Chains (Xiao, 2009) and can be used to infer interactions (Claussen et al., 2017). They can indicate steady states and illustrate strengths of different attractors. The development of probabilistic Boolean models has further enriched this modeling approach by incorporating variable dynamics into this simple framework (Malekpour et al., 2020; Shmulevich et al., 2002).

Nevertheless despite the expansion of the Boolean network toolbox these remain highly simplistic models and lack the ability to incorporate many of the details underlying a GRN. Another limitation of Boolean modelling is that while it does allow for the incorporation of causality, it fails to incorporate strengths of interactions. To address these limitations one

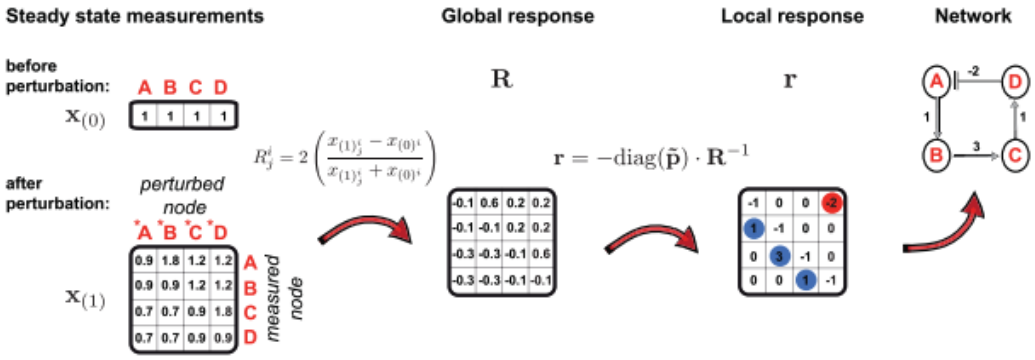


Figure 1.2: **A brief summary of the principles of Modular Response Analysis:** Outlining how the steady state quantities of gene expression are converted into interaction strengths. The quantities of the genes are compared between the unperturbed and the perturbed state under the context of each gene's independent perturbation. This is then converted through a jacobian matrix to a global response and the diagonals of the matrix showcase the strengths of the interactions from which a network can be inferred. The diagram was taken from Klinger and Blüthgen (2018)

can turn to techniques such as Modular Response Analysis (MRA) or Ordinary Differential Equations (ODEs).

1.1.2 Modular response analysis (MRA)

MRA was developed by Kholodenko et al. (2002). This approach decomposes the GRN into its individual components, evaluating how changes in each module affect the remaining components of the network (Kholodenko et al., 1999, 2002). By evaluating the impact of perturbations within each module and observing the response of the entire network, MRA infers quantitative relationships between genes. It integrates the strengths of individual perturbations and the network's overall response to predict the network's behavior. The MRA uses elements of metabolic control analysis (MCA) which was initially built for the understanding of enzymatic regulation and aims to establish quantitatively the degree of control that an enzyme has on concentrations of substrate and product (Moreno-Sánchez et al., 2008; Heinrich and Rapoport, 1974; Kacser, 1973). The method involves the formation of two Jacobian matrices. One that involves the specific quantities of each gene under each condition and the other the predicted strengths of interactions. It is summarised in Figure 1.2.

1.1.3 Ordinary Differential Equations (ODEs)

The discretisation problem found in Boolean models is circumvented by a modelling method known as Ordinary Differential Equations (ODEs). ODEs utilise continuous variables rather than discrete ones, however they rely on more detailed assumptions than a Boolean model. They describe the rate of change of a specific gene's expression as a function of factors such as transcription rate, degradation rate, gene expression of other factors and even environmental influences. This is mathematically represented by the equation:

$$\frac{dx_i}{dt} = f_i(x_1, x_2, ..., x_n, p, u) \tag{1.1}$$

Here x_i indicates the expression level of a gene at a specific time,n is the number of genes, p represents system parameters and u denotes external perturbation. The each gene x is commonly modelled using Hill equations. In the context of GRNs the Hill equation represents the relationship between the concentration of TFs and the transcription rate of a gene. Hence if we assume that the gene x_i's expression is dependent only on gene x_j, this would be mathematically expressed as:

$$\frac{dx_i}{dt} = K\frac{x_j^n}{{K_D}^n + x_j^n} - dx_i \tag{1.2}$$

Here x_i represents the gene expression in question, x_j denotes the concentration of TF J, n is the Hill coefficient and illustrates binding cooperativity, and K_D symbolises the affinity of a TF for a specific regulatory region, dx_i symbolises the degradation rate of gene x_i and K symbolises the basal transcription rate of gene x_i (Hill, 1910; Monod et al., 1965; Koshland Jr et al., 1966).

Quantifying these parameters typically necessitates either specific knowledge of biochemical properties or educated guesses. ODEs serve as a valuable tool in elucidating the strengths of various interactions and building GRNs (Matsumoto et al., 2017; Deng et al.,

2017). However, they may struggle with large GRNs containing numerous links and nodes. Moreover, if the parameters are uncertain, each ODE provides only a single numerical solution, thus, ODEs are often paired with sampling schemes, such as Bayesian inference, to address parameter uncertainties.

1.1.4 Bayesian inference

In the simplest of descriptions, Bayesian inference involves counting the number of ways an event could occur based on specific assumptions, relying on probability to determine the plausibility of said event. The foundation of Bayesian inference lies in Bayes' theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{1.3}$$

There are several terms that need to be defined to explain Bayes theorem. The first is likelihood. Likelihood is the relative number of ways that a specific conjecture could produce the observable outcome. This is shown as $P(B|A)$ in Bayes theorem and explains the probability of the observation given a specific conjecture. The next term to define is parameter. Parameters are events that would explain the behaviour of the data and are basically conjectures or hypotheses. In this equation the parameter is represented by A. Estimating the likelihood of each hypothesis or parameter enables us to assess which parameter is most probable and its likelihood of occurrence. This conditional probability is known as the posterior and is the probability of observations B given the conjecture A is true and in the equation is denoted as $P(A|B)$. However, estimating any of these probabilities necessitates knowledge of the priors. Priors are the probabilities of each parameter A occurring but also the overall probability of observations B.

Bayesian inference, may imply a form of causality through probabilistic dependency, enabling the modeling of uncertainty and variability in causal relationships. However this relies on the data fed into the model to be perturbational and thus have direct causality. Understanding causal relationships is pivotal for unraveling the mechanisms of gene regulation, facilitating the development of detailed quantitative models that help circumvent confounding

factors (Friedman, 2004). Without discerning causal relationships, comprehending complex processes such as stem cell patterning becomes challenging, as they are a result of a cascade of events that cause downstream changes (Sachs et al., 2005; Vipin et al., 2018; Mooij et al., 2016).

The Bayesian theorem treats the $P(B|A)$ as a relative plausibility of an event occurring. Consequently in models such as an ODE, instead of specifying the individual parameters directly, one can present a confidence interval and thus through sampling obtain different outputs of the equation, with accompanying probabilities (Bayes, 1763; Pearson, 1925).

Most sampling methods are variations on the Monte Carlo Markov Chain (MCMC) method. This method is an iterative process which aims to match its samples to a target distribution at equilibrium. MCMC can be divided into 4 steps.

1) Start. Initiating the system's initial state.

2) Proposal. Proposing changes to the system.

3) Acceptance. Determining whether the proposed change improves upon the initial state.

4) Update. Updating the network based on the acceptance step.

The probability of each new state depends solely on the previous state, not any preceding ones. This process iterates until a sufficient number of samples matching the desired target distribution are obtained (Hastings, 1970). Samples can be updated along a specific probability, eliminating the need to assume a Gaussian or any other distribution shape. One of the earliest MCMC algorithms is the Metropolis algorithm (Metropolis et al., 1953), which assumes the probability of the change from a proposed value to a new updated value is the same as the reverse. However, this approach may struggle with uneven parameters, such as standard deviations, when boundaries at 0 exist. Asymmetric proposals are a notable feature of Gibbs sampling (Geman and Geman, 1984; Casella and George, 1992). In Gibbs sampling, the distribution of a proposed parameter adjusts based on the present parameter values. This adjustment relies on pairs of prior distributions and likelihoods for each parameter, referred to as conjugate pairs, which offer analytical solutions for posterior distributions.

Overall, numerous methods exist for studying Gene Regulatory Networks (GRNs), with the choice of method depending on the available information and prior knowledge of gene

behaviors. For the purpose of this thesis, our attention will be directed towards understanding GRNs in epidermal stem cells in *C. elegans*. However, to achieve this objective, it is essential to grasp the development of GRNs within the broader context of the stem cell field.

1.1.5 GRNs in stem cells

Stem cells are defined by two primary properties. The ability to self-renew and thus produce more cells of its own kind, and the ability to differentiate and produce new cell types (Boveri, 1888; Haeckel, 1876). The potency of a stem cell is determined by the range of tissues it can produce. Totipotent stem cells can generate all tissues in an organism, while pluripotent stem cells can produce most cell types except for a few critical ones essential for early developmental survival. Multipotent stem cells, on the other hand, generate cells specific to particular tissue types, such as epidermal stem cells, hematopoietic stem cells, and neural stem cells.

Even before the precise players involved in stem cell regulation were identified, early models of cell fate determination laid the groundwork for the GRN field. Waddington's concept of canalisation compared cell fate choices to a ball rolling down a hill (Waddington, 1961, 1957, 1942). He illustrated each cell fate as a valley on the landscape. Cells with higher potency, situated higher up on the hill, possess greater flexibility in their fate choices and are more susceptible to external signaling. In contrast, cells with determined fates, nestled deeper in the valleys, exhibit reduced susceptibility to external signals (Waddington, 1961). Some of the first Boolean models also considered stem cell fate as an application (Preisler and Kauffman, 1999) and conversely the earliest studies focused on the core stem cell genes relied on boolean modelling (Boyer et al., 2005). These early models were expanded upon by incorporating new nodes, through linkage to new pathways, and the addition of new edges, like the auto-inhibition of NANOG (Chickarmane et al., 2006; Navarro et al., 2012). New models also took into account spatial localisation of the proteins and gene products.

However the exploration of stem cell gene regulatory networks (GRNs) intensified with the identification of core stem cell factors. These were initially identified as OCT4, NANOG and SOX2 and KLF4 and C-MYC (Nichols et al., 1998; Mitsui et al., 2003; Avilion et al.,

2003). Interactions between OCT4 and SOX2 were observed (Chambers and Smith, 2004), however the construction of a regulatory network governing pluripotency was initiated through investigations into Nanog's regulation of OCT4 and SOX2 (Rodda et al., 2005; Boyer et al., 2005; Loh et al., 2006). Overexpression and RNAi studies revealed that NANOG regulates both *Sox2* and *Pouf51 (Oct4)*. The conclusion that a few select factors lie at the core of the GRN driving the pluripotency of stem cells was driven by Shinya Yamanaka, who initially showed OCT4, SOX2, KLF4 and C-MYC are sufficient to generate induced pluripotent stem cells (Takahashi and Yamanaka, 2006). Later, replacing C-MYC with NANOG yielded improved results, and ultimately, OCT4, SOX2, and NANOG were shown to be sufficient for reprogramming (Okita et al., 2007; González et al., 2011).

The main focus of the community for over a decade was expanding the number of genes incorporated into the relatively simple network. This included adding individual links through mutagenesis to expand the network up to 13 different interlinked TFs and then linking known signalling pathways to the stem cell network (Boyer et al., 2005; Kim et al., 2008; Chen et al., 2008). The linkage to other signalling pathways once again relied on CHIP-Seq data. More recent work started integrating single cell data and allowed the uncovering of specific mechanisms underlying the function of this GRN amongst which a bi-stable switch between pluripotency and differentiation using PCA and clustering algorithms (Shalek et al., 2013; Eichler et al., 2003; Mojtahedi et al., 2016).

Attempts at more detailed modeling of the core factors were undertaken by Narula et al. (2013), aiming to elucidate the differentiation into specific fates through precise quantification of the genes involved in the network. The expansion of studies investigating the quantitative aspects of the stem cell GRN was partly driven by the advent of single-cell high-throughput studies, which facilitated the exploration of the transcriptomic landscape of individual cells within a heterogeneous population (Tang et al., 2009; Okawa et al., 2016; Wei et al., 2016). This deepened our understanding of the network dynamics, revealing the mechanisms underlying the bi-stable switch between pluripotency and differentiation, including the role of positive feedback loops from NANOG, the identification of the primed state in stem cells transitioning from pluripotency to differentiation, and numerous other detailed findings (Figure

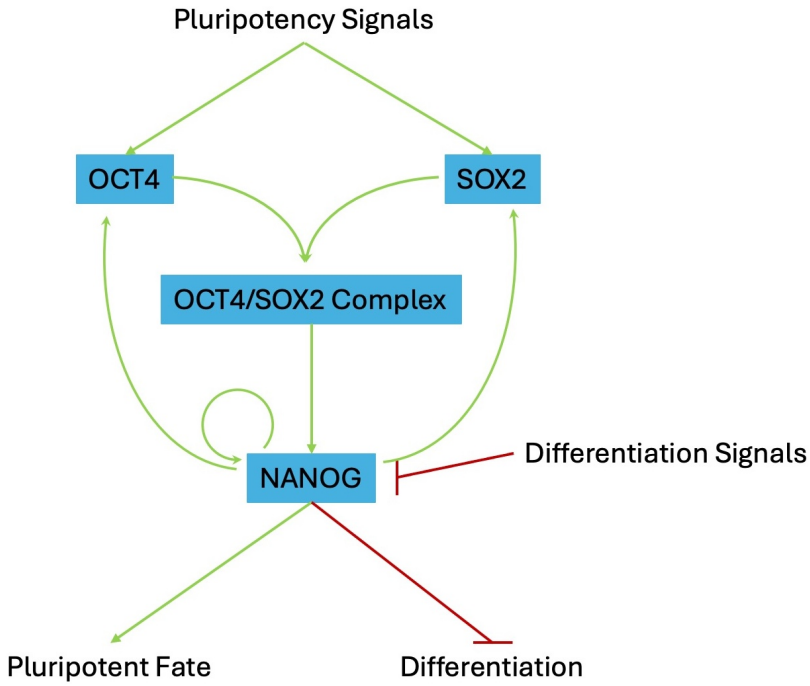


Figure 1.3: **The stem cell GRN in human stem cells:** The GRN model as described by Chickarmane et al. (2006) based on the core genes identified by Takahashi and Yamanaka (2006) showing the positive feedback loop centred around NANOG

1.3). Overall, the approach to modeling stem cell GRNs began with a focus on a few key factors, later expanding to incorporate signaling pathways and more detailed analyses of individual gene connections. We will try and replicate this approach in *C. elegans* stem cells.

1.2 Introduction to *C. elegans*

1.2.1 *C. elegans* development

Caernohabditis briggsae was described in a 1963 letter by Sydney Brenner as an ideal organism to study development and the nervous system (Brenner, 1974, 2001). However, Brenner later decided to focus more closely on the closely related *Caernohabditis elegans* as it was easier to grow in laboratory conditions (Félix and Wagner, 2008).

C. elegans is a tiny roundworm which hatches at $\approx$ 0.25mm in length and grows to 1mm when reaching adulthood. There are many advantages of working with this organism. Its life cycle from single cell to fully grown adult is $\approx$3 days when grown at 20℃ (Brenner, 1974, 2001). The longest developmental stage is the larval stage which is subdivided into 4 stages known as L1, L2, L3 and L4. These are separated by molting, during which the old cuticle is shed and replaced with a new one. Development of *C. elegans* can be accelerated by growing the animals at 25℃ and slowed down at 15℃ (Brenner, 1974; Hirsh and Vanderslice, 1976). However, extremes of high and low temperatures are prohibitive and cause developmental arrest and death (Hirsh and Vanderslice, 1976).

When the animal is subjected to unfavourable conditions, one of its survival strategies is to enter the dauer larval stage at the end of L1. This is an additional larval stage which which enables survival without nutrients and even in a dehydrated state. Upon removal of stressors, such as the addition of food or a return to normal environmental temperature, the animal transitions back to L2.

This dauer stage facilitates the maintenance of animals through freezing, making it highly advantageous for strain preservation. Furthermore, the worm's transparency allows for *in vivo* imaging of various tissues and cells. By utilizing fluorescent proteins, specific protein behaviors in developmental processes can be visualized without compromising spatial information or requiring tissue isolation (Chalfie et al., 1994; Huang and Chalfie, 1994; Boulin et al., 2006; Feinberg et al., 2008).

The majority of any *C.elegans* population consists of hermaphrodites, characterized by the sex chromosomes XX. Hermaphrodites exhibit female anatomy, with their sole male trait being

the ability to produce sperm during the L4 stage. The number of sperm in hermaphrodites is limited and is only used for self-fertilisation, allowing the production of $\approx$ 300 progeny. However, albeit infrequently, *C. elegans* are able to produce males with sex chromosomes XO. This happens due to X chromosome non-disjunction and occurs in $\approx$1/1000 cases under laboratory conditions. Males are identifiable by a hook-like structure on the tail, which becomes visible around the L3 stage and serves to attach to the hermaphrodite's vulva during mating. Mating between a male and a hermaphrodite yields more progeny than hermaphroditic self-fertilization, typically resulting in $\approx$ 1000 fertilized eggs. The presence of males significantly facilitates crosses between different strains.

One of *C. elegans*' most remarkable attributes is its ability to be used as a genetic model organism. At least 38% of *C. elegans* protein-coding genes are predicted to have orthologs in humans, while 60% of human protein-coding genes, including 40% of disease related genes are found to have orthologs in *C. elegans* (Shaye and Greenwald, 2011; Kaletta and Hengartner, 2006; Culetto and Sattelle, 2000). The advent of double-stranded RNA interference (RNAi) further solidified *C. elegans* as a premier model organism. First demonstrated by Montgomery et al. (1998), RNAi enables gene silencing by leveraging the siRNA machinery to bind and cleave specific mRNA targets. While this was Initially administered via injection, subsequent research by Timmons and Fire (1998) revealed that RNAi can be effectively induced by feeding animals with dsRNA-expressing bacteria or by soaking them in a dsRNA solution, facilitating systemic interference throughout the organism's body (Timmons and Fire, 1998).

Since the development of CRISPR (clustered regularly interspaced short palindromic repeats-associated) technology (Jinek et al., 2012), the manipulation of genes in *C. elegans* has surpassed even Sydney Brenner's initial expectations. Just a year after its development, CRISPR was successfully implemented in *C. elegans* (Friedland et al., 2013). The pioneering study demonstrated the generation of loss-of-function mutants for *unc-119* and *dpy-13*, showcasing the precision with which even essential genes could be targeted and their activity modulated.

Another useful property of *C. elegans* is its transparency which allows the tracking of

fluorescently tagged proteins without the need of isolating tissues (Chalfie et al., 1994; Zhang et al., 2002). It allows the visualisation of protein location, movement and conformational changes. The transparency of the animal can also be exploited for the visualisation of mRNA through techniques like smFISH (Raj et al., 2008, 2010). This technique is able to localise single mRNAs as diffraction limited fluorescent spots and was first used in *C. elegans* by Saffer et al. (2011) but has become since widely used for the quantification of transcripts (Barkoulas et al., 2013; Katsanos and Barkoulas, 2022; Koneru et al., 2021b). The technique requires fixation and permeabilisation of the animal to allow the dye-tagged RNA probes to bind the mRNA of the animal, and thus cannot be paired with live imaging. However, it provides detailed information on the location and intensity of transcription of specific genes. New techniques to be paired with live imaging have been developed in recent years, like the MS2 system (Tutucci et al., 2018; Kinney et al., 2023; Lee et al., 2019; Tocchini and Mango, 2024). This paired with a micro-fluidics devices can be utilised for true live imaging of transcription (Keil et al., 2017), however this is an extremely labour intensive process and thus is more aligned with detailed dynamic studies of individual genes.

Overall *C. elegans* is an excellent model organism for in-depth understanding of GRNs. It is both easy to maintain and to modify genetically. As such genetic perturbations can be introduced to easily establish causality in links between genes in the GRNs. The animal's robust behaviour means that any change in cell division patterns is easily identifiable. Despite all this there are still broad gaps in our understanding of *C. elegans* GRNs.

1.2.2 *C. elegans* epidermis

Skin is an indispensable component in every animal's physiology, crucial for survival and developmental processes (Fuchs, 2007). It is defined as the epithelial layer that surrounds an animal. In *C. elegans*, this layer is referred to as the hypodermis. Although simpler in structure compared to mammalian skin, the hypodermis of *C. elegans* shares several fundamental traits. Notably, both possess an epidermis supported by a basement membrane, a thin extracellular matrix providing essential structural support to the surrounding tissue (Figure 1.4). The outermost layer comprises a permeability barrier, characterized by the lipid-based

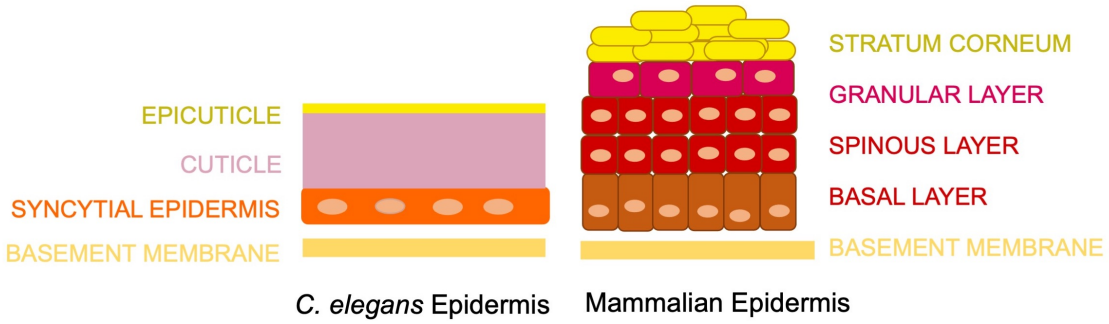


Figure 1.4: **The comparison between the structure of the *C. elegans* epidermis and the mammalian epidermis:** An illustration showing the similarities and differences in the general structure of mammalian and *C. elegans* epidermis.

epicuticle in *C. elegans*, whereas in mammals, it consists of the tough, keratinous stratum corneum. In both cases, the skin contains mechanoreceptors. (Schafer, 2015; Glauser, 2022; Chisholm and Hsiao, 2012).

A notable distinction between mammalian and *C. elegans* epidermis is that the main layer of the *C. elegans* hypodermis is composed of syncytia. Syncytia are multinucleated cells formed by fusion. The *C. elegans* hypodermis comprises numerous syncytia, with hyp7 being the largest among them (Figure 1.5). Originating from the AB and C blastomeres, the hyp7 syncytium expands during larval development through the differentiation of epidermal stem cells known as seam cells (Brunschwig et al., 1999; Cassata et al., 2005; Gilleard and McGhee, 2001; Landmann et al., 2004; Shao et al., 2013; Liu and Murray, 2023). Syncytia are one of the three epidermal cell types found in *C. elegans*, which are: syncytia, seam cells and P cells (Sulston et al., 1983). The focus of this thesis are the seam cells.

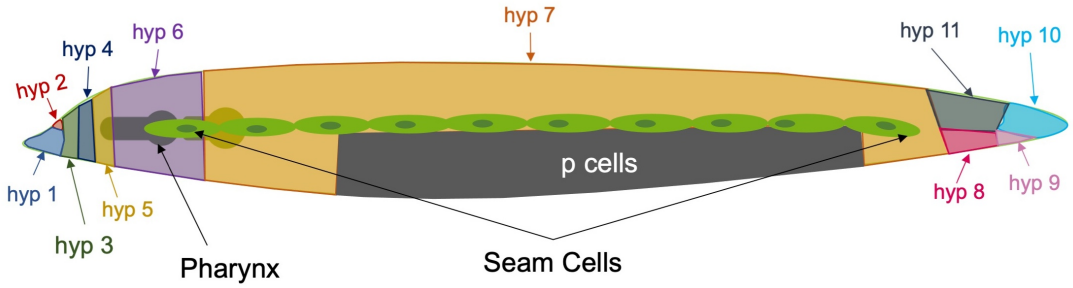


Figure 1.5: **Illustration of the epidermal syncytia of *C. elegans*:** The 11 epidermal syncytia of *C. elegans*, known as hyp 1-11 and their position within the nematode's body. Each syncytium is shown in a different colour that matches its label. Seam cells are shown in green, and the position of the p cells is shown in grey. The pharynx is shown in light grey.

1.2.3 *C. elegans* and stem cells

C. elegans maintains two types of cells that can be classified as stem cells post-hatching: germ cells and the epidermal stem cells. The epidermal stem cells, known as seam cells, are the sole autosomal stem cells found in *C.elegans* (Joshi et al., 2010). Upon division, seam cells can differentiate into neuronal and epidermal cell types while retaining the ability to self-renew; both defining properties of stem cells (Chisholm and Hsiao, 2012; Page et al., 1997b; Boveri, 1888). However, upon reaching adulthood, seam cells undergo terminal fusion and lose their ability to differentiate or divide, leading to their designation as stem-like cells rather than stem cells. Initially, each lateral side of the animal contains 10 seam cells, which form $\approx$ 4 hours after the first embryonic cell division. All 20 seam cells originate from the AB cell lineage within the embryo. Specifically, V3, V5, and T arise from the ABpxap lineage, while H0 and H1 originate from the ABarpa lineage, and H2, V1, V2, V4, and V6 from the ABarppx lineage (Liu and Murray, 2023).

After hatching all cells except H0 undergo multiple divisions (Figure 1.6). The *C. elegans* cell division pattern is highly robust meaning that all cells divide in the same way in all animals. This means that the terminal seam cell phenotype is highly robust and consists of 16 seam cells. Robustness is defined as phenotypes that remain invariant in the face of internal (genetic) or external (environmental) perturbation (Félix and Wagner, 2008) V5 is the first seam cell to divide, often before the animal hatches, which occurs approximately 8-10 hours

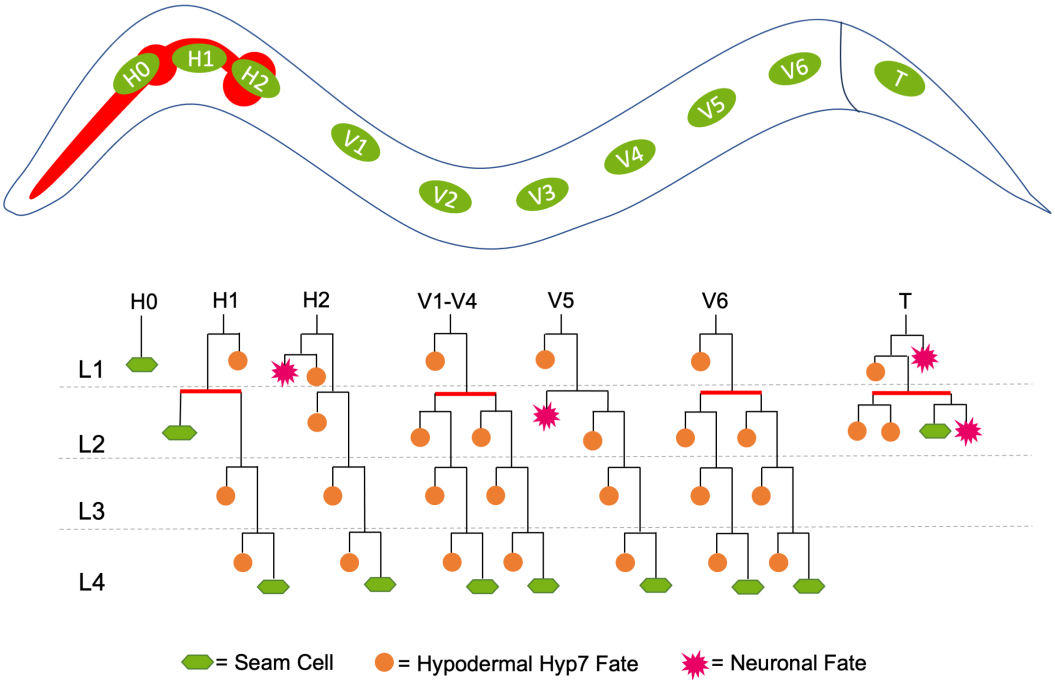


Figure 1.6: **Diagram showing the division patterns of the seam cells:** The position of the 10 seam cells that *C. elegans* hatches with and the division pattern of each of these cells including the expansion to 16 seam cells. The *C. elegans*' pharynx is shown in the head in red, and the rectum is denoted in the posterior with a green line. Seam cell fate is denoted as green circles, hypodermal cell fate is denoted in orange, and neuronal fate is shown in magenta. Red thick bars between daughter cells indicate a symmetric division and black thin horizontal lines indicate an asymmetric division.

post-bleaching. Most commonly, seam cells undergo asymmetric division, resulting in one seam cell and one hypodermal cell, which subsequently fuses to the hyp7 syncytium. This division pattern occurs in all cells except H0. During the L1 stage, V1-6 divide carry out this division, with the anterior cell adopting the hypodermal fate and the posterior cell remaining a seam cell poised for further differentiation. H1 follows a similar pattern, but with the posterior cell acquiring the hypodermal fate and the anterior cell remaining a seam cell. H2 initially divides to produce two seam cells; the posterior cell remains a seam cell and does not divide until L2, while the anterior cell gives rise to a hypodermal hyp7 cell and a neuronal anterior deirid socket cell. The T cell undergoes two divisions during L1: the first division results in one anterior seam cell and one posterior neuronal cell, eventually giving rise to five different neuronal cells including the PLN neurons (Sulston et al., 1983; Sulston and Horvitz, 1977; Chang et al., 2006). In the second division, one anterior hypodermal hyp7 cell is produced, while the second cell maintains its seam cell fate. These divisions occur between 10-16 hours post-hatching (Sulston et al., 1983; Harandi and Ambros, 2015).

The early L2 stage in *C. elegans* is marked by the occurrence of symmetric divisions. Specifically, V1-V4 and V6 undergo a single symmetric division approximately 25 hours post-bleaching, followed by an asymmetric division roughly an hour later, which results in one seam cell and one cell that fuses with the hyp7 syncytium. Similarly, H1 undergoes symmetric division at L2, but it does not proceed to asymmetric division afterward; instead, the anterior daughter cell remains undivided throughout larval development, akin to H0, until it eventually merges terminally with the seam line.

Also during the L2 stage, V5 generates the PVD and PDE neurons from the anterior cell following an asymmetric division (Sulston and Horvitz, 1977; Sulston and White, 1980). The remaining cell then undergoes another asymmetric division, with the anterior cell assuming the hyp7 fate and the posterior cell maintaining its seam cell identity. Likewise, the T cell undergoes symmetric division at L2 (Sulston and Horvitz, 1977), after which both daughters undergo further division. The anterior daughter cell produces two hyp7-fated cells, while the posterior daughter cell gives rise to the PVN neuron from its posterior cell, and a non-dividing seam-cell from its anterior cell (Hall and Russell, 1991; White et al., 1986). In subsequent

larval stages, from H2 to V6, each cell undergoes one asymmetric division per stage, yielding an anterior cell fated for hyp7 and a posterior seam cell. Upon completion of the larval stages, all 16 cells fuse and cease division (Sulston and Horvitz, 1977). Seam cells play indispensable roles in *C. elegans* morphogenesis and ensure proper growth by generating essential neuronal and epidermal tissues.

1.2.4 Gene regulatory mechanisms of seam cell patterning

The earliest gene expressed in the hypodermis to be implicated in seam cell specification is *elt-1*. ELT-1 functions as a GATA transcription factor and was initially identified for its role in activating the epidermal differentiation GATA factor cascade in the embryo (Koh and Rothman, 2001; Page et al., 1997b). ELT-1 has 60% homology to human GATA1, GATA2 and GATA3 (Page et al., 1997a). In humans GATA factors play crucial roles in stem cell regulation, particularly in hematopoietic stem cells, where they govern differentiation (Simon et al., 1992; Bresnick et al., 2010; Fujiwara, 2017). Specifically, GATA3 is essential for the function human epidermal stem cells and is implicated in repressing proliferation and promoting differentiation of epidermal stem cells into keratinocytes (Masse et al., 2014).

In the *C. elegans* embryo, *elt-1* is expressed in all epidermal precursors and is involved in the formation of almost all hypodermal cell types (Page et al., 1997b). Loss of *elt-1* function leads to failure in hypodermal formation and early developmental arrest (Page et al., 1997b). Abnormalities in cell behaviour in the absence of *elt-1* are noted even up to 2 cell divisions before any of the hypodermal cell types are specified (Page et al., 1997b). Knockout mutants of *elt-1* exhibit a redirection of epidermal cell fate towards non-epidermal lineages, highlighting the indispensable role of *elt-1* in epidermal specification. The ELT-1 protein features a double GATA-type zinc finger domain structure, reminiscent of vertebrate GATA factors, although unique to *C. elegans*. While expressed in multiple tissues during embryonic development (Smith et al., 2005), ELT-1 persists exclusively in seam cells and neuronal cells of the retrovesicular ganglion during post-embryonic stages. In fact ELT-1 is often portrayed as the master regulator of seam cell fate however, numerous genes are considered crucial for maintaining seam cell integrity. RNAi experiments targeting *elt-1* in the larva result in defective seam cell development, leading to severe phenotypic abnormalities such as dumpy phenotypes and impaired motility (Smith et al., 2005). *elt-1* has been shown essential for seam cell formation, and the absence of *elt-1* and therefore seam cells is fatal for animal development (Smith et al., 2005). Notably, repression of *elt-1* during larval stages culminates in the complete loss of structural integrity in animals post-L4 moult, concomitant with the absence of seam cells. However, the precise causality underlying these observations

remains to be determined. In order to carry out detailed studies on the role of *elt-1* in *C. elegans*, the *elt-1(ku491)* partial loss-of-function mutation was generated. This mutant allele contains a C-to-T substitution in exon 5, causing missense mutation of ELT-1, which unlike the *elt-1(ok1002)* allele, isn't lethal (Cohen et al., 2015).

Another significant GATA factor in seam cells is *egl-18*. In a study by ENCODE Project Consortium (2012); Luo et al. (2020) ELT-1 ChIP-seq analysis revealed peaks in the upstream region of *egl-18*, suggesting a potential binding mechanism. Although the exact nature of this interaction remains unexplored, and perturbational experiments have yet to be conducted, this observation hints at a plausible direct regulatory relationship.

In Koh and Rothman (2001), it was discovered that *egl-18* (then known as *elt-5*), along with its paralog *elt-6*, are essential for seam cell development. These two genes are found to be transcribed either monocistronically or dicistronically, depending on the cell type. They are $\approx$130bp apart which is characteristic of genes that reside together in an operon (Page et al., 1997a). However, subsequent studies have suggested that both genes are likely transcribed from independent transcription sites (Koh and Rothman, 2001). Deletion of both genes simultaneously results in frequent embryonic arrest (Koh and Rothman, 2001). However, even the removal of just *egl-18* causes a severe seam cell defects, and morphological abnormalities, particularly in the vulva, resembling those observed with the removal of both genes and indicative of epidermal defects. This is observed in the *egl-18(ga97)* mutation, which converts a CGA Arginine codon in exon 2 to a premature TGA stop codon, leading to a truncated protein. This results in an average phenotype of 12 seam cells and a malformed vulva (Eisenmann and Kim, 2000). Knockouts of *elt-6*, on the other hand, do not appear to exhibit any discernable phenotype. The losses of seam cells caused by the absence of *egl-18*, were found to be a result of inappropriate fusion of seam cells to the syncytium (Koh and Rothman, 2001). During the 1.5-fold stage ($\approx$ 7 hours after the first cell cleavage) (Chisholm and Hardin, 2005), the gene *elt-3* serves as a marker for all epidermal tissues except seam cells. In the absence of *egl-18* and *elt-6*, ectopic expression of *elt-3* is observed. As reported in Koh and Rothman (2001), ectopic expression of *elt-3* is most frequently observed in the V3 seam cells, occurring in 80% of cases. However, despite this observation, V3 is not the most

susceptible to fusion but is rather believed to acquire a fate closely related to the P-cell fate (Sulston et al., 1983). In the vulva, *egl-18* has been demonstrated to be crucial for proper vulval formation, indicating that its role extends beyond the seam cells (Koh et al., 2002). This underscores the importance of *egl-18* in both the determination and maintenance of seam cells, establishing it as one of the core genes in the seam cell gene network.

When discussing inappropriate fusion resulting from the loss of *egl-18*, it is important to consider the role of fusogens. Biological membranes are composed of lipid bilayers with highly repulsive energy between them due to the polar heads of their phospholipids. Thus, to facilitate the fusion process, they necessitate specialized proteins known as fusogens. Within seam cells one of the most important fusogens is EFF-1. In the absence of *eff-1*, even cells that successfully exit the seam cell program and differentiate into epidermal tissue are unable to form a syncitium (Podbilewicz et al., 2006; Koneru et al., 2021b). *egl-18* has been shown to repress *eff-1* expression, thereby preventing inappropriate fusion. However, during differentiation into the hyp7 syncytium, *eff-1* expression is activated.

eff-1 has also been shown to be repressed by CEH-16 (Cassata et al., 2005), a gene identified as a potential direct target of ELT-1 in studies by Luo et al. (2020); ENCODE Project Consortium (2012). *ceh-16* is the homolog of the *Drosophila* gene *engrailed* (Cassata et al., 2005; Eker, 1929), which plays a pivotal role in appendage development and animal segmentation. The human homologs of *engrailed*, EN1 and EN2, are implicated in cerebellum and hind-brain development, establishment of dorso-ventral patterning, and limb development (Loomis et al., 1996). Since its discovery, *engrailed* has garnered significant interest due to its involvement in several degenerative diseases, including Parkinson's and Huntington's (Rekaik et al., 2015; Rissling et al., 2009; Mugat et al., 2008; Simon et al., 2001) Its repression domain has been widely utilized in constructing artificial transcription factors across various organisms, ranging from *Drosophila* to plants, frogs, and even mice (John et al., 1995; Markel et al., 2002; Viger et al., 2008; Suzuki et al., 2022). Despite its predominant use for its repressive domain, *engrailed* has also been implicated in transcriptional activation (Alexandre and Vincent, 2003; Serrano and Maschat, 1998).

The identification of the homolog of this gene in *C. elegans* epidermal stem cells provides

an opportunity to further investigate its function. In a study utilizing extra-chromosomal arrays to knock out *ceh-16* in a subset of seam cells, it was demonstrated that cells lacking *ceh-16* exhibited incorrect positioning and fusion with the epidermis in the embryo, indicating the essential role of *ceh-16* in maintaining seam cell fate and preventing the acquisition of epidermal fate. The downregulation of CEH-16 is thought to prime a differentiation program in seam cells in the embryo by both derepressing the fusogen *eff-1*, involved in seam cell-epidermal fusion, and independently regulating other factors, as evidenced by the failure of cells expressing the fusogen to maintain seam cell fate in the absence of *ceh-16*. In the embryo, ectopic expression of *ceh-16* has been shown to activate *egl-18*, a known seam cell marker, in the dorsal epidermis (Cassata et al., 2005). Furthermore, seam cells that never expressed *ceh-16* in the embryo failed to develop into larval seam cells. Upon entering the larval stage of development *ceh-16* is restricted exclusively to seam cells and its complete absence leads to embryonic lethality (Cassata et al., 2005).

ceh-16 has been identified as an essential gene for the correct behaviour of seam cells during larval development. The two essential properties of a stem cell are the capacity to differentiate into different cell types, but also to self-renew in order to expand its own population. Given the lethality associated with complete loss of *ceh-16*, partial loss-of-function mutants, such as *ceh-16(bp323)*, have been developed. The *ceh-16(bp323)* mutation is temperature-sensitive and impairs DNA binding through the conversion of a Threonine to an Asparagine in the first of the three *ceh-16* DNA-binding helixes (Huang et al., 2009). Most of the self-renewal divisions of the seam cells occur during L2, however in the *ceh-16(bp323)* mutant the self-renewal divisions have been observed to fail (Huang et al., 2009). Interestingly, there are no temporal disruptions, and all asymmetric divisions after the L2 symmetric division proceed normally. However, loss of *ceh-16* results in the conversion of a symmetric division into an asymmetric one. One cell in which the loss of *ceh-16* doesn't cause the loss of proliferation, but rather causes an asymmetric division to become symmetric, is the V5 L2 division that results in the differentiation of the postdeirid neuron. In 65% of *ceh-16(bp323)* animals, the PDE neuron was absent, and a small proportion exhibited two V5 seam cells at the end of L4 (Huang et al., 2009).

Overexpression of *ceh-16* during late L1 and early L2 stages leads to seam cell hyperplasia, with an average seam cell count of approximately 22.5 (Huang et al., 2009). The additional seam cell divisions occur during the L2 stage, indicating that *ceh-16* overexpression alone is sufficient to induce seam cell duplication (Huang et al., 2009). This duplication is thought to be linked to the role of RNT-1 and BRO-1 which are involved with repressing the gene *pop-1* (van der Horst et al., 2019). POP-1 protein asymmetry has been shown to be highly important in the maintenance of asymmetry in asymmetric divisions (van der Horst et al., 2019; Harandi and Ambros, 2015; Gleason and Eisenmann, 2010), and the repression of *pop-1* by the RNT-1/BRO-1 complex is thought to be essential in the conversion of the L2 division to a symmetric one (van der Horst et al., 2019). Despite the role of *ceh-16* in the symmetric division, no link between these genes has been formally established. The *ceh-16(bp323)* strain phenotype has also been found to be counteracted by the absence of the gene *apr-1*. Under normal circumstances *apr-1(bp298)* missense mutation causes a symmetrisation of the L4 division. However, together with *ceh-16(bp323)* it allows the recovery of the L2 symmetric division (Huang et al., 2009). Overall, *ceh-16* emerges as one of the most crucial genes for seam cell maintenance, dictating both seam cell specification and their self-renewal capacity, a defining characteristic of these cells.

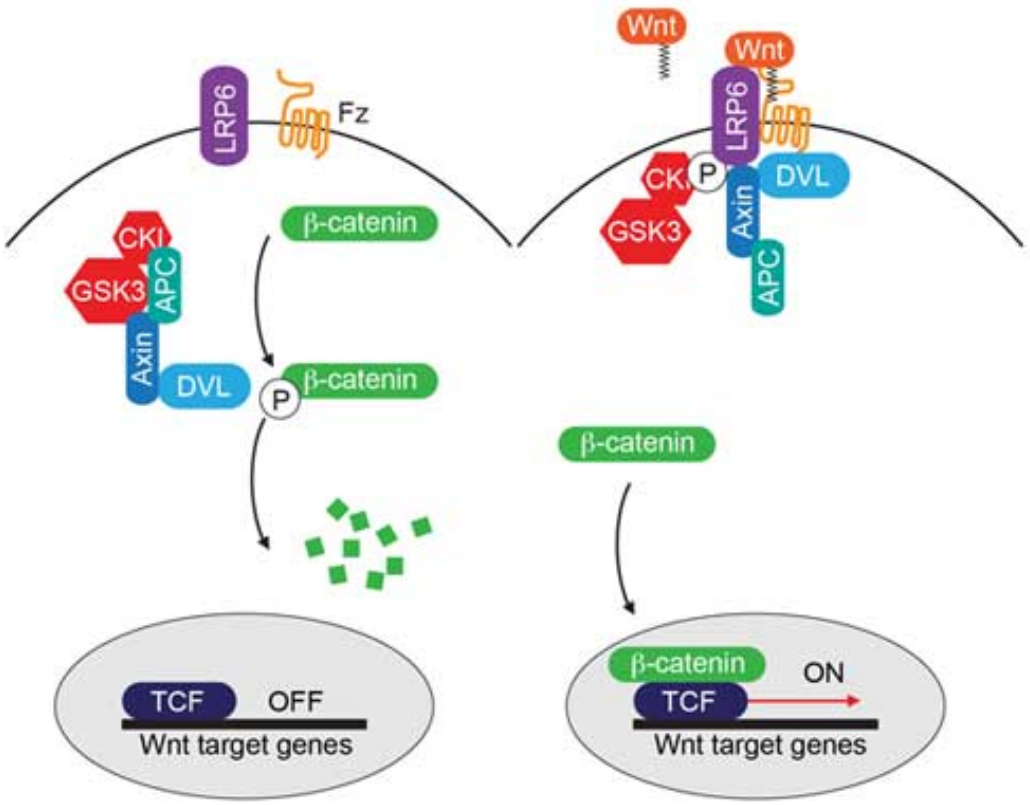


Figure 1.7: **Diagram showing canonical Wnt signalling components:** Wnt ligands are required to activate Wnt target genes. The signal from Wnt ligands is transduced through frizzled, dishevelled (DVL), LRP6, the APC complex and other Wnt components. In absence of Wnt signal β-catenins are phosphorylated preventing them from binding TCF whose role is to activate Wnt targets

1.2.5 *Wnt* regulatory system in C. elegans

The *Wnt* pathway is a conserved signaling cascade pivotal in the development of various animal species, ranging from mammals to sponges. *Wnt* refers to a conserved group of lipid-modified glycoproteins (Cadigan and Nusse, 1997; Clevers and Nusse, 2012), which play crucial roles in regulating diverse processes such as cell division asymmetry and differentiation (Angers and Moon, 2009). They function not only in short-range signaling but also provide positional cues by establishing concentration gradients across developing tissues (Zecca et al., 1996). Canonical *Wnt* signalling functions through the β-catenin group of effector proteins, and non-canonical *Wnt* functions independently of them. In canonical *Wnt* signalling, the absence of a *Wnt* signal triggers the activation of the β-catenin degradation complex (Akiyama, 2000). This complex, comprising scaffold protein Axin, (*APC*), kinases, and the *Dishevelled* protein, phosphorylates β-catenins, marking them for ubiquitination

and subsequent degradation via the proteasome (Nakamura et al., 1998; Huelsken and Behrens, 2002; Sokol, 1996). *Wnt* signals are received through the membrane-bound co-receptors known as *frizzled* and LRP6. In presence of *Wnt* signalling LRP6 recruits the *Dishevelled* and Axin proteins to the plasma membrane (Ren et al., 2021; Bafico et al., 2001). Derepressed β-catenins then activate TCF, which modulates target gene expression. In *C. elegans*, homologs of these proteins perform analogous functions. For instance, POP-1 is a TCF homolog, LIN-17, MIG-1, CFZ-2 and MOM-5 are Frizzled homologs, EGL-20, CWN-1, CWN-2, LIN-44, and MOM-2 are *Wnt* ligands, *BAR-1* and *WRM-1* are β-catenins, *APR-1* is an APC homolog, and LIT-1 is an NLK homolog and is one of the kinases involved in TCF nuclear export (Thorpe et al., 2000; Korswagen, 2002; Herman, 2003; Whangbo and Kenyon, 1999). These *Wnt* genes have been found to be essential in maintaining cell asymmetry, in particular in the seam cells (Mizumoto and Sawa, 2007a). The *Wnt* pathway has also been found to affect division timings and developmental advancement of the seam cells (Ren and Zhang, 2010), as well as regulating *egl-18*, a core seam cell gene (Gorrepati et al., 2013).

The gene *pop-1* plays a pivotal role in the differentiation program of the anterior cell during asymmetric division and is crucial for maintaining asymmetry. It functions via the non-canonical *Wnt* pathway. *pop-1* exhibits three main expression level thresholds (Calvo et al., 2001; Kidd et al., 2005; Shetty et al., 2005; Huang et al., 2007). At high expression levels POP-1 forms a complex with the Groucho homolog *unc-37* and acts as a repressor of seam cell fate (Calvo et al., 2001; Bekas and Phillips, 2022). At intermediate levels it interacts with the β-catenins *sys-1* and *wrm-1* to activate seam cell fate (Kidd et al., 2005; Shetty et al., 2005; Huang et al., 2007). The complete absence of *pop-1* results in a lethal phenotype (van der Horst et al., 2019). In asymmetric divisions, with the anterior cell destined for differentiation, POP-1 protein levels are significantly higher in the anterior cell than in the posterior (van der Horst et al., 2019). In the asymmetric division POP-1 itself is regulated by many different factors, key of these being APR-1 and LIT-1, and in turn it regulates many important seam cell factors, like *egl-18* (Gorrepati et al., 2013).

The APC homolog, APR-1, plays a pivotal role in *Wnt* signalling. APR-1 mutants were identified through abnormal seam cell phenotypes, as the gene's absence causes a con-

version from an asymmetric L4 division to a symmetric one (Mizumoto and Sawa, 2007b). APR-1 facilitates the export of the WRM-1 and SYS-1 β-catenins out of the nucleus (Baldwin and Phillips, 2014). Furthermore the *apr-1(bp298)* knock-out, has been demonstrated to alleviate the loss of symmetry in *ceh-16(bp323)* mutants (Huang et al., 2009).

LIT-1 is also integral to asymmetry and functions as a protein kinase activated by the β-catenin WRM-1 (Rocheleau et al., 1999). Under normal circumstances, LIT-1 localizes to the anterior cortex of cells, but following an asymmetric division, it relocates to the nucleus of the posterior cell. LIT-1 is crucial for phosphorylating POP-1 in the posterior cell, triggering the export of POP-1 from the nucleus and maintaining the seam cell fate (Lo et al., 2004). *lit-1* expression has also been found to be repressed by EFL-3 (unpublished data from the lab). This gene has been found to be expressed in seam cells (Marshall et al., 2016) and be involved in the maintenance of robust seam cell division patterns in larval development (Katsanos, 2021).

During a symmetric division, both POP-1 and APR-1 are asymmetrically distributed, similar to an asymmetric division (Wildwater et al., 2011; Baldwin and Phillips, 2014). However van der Horst et al. (2019) identified the homologs of RUNX and CBFβ, *rnt-1* and *bro-1* respectively, as potential candidates for converting asymmetric to symmetric cell divisions. Loss of either gene results in failures in seam cell divisions, whereas overexpression leads to hyperplasia (Nimmo et al., 2005; Kagoshima et al., 2007). The BRO-1/RNT-1 complex has been shown to promote seam cell fate, by down-regulating *pop-1* expression during the L2 symmetric division. The repression of *pop-1* expression results in the anterior cell expressing *pop-1* below the threshold required for the cell to exit into the differentiation program.

1.3 Aims

The objective of this thesis is to understand the genetic mechanisms underlying epidermal stem cell patterning in *C. elegans*. Despite being a very well characterised organism, our knowledge of the *C. elegans* epidermal stem cell GRN lags behind that of other organisms. As such our ultimate goal was to describe in detail the mechanisms governing epidermal stem

cell patterning. The state-of-the-art of the field so far describes three main factors as essential for seam cell maintenance (*elt-1, egl-18* and *ceh-16*). Several studies propose ELT-1 as the master regulator of the other two genes as it is the first gene to be expressed solely in the seam cell (Koh and Rothman, 2001; Baugh et al., 2005; Smith et al., 2005; Thompson et al., 2016). This however, has never been tested through a perturbational approach. There has also been a proposed positive interaction between CEH-16 and *egl-18* (Cassata et al., 2005). However, this was only shown at embryonic level and failed to consider that the main seam cell functions occur at the larval stages.

We selected a time point we assumed to be steady state and developed a detailed quantitative model of the interactions between these three genes. We relied on smFISH data coupled with a selection of mathematical approaches, starting from a modified version of the MRA and finally developing a minimal model using a new bayesian inference technique called SLING. We test our minimal model by making predictions on the behaviour of the three core genes in a novel mutant, resulting from a cross between *ceh-16(bp323)* and *egl-18(ga97)*. In this cross we observe a highly robust phenotype where the modified nematodes reach adulthood with only 11 seam cells in over 75% of the animals with a seam cell number range of 10-12. This very narrow distribution is unusual for strains with an erroneous seam cell number as most seam cell specific mutants exhibit high variability. As such in Chapter 5 we investigate the *ceh-16(bp323);egl-18(ga97)* strain more closely.

An assumption that is widespread in the seam cell community is that all the cells are the same and that the core genes interact in the same way at all time points. In Chapter 4, we question this assumption by investigating the seam cells behaviour at different time points and by questioning the similarities between them. We start by observing these interactions at 3 additional time points: pre-L1 division, post-L2 symmetric division, and post-L2 asymmetric division. We observe some changes from what we observe at late L1 and some recurring patterns. We also test our assumption of the 17 hour time point being a steady state and we reach the conclusion that while 17 hours is most likely the best candidate for a steady state, the core gene expression is highly variable over time and thus every state appears to be highly unstable. We theorise that later time points could have lower and more stable core

gene expression. We further delve into spatial differences in core gene expression, as well as the different sensitivities of seam cells to the loss of the core genes. We identify patterns that strongly indicate that further studies of the differences between the individual seam cells are essential for the full understanding of the mechanisms underlying these highly robust stem cells. We conclude that the frequencies of loss of different seam cells correlate with the core gene expression patterns, especially in the late L2 stage. We also describe a connection between *elt-1* spatial patterning and the *Wnt* signalling pathway.

In Chapter 5 we address the *ceh-16(bp323);egl-18(ga97)* mutant we generate in Chapter 3. A seam cell phenotype where the terminal number of seam cells is erroneous yet consistent has never been identified before as seam cell division errors tend to disrupt robustness. We aim to understand whether the phenotype is truly robust and to uncover the mechanism underlying its extremely consistent phenotype. We compare it to a *ceh-16*$_{LoxP}$ mutant as both strains show a terminal seam cell phenotype of $\approx$ 11 seam cells with low variability. We assess *ceh-16(bp323);egl-18(ga97)* and *ceh-16*$_{LoxP}$ resistance to both genetic and environmental perturbations and note that not only is the double mutant's number of seam cells at the end of development more consistent than either of its parental strains, *ceh-16*$_{LoxP}$ and in fact most seam cell gene mutants, but also the specific seam cells that are lost are consistently the same. The resulting phenotype appears to be independent of the WBA pathway, as the *Wnt* genes appear to behave in a way that would lead to additional symmetrisation in non mutant animals. We did however observe a partial rescue using *apr-1* RNAi. We propose that *apr-1* has a larger role in symmetry than previously thought. Finally, we are able provide the core seam cell GRN with a connection to the genes involved in symmetry via *ceh-16* and *egl-18*.

Overall, we start by identifying core seam cell factors. We then seek an in depth analysis of the quantitative relationships between them. We interrogate this network over time and space and while we note that our minimal model is not sufficient to explain all the behaviours of the network at certain time points, we further strengthen several of our newly identified connections. We then expand our network to incorporate *Wnt* signalling with the help of a new mutant with a highly robust erroneous stem cell phenotype.

Chapter 2

Methods

2.1 *C.elegans* specific methods

2.1.1 Maintenance

C. elegans is maintained on 60 mm petri dishes of Nematode Growth Medium (NGM) seeded with 300 μl of OP50 *Escherichia coli* strain as described in (Brenner, 1974). The worms were transferred onto new plates using one of four methods: 1) Picking using a piece of platinum fused to a Pasteur pipette directly onto a new NGM plate. 2) Transferring a small square of agar to a new plate by cutting it with a scalpel. 3) Picking using a piece of platinum fused to a Pasteur pipette into a 10 μl drop of bleach (alkaline sodium hypochlorite solution). 4) Washing adult animals off a plate and into eppendorph and then resuspending in bleach till the eggs are extracted, then resuspending in M9 and placing eggs on the plate.

The last two of these methods allow the synchronisation of the animals. Unless stated otherwise all strains were maintained at 20℃. All buffers were produced as described in (Theresa, 2006). All strains used are listed in table 2.1 and are in the N2 Bristol background.

2.1.2 Freezing

Strains were frozen after being synchronised, put on standard NGM plates and left at 25℃ for 6 days to induce starvation and entry into dauer phase. Some slow growing strains

Table 2.1: **List of all strains used**: We include the names used in the text, the genotype of the strain and the official strain name.

Strain List		
Name	Genotype	Strain Name
Wild Type	unc-119(e2498::Tc1)III;(wIs51) V	JR667
elt-1(ku491)	elt-1(ku491)IV; wIs51[SCMp::GFP+unc-119(+)]V	MBA650
egl-18(ga97)	egl-18(ga97)IV; wIs51[SCMp::GFP + unc-119(+)]V	MBA290
ceh-16(bp323)	ceh-16(bp323ts); wIs51[SCMp::GFP + unc-119(+)] V	HZ620
ceh-16(bp323); *egl-18(ga97)*	ceh-16(bp323ts) III ;egl-18(ga97) IV; wIs51[SCMp::GFP + unc-119(+)] V	MBA1216
ceh-16$_{LoxP}$	icb128[LoxP:ceh-16(ex1-3):LoxP: ceh-16(ex4)] III ;wIs51[SCMp::GFP] V; icbSi218[pHIN32(ttTi5605 right::srf-3i1:: delta-pes-10::woCRE::unc-54-3utr::unc119)]	MBA1428
apr-1(bp298)	apr-1(bp298) I; wIs51[SCMp::GFP + unc-119(+)]V	HZ194
Triple Hox	lin-39(n1760);mab-5(e1239);egl-5(n945) III; wIs51[SCMp::GFP + unc-119(+)] V	MBA748
egl-5(n945)	egl-5(n945) III; wIs51[SCMp::GFP + unc-119(+)] V	MBA757
lin-39(n1880)	lin-39 (n1880) III; wIs51[SCMp::GFP + unc-119(+)] V	MBA747
bar-1(ga80)	wIs51[SCMp::GFP + unc-119(+)] V; bar-1(ga80) X	MBA836
egl-20(hu105)	egl-20(hu105) IV; wIs51[SCMp::GFP + unc-119(+)] V	MBA521
lin-17(n671)	lin-17(n671) I; wIs51[SCMp::GFP + unc-119(+)] V	MBA450
ELT-1::EGFP	OP354[unc-119(tm4063) III; wgIs354[elt-1::TY1::EGFP::3xFLAG + unc-119(+)]]; ouIs91[ajm-1::mCherry + unc-119+]? ; him-5 (e1490) V	MBA1465
CEH-16::AID::nG	ceh-16 nG AID::3xFLAGIII; ouIs91[ajm-1::mCherry + unc-119+] ; him-5 (e1490) V	MBA1464
EGL-18::nG	ujIs113 II; egl-18(dev210([mNeonGreen::egl-18])IV ;ouIs91[ajm-1::mCherry + unc-119+]?; him-5 (e1490) V	MBA1615

Table 2.2: **List of all RNAi used**: We include the names used in the text, the official gene name and the sequence used for the RNAi. All of these sequences were cloned into the control vector vector is L4440 in the *ht115* strain, which we refer to in the text as control vector and were generated by Kamath and Ahringer (2003); Kamath et al. (2003)

RNAi List		
RNAi	Gene ID	Sequence ID
elt-1	WBGene00001249	W09C2.1
egl-18	WBGene00001186	F55A8.1
elt-6	WBGene00001253	F52C12.5
ceh-16	WBGene00000439	C13G5.1
rnt-1	WBGene00004393	B0414.2
lit-1	WBGene00003048	W06F12.1
efl-3	WBGene00009899	F49E12.6
eff-1	WBGene00001159	C26D10.5
pop-1	WBGene00004077	W10C8.2
apr-1	WBGene00000156	K04G2.8

were left for 13 days. After the plates were full of dauer larvae the animals were washed off into an eppendorph, in which they were washed repeatedly in M9 and resuspended also in M9. Finally 1ml of M9-*C. elegans* solution transferred into a cryo-tube with 1 ml of Freezing solution. Each strain was frozen in 3 different cryo-tubes, which were placed at -80. If the test-thaw was successful, two of the three cryo-tubes were placed in liquid nitrogen.

2.1.3 RNA interference (RNAi) by feeding

RNAi was carried out by placing *C. elegans* on plates of *E. coli* expressing double stranded RNA (dsRNA). RNAi bacterial clones for the genes of interest were streaked onto a Lysogeny broth (LB) agar plate with 50μg/ml ampicillin and 12.5μg/ml tetracycline. The RNAi clones in question are listed in table 2.2. The clones were then inoculated in 12 ml of LB medium with 50μg/ml ampicillin and 12.5μg/ml tetracycline. In mixed RNAi treatment the cultures are then tested through a spectrophotometer to test their OD600 and diluted to equal concentration before they are mixed to the specified dilution with bacteria containing empty vector RNAi. For example, *50% egl-18* would be produced by mixing 6ml of *egl-18* RNAi with 6ml of empty vector RNAi, both diluted to equal concentration.

RNAi plates were prepared through the methods of preparation of NGM as described in

Theresa (2006) but with the addition of 25μg/ml ampicillin, 12.5μg/ml tetracycline and 1mM filter-sterilised Isopropyl β-D-1-thiogalactopyranoside (IPTG). The plates were to be used within three weeks. Animals were spot bleached onto a plate for a 1st generation treatment and if a two generation treatment is required then those animals, once they reach adulthood, are bleached onto a second plate of the same RNAi.

2.1.4 Genetics

Males were induced by subjecting 5 plates L4 hermaphrodites to heatshock by placing them at 30℃ for three hours and then 37℃ for half an hour. These were then allowed to lay eggs and 3 days later screened for males. To set up crosses we picked males and hermaphrodites onto a plate, maintaining a ratio of 10:1 males to hermaphrodite, with at least 3 hermaphrodites on the plate to ensure a successful mating encounter. For genotyping of the *ceh-16(bp323);egl-18(ga97)* cross we relied on the primers listed in (Table 2.4) for sequencing, as well as visual observations of *egl-18(ga97)* induced deformities in the animals.

2.1.5 Temperature treatment of animals

Three sets of 20 gravid hermaphrodites were spot bleached onto a regular NGM plate and placed at 15℃, 20℃, 25℃. They are allowed to reach adulthood and spot bleached onto three new plates. These plates are placed back at the same temperatures that the mothers were kept at. These animals had their seam cells counted once they reached L4.

2.2 Microscopy and Image analysis

2.2.1 Seam cell counts

Seam cell counts were carried out at the end of L4/early adult stage unless otherwise specified. They were counted using the SCM::GFP (seam cell marker) fusion signal at the Zeiss compound microscope using a 20x objective. The specific cells that were missing or

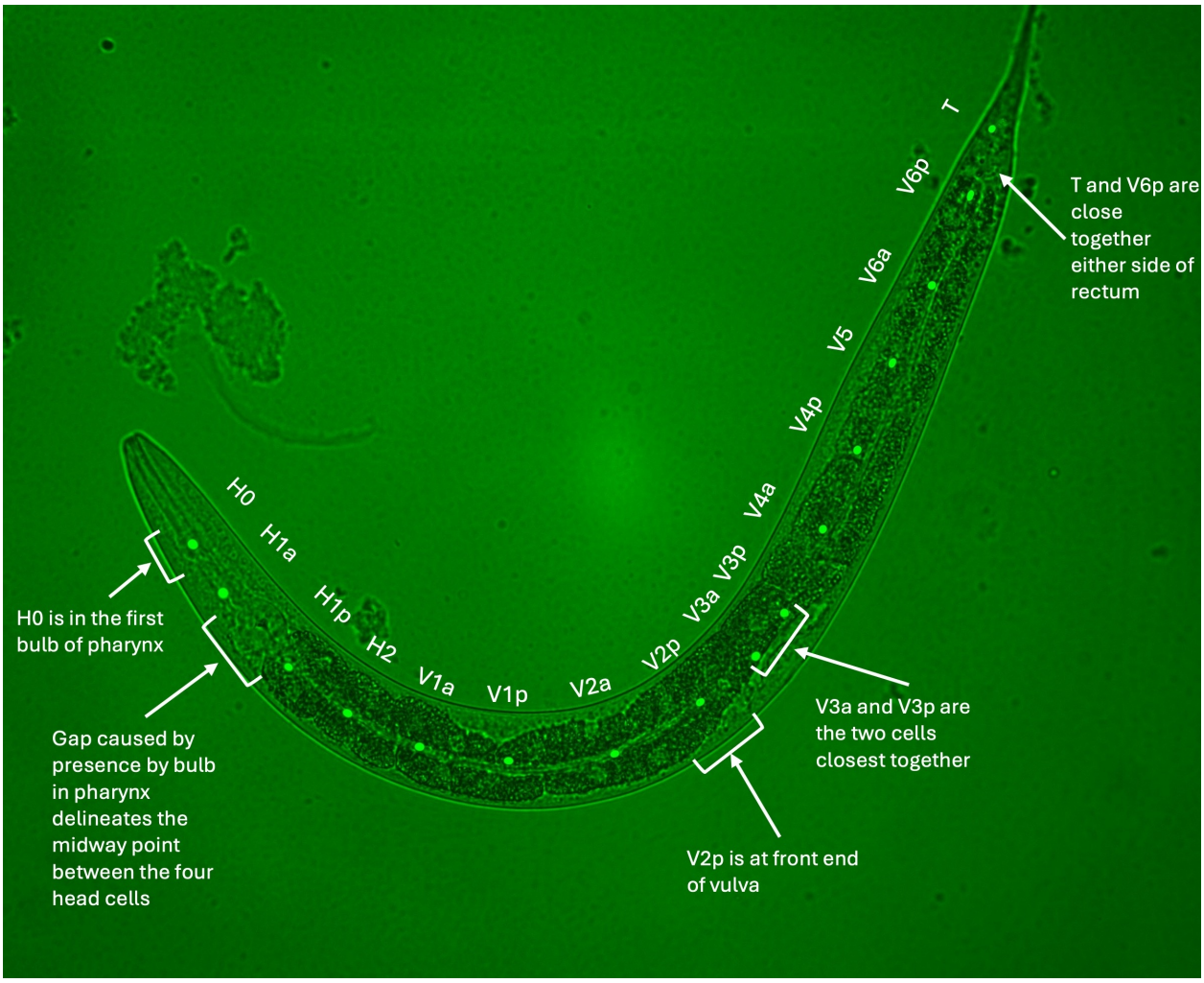


Figure 2.1: **Description of how seam cell-specific abnormalities are identified**: Here we highlight the anatomical landmarks used to determine which cells are lost or gained in mutant strains. The seam cells here are identifiable through the SCM::GFP fusion

gained were determined as shown in Figure 2.1.

2.2.2 Lineaging seam cells

We synchronised mutant animals alongside wild type animals and observed them at 10h to check for hatching abnormalities, at 15 to check the first asymmetric division, at 25 to check for the occurrence of the L2 symmetric division, 27 for L2 asymmetric, 36 hours for L3 asymmetric and 42 for L4 asymmetric.

2.2.3 Single molecule mRNA fluorescent in situ hybridisation (smFISH)

Synchronised animals are fixed in 4% formaldehyde dissolved in PBS. smFISH was performed using custom-made probes labelled with Cy5, listed in table 2.3. Z-stacks were taken with 30 slices 0.8μm apart at 100x magnification using an oil objective using a Nikon Ti Eclipse epifluorescence inverted microscope. Each slice was exposed to Dapi, Far red and GFP fluorescence at 100ms, 3s and 300ms respectively and viewed through the NIS Nikon imaging software.

The data analysis involved using MATLAB software to draw ROIs using the Dapi and GFP signal as markers for seam cells and to automatically quantify the mRNA spots by manually setting a brightness threshold. The thresholding was carried out independently for each worm to consider the noise levels for each image separately. The code used for spot detection, thresholding as well as all the statistics is available in the GitHub repository $https : //github.com/Allie3004/smFISH_{C}ode_{A}licja_{P}hD.git$. We opted against using published tools like FISH-Quant in order to manually inspect the images independently, and to maintain consistency with previous work in the lab.

2.2.4 CRISPR-Cas9

The LoxP sites were inserted using CRISPR-CAS9 using guide RNAs to cut two regions in the target gene (in our case *ceh-16*) to create homologous recombination sites between the target gene and the repair template (Table 2.4).

For *ceh-16*$_{LoxP}$ we used sgRNA that were located before exon 1 for the 5' cut and between exon 3 and exon 4 for the 3' cut (Figure 2.2). The injection mix contained:

- 5µl of tracRNA (from 18uM stock)
- 1.4µl of each of two custom sgRNA (shown in (Table 2.4)
- 1µl aliquot of Cas9 (2.5ug)
- 2.2µl of single strand templates (from 25uM solution) (Table 2.4)
- 5ng/µl of myo-2::DsRED
- H_2O to bring final volume to 20µl.

Table 2.3: **List of all smFISH probes used**: We include the specific binding sequences of each smFISH probe we used. For the purpose of this thesis we focus on those with cy5 as a fluorophore.

smFISH Probe List								
elt-1::cy5	egl-18::cy5	ceh-16::cy5	apr-1::cy5	efl-3::cy5	lit-1::cy5	rnt-1::cy5	eff-1::cy5	pop-1::cy5
gttagcatcacgataatgca	cgtcattatgctgatcgaca	gttcaatacccgaatttcaga	gttctgtgtatcgttgtttc	agagccgaagatagttgtca	aaactgtccctcttgatttc	gagcacaggaatgacgtcat	aactggggagaccactcaaaacggaacaccttcacctcat	
agattcactttatttcggga	agcacttcgtggtgttgttg	gggtaggggggaagacaggat	gttgtgagtagattccactt	ccgggtggaatattttcttt	gaactacagccggatgaact	ccatcggattcaatttcagt	aatccgtaggttgtaactgc	ctgttgttctgacgtttcac
ccgacaactccatctaacat	ctacacggctcatctgacgg	tgaaatagttggagatgggc	cgcacattcgatgttctttt	tctggatctgttgttcgatt	ctgttgctgtcgggatataa	atcttcttgcatctggagat	atcgaatttctcctcgagtg	gtctaattctgcttccatta
cattcgtgttctgcatatca	cttctgtaactgtttgcaac	ttgatggacttgtagcagga	tgctcttcttcatcatctac	ctggagatggtcttggaaca	gtattattgggattgcctga	tggacagctgtgaaattaca	tgtcttggaacagtgtggtg	aatcgatggatttcgaccgg
gacggagcaaagagtccaactg	ctgattgtctttgcaaca	ggagatgcaaaagtaggaga	atgttatagcatcgtcaacct	gaagcatcggagagttgtga	cttgcggctgatgatgattg	tctccattagaggtttgttt	gagatgtttgagcacggaca	aaccgggaaacgatggtgac
gatggagtttgtgtcaggat	ttgtccattcgctccataac	cgcaatatttggtgttccat	gacgcaggaagagctttttt	gtcacttctagatcgtcatc	ctgctgaatagtgtggttgt	tttattcgagctctccgttt	gaattgcatttgcattcctc	atccaggactgaagtgtgac
atgttttcctgctcaattgg	ctcgtcgagccgatactgaa	tgaacacccatgcaggatac	gttgatggtttccttcgaag	cccaaactctttttttct	tattacgtcagatcgtggca	gttttagttgtgacagtgac	atctgaagcagactgcagtg	aacatgggtaaagcagctgc
aattcaacgggttttccttc	gctcattgttctctttgagc	ggtcgatcggaatatcttgt	acgattgtgtgaagagctgc	ctgtctcttcattgatagca	agagcaaagtgtggtggttg	cgataccttatgctcatctt	tcattgatctcttgggatgc	gctgcgtatggattcataaa
tccagaagaagtgccgagag	gatgagaccgatgagctttt	ctgaaaagtcgcatacctgc	tagtccagcgtatacttgag	attcattttccttgcaacgg	tgtatgtgaaactagggcca	acttgtttttcggaggtat	atgtctgatttccagcattt	caggcttggtgatcgtagtg
aatggtgcaatggatgcagc	cggtgatggtgaggcttttc	ctactcagagaccttctaga	gtgcatgaactgttggttga	cgtagattcgtcgtttttca	tttttcgtagcacgtcgtag	cctgatctgaagcaatgcta	gttcaagcttttccaatcga	tcattgctcccattggaaac
gttggttgctgatgtattgt	tctcgacaagcttcggagag	agttactcctggagtgtata	tgtatgtgcatgtactgcag	ttgttagttttctgcatggc	agcgaggatcgcatttgaag	taattccggcaacttgccga	aagtgtaccgctgagttatc	gattgctgctccatagacag
tggataactactcggtgtgg	cctgatactggagcgactac	cacttatatccgatgtgatt	ctgagtgaccagcaagattc	aagagactctaatccttgcc	tggagggtagaatggttgtg	cacagccctagttgtaattg	tggcatgaacttcagggatt	aaagtgttgtttggcggctg
cctccagcatattgatagtt	aagtctggaagtggactcgc	gattctcgttttcgggattt	acttcagaattccagtagct	ctctggaagaccttcttcaa	tctggctgactatcttgtac	ttttttctaaaaccggccgt	tggcatcacactcacagata	atgttcagtggagccatatt
atgcatccatatcagttgtc	ggatcaaacatgaatccgtt	tgaacttccactacttccag	cgtggaagtattcgcgagat	cacgtggtgatgacaactca	atgcaccataaccaataggc	acaacatacccctcggaaaa	agattctgcggtacatgttg	tctgattcattggatgaccg
attgacatatccattccact	gcatcattccatttggattt	tacgtggtctttttttcttct	gagtgatgagcaagcttctt	gagccgacaatatccttgaa	gttactgaccatacaactcc	aactcggaaattgtccgccg	agactctggacaagcggtaa	ttcgatcgacattctgtgga
cggattgttttgtgttccaa	ctcacggattgttgattctc	ccaattgatctccagtgaat	acaagtgaacgaggcactct	ggtcggtgttgaagtacatg	cacgttgggcatttttttga	gatgaagtatagcgagtggc	gcggtagcatgaagatgttt	tgaacgcgtttaatggcttc
gagttgtgtcgttggataga	cacggatttcgattgtggtg	cggaattcagttttgagccg	attaggtttgcaatagctcc	acgagtgtcaactctcgatc	gatttctcgaaacactcgct	gagtttggccagtcattata	tggtgtctgatttgggaaga	aatctcttccagcagtgctt
gcgagtgtatcatatccata	gatccattggatcttcaatt	gtaatcatgattacctgctt	tcacgattcgagtttcggag	tcaaaaaccgtctgcacagc	aaaaagctcggattcgcggg	ttcacggcggcggacaattt	tcgaacgtcacagcaaagct	gcactctgtttctcattatt
cagtaattccagatgctgtc	gactcttcctgtttcacatc	tttgccgagttaggacatta	gcagttccatgataagcatg	cttttcggatattcttcggg	ataatcttgtgtaggtcgct	aattttccgaggggtatgtt	gtttgactgcgaggaatgtc	tccaagctctttattcaact
tacgatagttccatttcctt	agtgcatccaaaaggttctg	gttgttgctggtaatcggaa	gtggagtctgtatcaactgt	aactgtgctagctacatcga	tgtcaaagcttgcggggaaa	ggtcgtccggttttagaaaa	atgttgcatacgttgtaggt	tcgacaaatcatgccacctc
tgttttgagtgattggctgt	gttgctgctaaactgtgctg	ccgcttttctgttagatatc	aatccatttcgttgaacggc	aaaaccttctgtttccggat	ggatgtttgcagtgtgcaaa	tataagagctgtgcaaccgg	tttatctttttccacccaat	aaatatttcgcctgttcttc
ggtgagaggttgactgttat	tggtggaggtctggaagaacagctcgtgagccaattcttg		gttgttgttgctgcaattga	tataagcgtctgcatcgact	aagagatttcccggtttgat	tcaaacctgcctacggaata	tgtgttccaccatctaattg	cttatccttctttgccattt
gatgatcctgaagacgtcga	gatgatgatgatggtggaga	gatctgcgattcgttgagtc	ctgagcacacattgttggag	aaccaggacattggcaatgt	ctctatagtactgggtcaca	atctctcgcgctacttaaaa	cgacgtttttggtcgagatg	gataccgttccttgtgtgtt
gagtttgctgatgagcttga	ttcgatgaccgcttgtactt	gcttgttctggaaccagatt	tccgattcagtggatttagg	acctttttaatgagcccaag	aaccgacggaccaaatatcc	tgcacgttattcgcgaattc	ggcagttacagccaatgaaa	ttgttaactccgaatctggc
ggtgttcttcggtgtagaag	ttgaactgtcttggctttgg	gtcgacttttttgagtttggc	aatgtgtagtgttctgctgt	cgggattttcttttgttccaa	tgtagcagctcggcaaatat	gatttttttacaacttgcca	cagttgatgagatgctcgtc	gcttgcagaatttacaccac
atgatcggtttgctttggag	ctcgtctcggtttttctcaaa	gcatcgatctcttggaacag	actccgagtcatagtgttag	ggttcaggaccacagtaaac	ccggcggcttgaaatagaat	aatagaggcggaggctatta	ctccattactttgttcttgag	aacgatcagttgcgtactcg
atcttccgtgctgaattgag	acggattgcagacaagcttc	tgaggatttggagtgacgga	tgtctgttcgtgtaccaata	aaacatcgaagcttccggtt	cgatggggttcccaataaat	tggaatacccaaaagaagcctcaatggttgcattctcagt		cctgactgtcagttatatcg
ctccacaattgacacactca	gtgcaatcgatagtagagcc	cctccatgaattgacggatg	tcttctgaagtctttcgtgg	gtgaactactcagaaggcgt	cacaagcgtatttcatagcc		accatccaagacggtcaaag	tactgctatacgcaccactt
cgacgccataatggagtatt	atttctattggtcggcgaac	ccaactgagccataagttga	acctgggagaacatcatcag	caattttgtcagtctgtgct	gcccttcttctaaatatcga		atgaccagaatcgtccattc	tgctgtgtggtcagtgcaat
atgcgttgcagaggtagttt	ttgttggatgtggtttttgc	atgtaagctcgagcctgaac	gtcaaaagctcgtcatctgt	ttttgttctgagagcgttct	catacacgagtggaatcgca		tccattttcacaactccatt	catcgtatgcatcattgctg
ttcatcttgaagtagaggcc	actctttttccttgttcttt		atcttcggagtagacgacga	tgcgatttttggagtattcc	gtttcgtatagcagcacgag		gcaatttttcactttagcct	atcctagacgcatttgcatc
ttcaccaatggacgagcatg	tggttgaagatctgtgttgc		atgcaggagacgatgttgtt	tggtgatgatctgacagctg	aacaacctcgacggcatatt		tcgagcagattgaatccacg	tgatggaacgtgcgtcgatg
tttttgagcgttctgctgtc	atcgtcggcatttgtgtgag		gcatcagaagcatcttgacg	ccacctctgcgaacataaac	atcgatagtcgggacatgtc		gtgttacaacagcttgtctg	cctcaacatcagattccgaa
acgacagttgacacactcga	ttgaatgtgttgatggctcc		gcttcctcgtttttgaatg	tatctcattttctcagctgc	aaagggcgacaccgtacaac		tgaagattagttccttcggc	taaccgtcggatcaatctgc
atttcttctccagagtgttg	ctgcgtgaattgcgagattt		gagatgcatcgatctcatcg	attgatgctactgatcccaa	gcaactgacgagcttgcaaa		gacaaggttttgactttcca	aagcatatcatgtgtctgct
acaagcattgcacactggat	tgctattcatgagctcttga		agatcctttggtgtggaag	gatgtcattggatgctgtga	gatggaggtaattctgaggc		gatggatccactgaagtcac	tagtacacatcgattcctgc
cgttccaccttatgaagttt			gaactcctgggagtgaagat	ggaagttgaagaggagccat			agcctcatatactgtcaagt	
atccttcttcatggtgattg			tcttgactttccattcgttg	ggctttggaatctctggaag			ctgatccatcaatttttcca	
ttgcgattacgggtctgaat			agtgatccattcgtttttgg	tgtaaactgcttttgagcct			tccaaatccagttgacatct	
tttcttcattcttcgcgatc			tgattggcttgggattctta	attgaatcgaggcctagacg			atggctggcagtggaataat	
gaagatgttgtaggcattcc			ttaaaagtgtcggcttcggt	ttgcacaggtgtatagtcag			tgaatctgctcggagacaga	
ccactttgatcaagttcgat			caatcgaatcctcttctgga	agatacgagtgggcgtattg			tttccaaagggctctcgaaa	
agtgtttttcattccccaaa			gtatcatctctcgtctgatt	gttgcgaagattcttgtgga			ccatttttccctcaacaagat	
tgtcggtgtcatgagcattg			aaattcgttcttgttctgcc	aatgttttggcttgacgtct			attacaagttgggcaggttc	
tggatgctgggaatgcatat			tgacggtgattgctcaatgt	cttcgattcgccaagaatgt			gatccaatgagctggattca	
tcctcgatcgagttgaagta			tagggcgttacaattgcaga	tgtgctcgaacgtattctga			caccaatcgaactgattcct	
ggcacgttttatactccaat			atttcgaccagtgaaaggcg	cacctggaatgcggaagatg			atccgtagcaatcataacgc	
tccaaagtccaccatcattg			ttgtttaggattcggtgtca	ctccgaagacttttttggt			gacaaatttgtcccaacgga	

Table 2.4: **List of all primers, sgRNA and repair templates used**: We include the names used in the text, and the sequence of all the primers, sgRNA and repair templates used.

Primer List	
Primer	Sequence
ceh-16 5'F	tttgcggttaccagagatcg
ceh-16 5'R	aatagtgcaaaaccccagtaaaa
ceh-16 3'F	aaaacgagaatcaactggaagt
ceh-16 3'R	agaatctcctgaggttaaaccattt
ceh-16(bp323)F	caggccctcgacatcgaaaa
ceh-16(bp323)R	gtgagccaattcttgtcgcc
egl-18 ga97 F	caacaccaccagcacattga
egl-18 ga97 R	tccgttgctgctaaactgtg
M13F 43	agggttttcccagtcacgacgtt
M13R 49	cctgtgtgaaattgttatccgctc
sgRNA	Sequence
ceh-16 guide RNA1	gtaattaggtgtgaagaagg
ceh-16 guide RNA 2	tcttgtatcacgaatttgtt
Repair Template	Sequence
ceh-16 5' Repair Template	ctctttttactccctctctccttcacctcattcactt gtcacccgtaattaggtgtgaagaaggcgg agggttttcccagtcacgacgttcccgggat aacttcgtatagcatacattatacgaagttat gaaaagaggcggagcttccttagagcggg gatggacgatgagg
ceh-16 3' Repair Template	tcataaaatataataaatcctttcatcttgtatc acgaatttgttagggagcggataacaatttc acacaggcccgggataacttcgtatagcat acattatacgaagttattagaactcggaaat gtgtacggttttcaatgataaatcctga

This mix was administered by microinjection. 2µl of the injection mix was loaded into the injection needle. 3 young adults were placed in a drop of hydrocarbon oil on a 2% dried agarose pad and injected. These animals were then transferred to an NGM plate in a droplet of M9 buffer. The animals were then allowed to lay one generation of animals and then those animals were allowed to lay eggs and then were lysed to extract genomic DNA.

2.2.5 Worm lysis and PCR

Worm lysis was carried out by placing the a singular animal in 8μl of worm lysis buffer containing 0.2 $\mu g/\mu l$ of Proteinase K. This was then incubated for an hour at 65℃ and then 15 mins at 95℃. 1μl of this solution was then used in a 50 μl PCR reaction. Diagnostics were carried out using Gotaq® DNA Polymerase from Promega with extension time of 1kb/min. Primer annealing temperature was set to 56℃ unless otherwise stated. If sequencing was carried out the primers used are listed in Figure 2.4 and purified using the Promega PCR clean-up kit. Sanger sequencing was sent to eurofins genomics (https:// www.eurofinsgenomics.eu). The sequencing was analysed using ape and benchling.

2.3 Mathematical methods

2.3.1 Statistics

Several different statistical methods have been used over the course of this thesis. Mann Whitney U was used to test for significance of smFISH, T-tests were used to determine changes in seam cell number, Levene's tests were used to determine significance of variability. P-values in graphs were denoted as * = p $<$0.05, ** = p $<$0.01, *** = p $<$0.001, **** = p $<$0.0001. The code for all of our statistics calculations is available at $https://github.com/Allie3004/smFISH_Code_Alicja_PhD.git$.

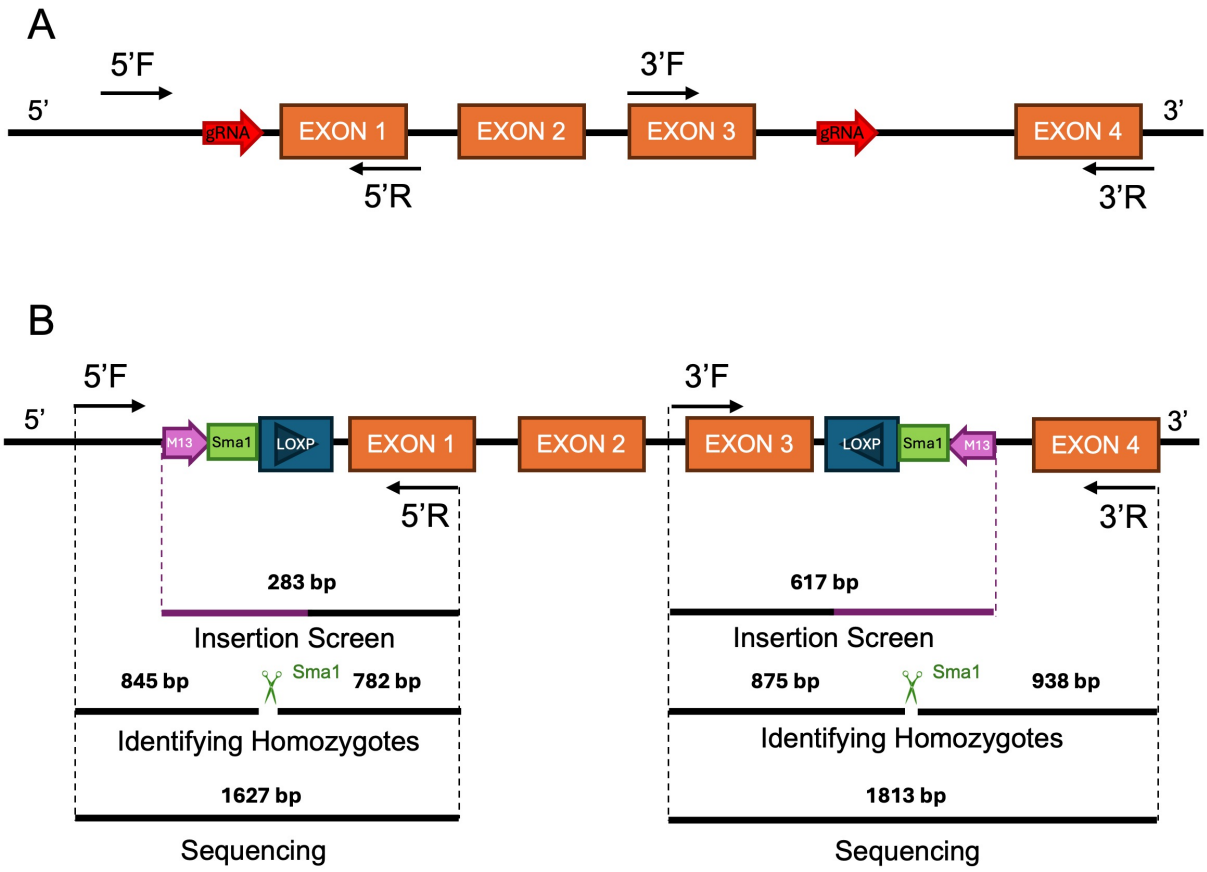


Figure 2.2: **Diagram showing the layout of the LoxP insertions in *ceh-16***: **A)**We show the location of the guide RNA complimentarity within *ceh-16*. **B)** We also show the location of the LoxP sites and the primers used to screen them both for the initial insertion as well as for homozygosity.

2.3.2 MRA

Together with masters student Arianna Ceccarelli we implemented the Modular Response Analysis algorithm by (Kholodenko et al., 2002) and adapted it by multiplying the amount of available protein by its expected functionality (0.25 for partial loss of function mutants and 0.01 for complete loss of function mutants). The algorithm is applied to bootstrapped data following the formula:

$$R_p = [R_{ij}]_{ij},$$

where

$$R_{ij} = 2\frac{x_i^{(j)} f_{ij} - x_i^{(0)}}{x_i^{(j)} f_{ij} + x_i^{(0)}}$$

and

$$f_{ij} = \begin{cases} 1 & if \quad i \neq j \\ \gamma_i & if \quad i = j \end{cases}$$

$x_i^{(0)}$ describes the mean expression of gene i in wild type and $x_i^{(j)}$ is the same gene's mean expression in the mutant j. R_{ij} represents the difference between transcript levels, $x_i^{(j)} f_{ij} - x_i^{(0)}$, divided by the mean $\left(x_i^{(j)} f_{ij} + x_i^{(0)}\right)/2$. f_{ij} indicates the fraction of activity of the protein carrying out the regulatory function. It is 1 in a non-mutant context and γ_i in a mutant context. γ_i is 0.25 for *elt-1* in *elt-1(ku491)* and *ceh-16* for *ceh-16(bp323)*, whereas for *egl-18* in *egl-18(ga97)* it is 0.01. Further details can be found in Brożek et al. (2023).

2.3.3 ODE

The ordinary differential equations were derived using Hill equations (Hill, 1910) to denote the interactions of the other genes of the network on each specific gene. The process of deriving ODEs for GRNs is described in many different pieces of literature such as Hale (2009); Ando et al. (2002); Delgado and Gómez-Vela (2019).

2.3.4 SLInG

The SLInG algorithm was implemented following the process illustrated in (Jørgensen et al., 2023) by Arianna Ceccarelli and Andreas Christ Sølvsten Jørgensen. The priors were based around the range of values provided for each interaction by the MRA. The δ_{ABC} described in (Jørgensen et al., 2023) was set to 0.0125 which would be applied if the distance measure of the best-fitting model within the burn-in divided by 0.0125 is lower than 35. However if not the δ_{ABC} is reset by the model itself to encourage convergence.

Chapter 3

A quantitative Gene Regulatory Network underlying seam cell patterning

3.1 Introduction

The work in this chapter was part of a project developed together with MSci student Arianna Ceccarelli, Andreas Christ Sølvsten Jørgensen and Mark Hintze, and is showcased in the paper (Brożek et al., 2023). My role was the generation of the data and its analysis as well as the testing of the newly developed models through the generation and analysis of the *ceh-16(bp323);egl-18(ga97)* strain. I was involved as Arianna Ceccarelli's supervisor in the modification of the MRA and the development of ODEs.

Despite the breadth of knowledge about the behaviour of the *C. elegans* seam cells and the robustness of their behaviour no study has actually delved into understanding the links between the GRN of these highly important cells throughout larval development. The focus in the GRN community has largely been on building GRNs in human and mouse stem cells. While this is a highly valuable approach, leaving behind model organisms' GRNs does result in underuse of a very important tool. The robustness underlying the behaviour of *C. elegans* seam cells makes it the perfect candidate in which to study and model robust stem cell behaviours, however in order to do this we must understand the basics of their GRN. As such we selected three genes as a starting point of our network. These were *elt-1*, *egl-18* and

ceh-16 (Baugh et al., 2005; Smith et al., 2005; Page et al., 1997a; Gorrepati et al., 2013; Koh and Rothman, 2001; Koh et al., 2002; Huang et al., 2009).

Given the small number of genes we were working with, we decided to take on a quantitative approach to construct this network. We started by utilising the MRA (Kholodenko et al., 2002). This mathematical approach has been utilised for the inference of a variety of different networks (Alsina et al., 2014; Santra et al., 2013; Klinger and Blüthgen, 2018), and has even been integrated with bayesian approached for the inference of the dynamics of a GRN (Santra et al., 2018). However, one aspect that this algorithm lacked is adaptation for use with loss-of-function mutants rather than knockdowns. This is important when working with essential genes whose full knockdowns cause lethality. MRAs can be used to inform parameters of ODEs (Santra et al., 2018). However, this is greatly limiting as, ODEs provide a single numerical solution and thus formulation of individual ODEs to evaluate the validity of different parameters would be slow. As such it is important to incorporate GRN inference methods that allow for a parameter search. This is frequently based on Markov Chain Monte Carlo methods that test different parameters and evaluate whether it improves the model relative to the data (Hastie and Green, 2012; Ram and Chetty, 2009).

A novel approach for GRN inference called SLING (Sparse Likelihood-free Inference using Gibbs) was described by Jørgensen et al. (2023). This approach relies on a bayesian framework and Gibbs sampling to identify sparsity inducing priors to find the simplest model that is able to explain the behaviour of a specific dataset. While this approach has been tested on synthetic data (Jørgensen et al., 2023), it remains to be tested on a real dataset.

The aims for this chapter was to construct a detailed quantitative model of the core seam cell genes at a time point where the cells were not actively dividing. We aimed to test this network mathematically and then utilise the newly developed SLING algorithm to eliminate non-essential interactions from our model. We then aimed to use the minimal model to formulate predictions on the response of these genes to further perturbations. We tested these predictions on a double mutant of the network specifically the *ceh-16(bp323);egl-18(ga97)* and used our results to refine the network.

3.2 Results

3.2.1 Gene expression analysis using smFISH data identifies new interactions between core seam cell TFs

In order to elucidate the interrelationship between ELT-1, EGL-18 and CEH-16 which, due to their importance in seam cell maintenance and function, we consider the core seam cell TFS, we compared these core genes' expression profiles in wild-type nematodes to those of corresponding loss-of-function mutants. Given the lethality associated with complete loss of either *elt-1* or *ceh-16*, we utilized partial loss-of-function mutants for these two genes. Although these mutants exhibited severe structural and growth defects, they remained viable, enabling genetic studies. Notably, these hypomorphs exhibited a loss-of-seam-cell phenotype almost comparable to that of the complete loss-of-function mutant *egl-18(ga97)* (Eisenmann and Kim, 2000) (Figure 3.1A).

As this study marked the first endeavor to quantify the relationship among these three core genes in *C. elegans*, we opted to commence our observations from a time point we believed to be representative of a steady state. Specifically we selected the 17 hour time point, after all the cells have completed their L1 divisions but before the cells carry out the subsequent expansion division increasing their number from 10 to 16. Our focus was directed toward V1-V4 seam cells due to their close positional proximity along the nematode's body and their highly similar behavior (Figure 7.1), particularly in *egl-18* and *ceh-16* expression. To quantify the gene expression levels in these animals, we employed single-molecule fluorescence in situ hybridization (smFISH) (Figure 3.1 B-C).

elt-1 exhibited no changes in expression levels in from the wild type in both the *egl-18(ga97)* as well as the *ceh-16(bp323)* (Figure 3.1 C). Similarly, there were no changes in expression observed in its own mutant *elt-1(ku491)* (Figure 3.1 C). This mutant also didn't affect the expression of *egl-18* relative to the wild type and neither did *egl-18(ga97)* (Figure 3.1 C). However, a strong increase in *egl-18* expression was observed in the *ceh-16(bp323)* mutant indicating a significant regulatory role of *ceh-16* on *egl-18* expression (Figure 3.1 C).

Furthermore, *ceh-16* expression levels were markedly decreased in all mutants compared to the wild type, particularly evident in the *egl-18(ga97)* and *elt-1(ku491)* mutants, although a notable response was also observed in the *ceh-16(bp323)* mutant (Figure 3.1C). Collectively, these findings suggest that *elt-1* likely acts upstream of the other two genes. We also hypothesise that CEH-16 is a likely repressor of *egl-18* and that the expression of *ceh-16* is activated by both EGL-18 and ELT-1 with both being required for its expression.

3.2.2 Constructing a gene regulatory network using Modular Response Analysis

The naïve network proposed in (Thompson et al., 2016), depicting ELT-1 activating *ceh-16* and *egl-18* and CEH-16 activating *egl-18*, had very significant disparities compared to our smFISH data. To ascertain the strengths, directions, and significance of the new interactions in the core seam cell GRN, we employed statistical inference and mathematical modeling. Our first approach was to utilise the Modular Response Analysis (MRA) to estimate the strengths of all potential links between the genes. Despite the original MRA being designed for low noise, low perturbation response matrices, the work of Borg et al. (2023) has shown that when bootstrapped data with a 95% confidence interval is input into the MRA, it performs well even when the perturbations are complete knock outs even in 10 gene networks. Due to our use of loss-of-function mutants (hypomorphs *elt-1(ku491)*, *ceh-16(bp323)* and the complete loss of function mutant *egl-18(ga97)*) and the original MRA being based on knockdowns, where the quantity of protein corresponded to the amount of functional protein, we had to modify the existing algorithm found in (Kholodenko et al., 2002). Since the transcriptional levels of genes in loss-of-function mutants do not necessarily correlate with the amount of functional protein, as presumed in the original MRA, adjustments were necessary. Thus, we estimated the percentage of function retained in each gene's hypomorphs relative to the wild type. Partial loss-of-function mutants were estimated to function at 25%, considering that heterozygotes exhibited wild-type behavior, while complete loss-of-function mutants were assumed to function at 0.1% capacity. The output of the MRA is an interaction matrix where

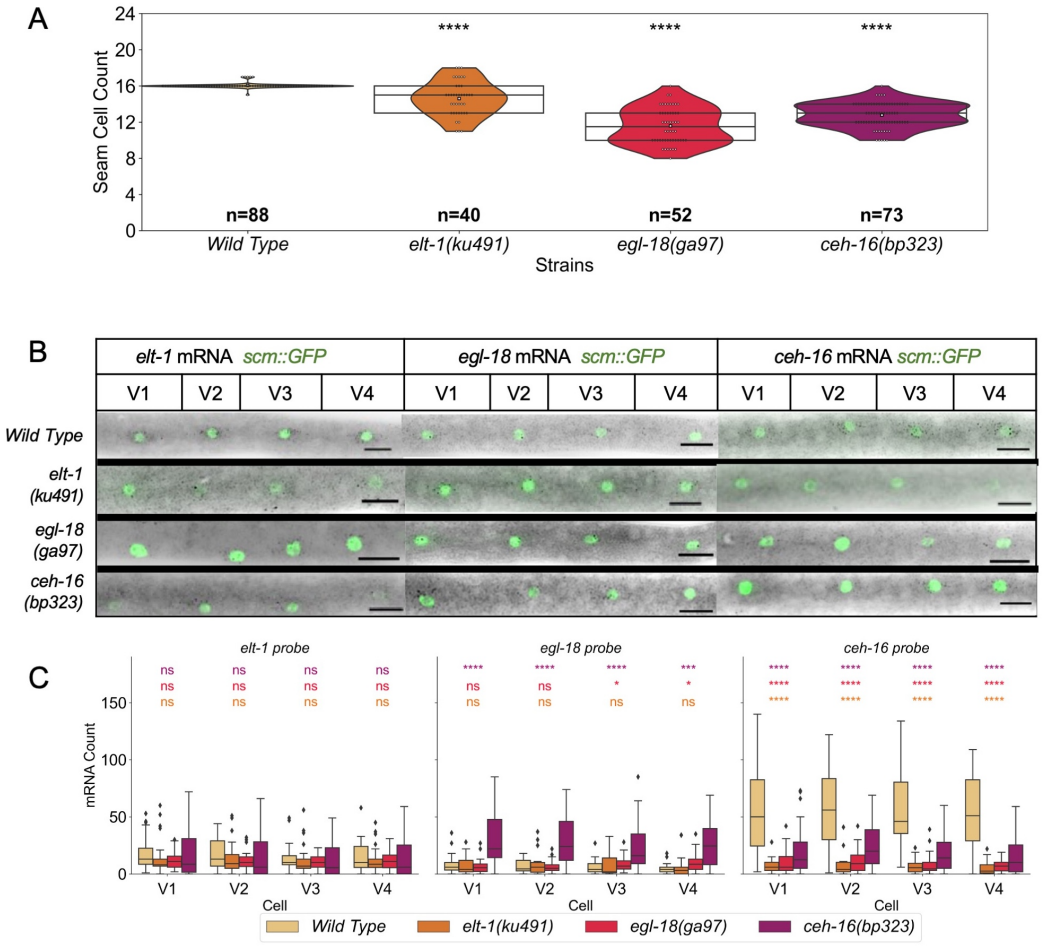


Figure 3.1: **Phenotypic and gene expression analysis in wild type and TF mutant backgrounds: A)** Seam cell number comparison at the end of the L4 larval stage between the wild type and core seam cell gene mutant animals. The lines of the box represent the quartiles of the data, and the white square in the middle represents the median. The p-values are calculated using the Z-test and are shown as described in methods. **B)** Representative smFISH images in wild type and TF mutant backgrounds. The mRNAs are visualised as black spots, and the nucleus of the seam cells appears green because all animals carry SCM::GFP in the background. **C)** Quantification of smFISH signal in V1-V4 seam cells. The significance of the changes between mutant and wild type backgrounds was calculated using the Mann-Whitney U test and is presented as described in methods. Strains are colour-coded as indicated in the key underneath panel C. We show quantifications of *elt-1* (in wild type (n = 108), *elt-1(ku491)* (n = 100), *egl-18(ga97)* (n = 100), *ceh-16(bp323)* (n = 104)), *egl-18* (in wild type (n = 108), *elt-1(ku491)* (n = 224), *egl-18(ga97)* (n = 60), *ceh-16(bp323)* (n = 104)) and *ceh-16* (in wild type (n = 153), *elt-1(ku491)* (n = 80), *egl-18(ga97)* (n = 87), *ceh-16(bp323)* (n = 120))

each column indicates a gene carrying out a function and each row indicates the gene being acted upon. The value in the cell at the overlap of a row and a column indicates the strength and polarity (positive for activatory, negative for repressive) of the specific interactions. The analysis, applied MRA to bootstrapped V1-V4 smFISH data, yielding a range of coefficients (Figure 3.2A). These coefficients were compared with the interaction map from the naïve model (Figure 3.2B) and average coefficient values were utilized to construct the interaction network (Figure 3.2C). Some of the interactions we have uncovered are consistent with the naïve model, notably ELT-1 exhibits very strong positive interactions with both *ceh-16* and *egl-18*. However, contrary to the naïve model, CEH-16 appeared to strongly repress *egl-18*, and a novel positive interaction from EGL-18 to *ceh-16* was identified. There are two additional interactions that the analysis produced: a weak repression of *elt-1* by CEH-16 and a weak activation of *elt-1* by EGL-18. These are however very weak and the error bars of the coefficients produced by the MRA do overlap 0. In summary, our predictions suggest that *ceh-16* and *egl-18* are downstream of ELT-1, which is consistent with previous findings (Huang et al., 2009; Koh and Rothman, 2001). Furthermore, CEH-16 and EGL-18 appear to regulate each other's transcription through a negative feedback loop that hasn't previously been observed.

3.2.3 Using the MRA results to infer a set of ODEs to explain expression patterns of the core seam cell genes

The MRA facilitated the construction of a model encompassing both strong and weak interactions. However, to determine which interactions were essential for the observed behavior of the gene network at an assumed steady state, we deemed it necessary to develop a minimal model. In order to understand which interactions could be eliminated we devised a set of Ordinary Differential Equations (ODEs) representing each gene's expression levels, given the expression quantities of the other genes in the network. Since the MRA does not generate self-regulatory loops, we included our strongest candidate, based on changes in expression levels, for such a loop, namely *ceh-16*. Additionally, we accounted for all potential

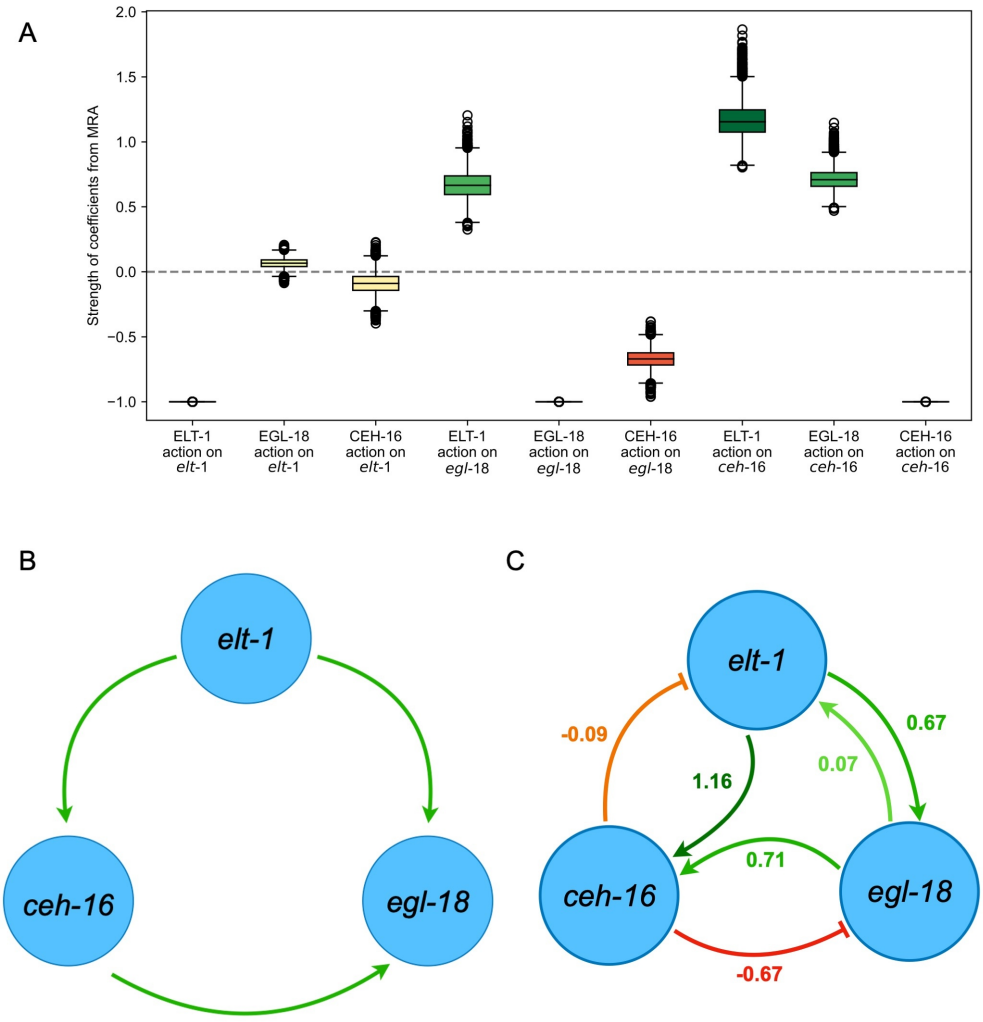


Figure 3.2: **Using MRA to construct a seam cell gene network: A)** Boxplots of network interaction map coefficients from the MRA performed on the bootstrapped smFISH data of V1-V4 seam cells. A positive number indicates activation, a negative number repression, and 0 indicates a lack of an effect. The coefficients of a gene acting on itself are set to -1 as default in the MRA algorithm. **B)** The model proposed in the literature that we use as a benchmark that must be improved upon. **C)** Network summarising the results of the MRA algorithm using the average of each coefficient value shown in the boxplots in panel A. Green arrows represent activation, while orange/red blunt-end arrows represent repression.

forms of co-regulation by incorporating AND and OR gates for each interaction, represented by multiplications ($\cdot$) or additions (+), respectively. We examined all possible combinations, resulting in 20 different equations. To denote OR or AND links, we used the $\bigoplus$ symbol. We denoted *elt-1*, *egl-18*, and *ceh-16* as 1, 2, and 3, respectively, with g1, g2, and g3 representing the quantity of each gene.

The equations (3.1), (3.2) and (3.3) were used to describe the change in expression of *elt-1*, *egl-18* and *ceh-16* respectively, and were derived as:

$$\frac{dg_1}{dt} = b_1 + v_{1a2}\left(\frac{(g_2f_2)^{\alpha_2}}{k_2^{\alpha_2}+(g_2f_2)^{\alpha_2}}\right)\bigoplus v_{1r3}\left(\frac{k_3^{\alpha_3}}{k_3^{\alpha_3}+(g_3f_3)^{\alpha_3}}\right) - d_1g_1, \tag{3.1}$$

$$\frac{dg_2}{dt} = b_2 + v_{2a1}\left(\frac{(g_1f_1)^{\alpha_1}}{k_1^{\alpha_1}+(g_1f_1)^{\alpha_1}}\right)\bigoplus v_{2r3}\left(\frac{k_3^{\alpha_3}}{k_3^{\alpha_3}+(g_3f_3)^{\alpha_3}}\right) - d_2g_2, \tag{3.2}$$

$$\frac{dg_3}{dt} = b_3 + v_{3a1}\left(\frac{(g_1f_1)^{\alpha_1}}{k_1^{\alpha_1}+(g_1f_1)^{\alpha_1}}\right)\bigoplus v_{3a2}\left(\frac{(g_2f_2)^{\alpha_2}}{k_2^{\alpha_2}+(g_2f_2)^{\alpha_2}}\right)\bigoplus v_{3a3}\left(\frac{(g_3f_3)^{\alpha_3}}{k_{3s}^{\alpha_3}+(g_3f_3)^{\alpha_3}}\right) - d_3g_3, \tag{3.3}$$

where

$$f_1 = \begin{cases} \gamma_1 = 0.25 & \textit{elt-1} \text{ mutant } (\textit{ku491}) \\ 1 & \text{otherwise} \end{cases}$$

$$f_2 = \begin{cases} \gamma_2 = 0.01 & \textit{egl-18} \text{ mutant } (\textit{ga97}) \\ 1 & \text{otherwise} \end{cases}$$

$$f_3 = \begin{cases} \gamma_3 = 0.25 & \textit{ceh-16} \text{ mutant } (\textit{bp323}) \\ 1 & \text{otherwise} \end{cases}$$

The complete set of parameters utilized in these equations is detailed in Table S7.1A. Assuming that the late L1 time point represents a steady state, we set the degradation rates d_1, d_2, and d_3 to 1, as we are not considering the system's progression over time. Activatory and repressive actions on each gene were modeled using Hill functions, with each action multiplied by the fraction of fully functional protein involved (denoted as f_1, f_2, f_3), which is 1 in the wild type, 0.25 in the partial loss-of-function mutants, and 0.01 in the strong loss-of-function mutants. While these may be considered strong perturbations, and the MRA was designed around weak perturbations, previous studies have relied on full knock outs as perturbations to inform the MRA and have shown this approach to be successful (Borg et al., 2023). Interactions onto the same gene were linked with either AND or OR to indicate co-regulation, with actions on each gene either added or multiplied, respectively. This formulation resulted in 20 equation sets labeled 1-20, with their AND/OR links described in Table S7.1B. The initial set of equations assumed only independent links denoted by OR, while the final set assumed full co-regulation with all connections denoted by AND.

In these models we included all potential interactions suggested by the MRA. As each Ordinary Differential Equation (ODE) solution yielded a single numerical output, we sought a modeling approach allowing parameter search to optimize the strengths of network interactions and achieve a minimum number of required connections. In order to do this we employed the recently developed Bayesian model discovery approach, SLING. SLING aims to find the simplest possible model by fitting and optimising parameters of the model while also setting all non-necessary parameters to zero (see Jørgensen et al., 2023, and Methods for more details). This approach prevents the formation of unnecessarily complex models and avoids over-fitting. Notably, it does not determine whether links exist but rather assesses whether they are necessary for the model to explain the data and make accurate predictions. SLING simulates various parameter landscapes and compares model outputs to summary statistics of the data using a distance measure. We relied on the mean expression of each of the three TFs under wild type and mutant contexts and compared them to the numerical solutions of the ODEs with different parameters using a normalised Euclidean distance.

The parameters investigated in the model included b, v, and k, where b represents the

basal production rate of a gene, v denotes the rate of gene production given the action of another gene in the network, and k indicates the amount of a gene required to achieve half of its maximal activation or repression. Each of the 20 models described in Table S7.1B underwent 3 independent Markov Chain Monte Carlo runs using SLING. The presence of each parameter was quantified as a proportion of the runs in which it was kept (Figure 3.3A), which provided the relative importance of each interaction as described in (Figure 3.3B).

ceh-16 exhibited a significant dependence on activation from other genes ($>$80% of the Markov Monte Carlo Chain runs kept these links) or self-activation (kept in 25% of the runs) and thus relied less on basal production which was kept in none of the simulations (Figure 3.2 B). Conversely, *elt-1* demonstrated low dependence on activation by other genes (kept in $\leq$ 10% runs)but showed a substantial reliance on basal production, which was maintained in 97% of the runs (Figure 3.2 B). *egl-18* is highly regulated by the other two genes, however some of this regulation is repressive and thus while this gene shows reliance on basal production it was retained in only a few of the simulations, with the b_2 parameter being kept as non-zero in 18% of the runs (Figure 3.2 A).

To assess links previously suggested by the MRA, we compared our model's behavior to the data and to the output of the naïve model proposed by the literature (Figure 3.2B). The output of the naïve model (shown as a black dashed line in Figure 3.3C) showed a large distance from the data (dashed red line in Figure 3.3C). The data itself had a non-zero distance measure due to variability stemming from the data and bootstrapping. Our objective was to develop models that outperformed the naïve model, but had an error measure higher than the data itself to avoid over-fitting. All of our models met these criteria for all three of their Markov Chain Monte Carlo runs (Figure 3.3 C). Model 11 (with actions on *egl-18* and *elt-1* linked with OR, but actions on *ceh-16* linked with AND) and model 16 (with actions on *elt-1* and *ceh-16* linked with AND, but actions on *egl-18* linked with OR) exhibited the lowest distance measure in all three runs. However, this could mean that the models fit too closely to the data and thus could result in over-fitting. As such we can consider all the models to be equivalent.

This allowed us to calculate the frequencies with which each parameter was retained in all

of the runs for all of the models (Figure 3.3D). The connection that was deemed necessary in all the models and all the Monte Carlo runs was the activation of *ceh-16* by EGL-18. Similarly the repression of *egl-18* by CEH-16 was kept in 92% of the runs, underscoring the importance of the negative feedback loop between *ceh-16* and *egl-18* predicted from our data. The activation of *ceh-16* by ELT-1 is also conserved in the majority of the runs (87%). Conversely, despite the MRA prediction that the activation of *egl-18* by ELT-1 is as strong as the repression of *egl-18* by CEH-16, this interaction was only preserved in 47% of the models, indicating that whilst it is a strong interaction, it is uncertain whether it is necessary for the function of the network at steady state. Similarly the self-activation of *ceh-16* is only kept in 35% of the runs. The two actions on *elt-1* were described as weak by the MRA and were also considered necessary in very few models ($\leq$10%).

Overall using a combination of smFISH data, a modified MRA and the novel SLING algorithm, we have developed a GRN for the core seam cell genes. This network however requires testing at both the mathematical and the biological level.

3.2.4 Mathematical testing of the newly developed GRN

Before applying our model to new data, we first assessed several underlying assumptions. The first was that there was no cooperative binding to the same site between the proteins and thus the Hill coefficients of each interaction within each ODE (α) could be set to 1. To explore the impact of cooperativity, we tested models with α set to 3. For simplicity, we focused on two models rather than all 20: model 1 (OR interactions only) and model 20 (AND interactions only). We repeated our previous SLING approach and ran each Monte Carlo Markov Chain 3 times. This resulted in minimal parameters variation (Figure 3.4A). Interestingly, simulations with α = 3 showed negligible deviation from the data and often exhibited a tighter fit than the data's inherent error (Figure 3.4B). As such we decided that α = 1 was less prone to over-fitting.

Next, we examined whether the MRA's predicted interaction types were optimal. We allowed the model to consider each interaction as both repression and activation, selecting the one that best fit the data. We simulated this in the same way as we did when testing

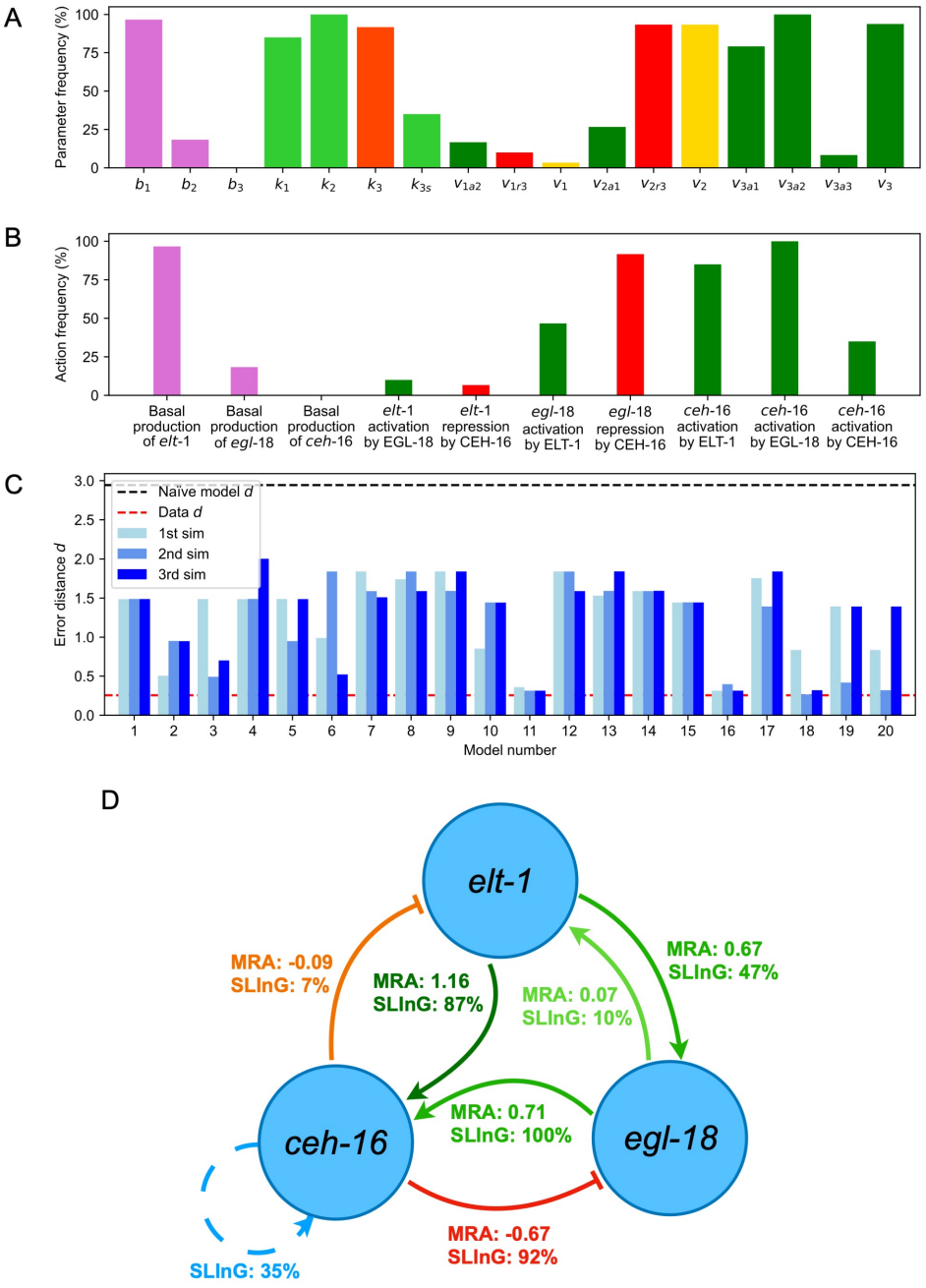


Figure 3.3: **Model fitting using SLInG: A)** Bar graph showing the presence of each parameter in the final models, computed as the percentage of all SLInG runs **B)** The presence of each effect (in percent) in the final models is shown. These are derived from the combinations of parameter presence in each SLInG run. **C)** Distance measure of the median parameter set d, grouped by model type (3 SLInG runs for each). The red line denoting the data value indicates one standard deviation of the original data. **D)** GRN summarising the MRA and SLInG results. The percentages for the SLInG reflect the presence of actions in final models.

the different αs and then checked whether any interactions tended to change signs between activations and repressions of the v parameters in the equations (Figure 3.4C). This analysis, denoted as a/r, revealed minimal changes in interaction signs, with only the *elt-1* repression by CEH-16 showing occasional switching, albeit being weak and frequently disregarded in the models (Figure 3.4C). Additionally, we assessed the model's performance with relaxed sign constraints but observed no significant difference compared to standard runs. As no strong interactions changed signs and the models performed equally well under both constrained and unconstrained conditions, we concluded that our chosen interaction types were appropriate, suggesting consistency between the MRA and the SLInG algorithm in identifying interactions from our dataset.

3.2.5 Biological validation of new model using the *ceh-16(bp323);egl-18(ga97)* mutant

The GRN derived using the SLInG algorithm and the MRA provided insights into the behaviours of the core genes, enabling us to make predictions about their responses under different conditions. To assess the network, we investigated the simultaneous loss-of-function of two genes. Given our strongest predictions were centered around *ceh-16* and *egl-18* we crossed the *egl-18(ga97)* mutant to the *ceh-16(bp323)* mutant. The resulting *ceh-16(bp323);egl-18(ga97)* strain showed a highly consistent phenotype with an average count of $\approx$ 11 seam cells at the end of the L4 larval stage (Figure 3.5A). When compared to the wild type using the Levene's test we did note that it was significantly more variable than the wild type (p-value = 0.002), however it was significantly less variable than either of the parental strains (p-value compared to *ceh-16(bp323)* = 3.36 x 10^{-10}, and compared to *egl-18(ga97)* = 2.37 x 10^{-16}). The 11 seam cells were consistently evenly spaced with two cells in the tail, implying that they were the original 10 that the animal hatched with, with an additional cell in the tail. However given that the definition of robustness includes resistance to both genetic and environmental perturbations (Félix and Wagner, 2008), in order to truly describe the strain as robust we must further subject it to perturbations (see Chapter 5).

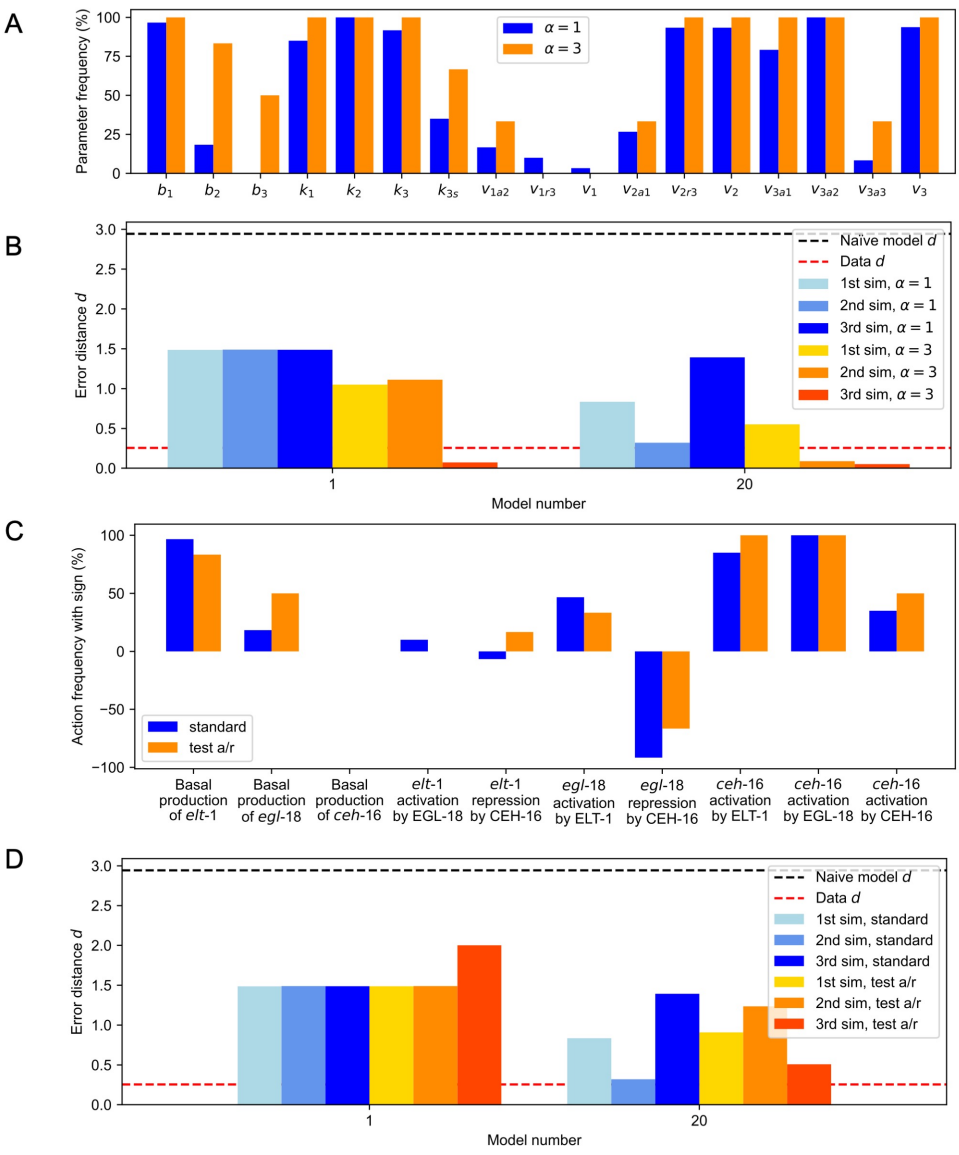


Figure 3.4: **Testing additional conditions for SLING: A)** Bar graph showing parameter presence in the final models comparing α = 1 (60 SLING runs, 3 for each one of the 20 models) with α = 3 (6 SLING runs, 3 for models 1 and 20). **B)** Distance measure of the median parameter set in the SLING runs for models 1 and 20 comparing α = 1 and α = 3. **C)** Bar graph showing the percentage and sign of the actions of the standard model compared to the activation/repression (a/r) model. **D)** Distance measure of the median parameter set of the standard model compared to the activation/repression (a/r) model.

To validate our model, we quantified the expression levels of the three core seam cell transcription factors (TFs) in *ceh-16(bp323);egl-18(ga97)* at the 17-hour time point and compared this experimental data with our predictions. While we observed no significant change in *elt-1* levels, there was a significant increase in *egl-18* levels and a decrease in *ceh-16* levels compared to the wild type. By utilising the ODEs, we assessed the behavior of our models under double mutant conditions and calculated the distance from the experimental data.

While all the combinations of AND/OR links performed well at single mutant level, we noticed that with the double mutant 24 out of 60 runs showed a decrease in performance, with the models closest to the single mutant data showing the highest error (Figure 6C). We categorised poor performance mean expressions with a distance greater than 2.5 in any of the simulations. Examining the prediction accuracy of specific gene expressions, we observed that *elt-1* expression was accurately predicted in all models and simulations (Figure 3.5D). However, predictions for *egl-18* were more variable, with certain models overestimating its levels while others underestimated them (Figure 3.5D). Finally the SLING runs that performed poorly tended to underestimate *ceh-16* expression levels (Figure 3.5D). Models that performed poorly on any simulation were discarded. We then examined which parameters were prevalent in the retained models compared to those that were discarded. Interestingly, *ceh-16* activation by EGL-18 (v_{3a2}) was retained in both the models that were kept and those that were discarded due to large errors. However, the *ceh-16* self-activation parameter (k_{3s}) was only retained in the discarded models, suggesting that the *ceh-16* self-regulatory loop may contribute to over-fitting.

Testing our models on the *ceh-16(bp323);egl-18(ga97)* double mutant allowed us to narrow down which AND/OR combinations would function best within this network. We used this to create the minimal network depicted in (Figure 3.5F). Remarkably, the activation of *ceh-16* by EGL-18 and the repression of *egl-18* by CEH-16 were consistently retained in all Markov Chain Monte Carlo runs that yielded accurate gene expression estimates in the double mutant, irrespective of whether other actions on the same gene were linked by an AND or an OR. This highlights the importance of this negative feedback loop for the correct

functioning of the core network. Similarly, *ceh-16* activation by ELT-1 was preserved in the majority of runs, regardless of whether it was linked to the activation of *ceh-16* by EGL-18 with an AND or OR link. However, the primary uncertainty within this network lies in the activation of *egl-18* by ELT-1, which appears to be present only when linked by an AND to the CEH-16 repression of *egl-18*. This suggests that the activation of *egl-18* by ELT-1 is conditional on *egl-18* being repressed. By removing links that were discarded in over 90% of the cases in low-error models, we derived the final network shown in Figure 3.5F. Consequently, the final equation for *elt-1* expression is:

$$\frac{dg_1}{dt} = b_1 - d_1 g_1. \tag{3.4}$$

There are two forms of equations for *egl-18* expression:

$$\frac{dg_2}{dt} = v_{2r3}\left(\frac{k_3}{k_3 + (g_3 f_3)}\right) - d_2 g_2 \tag{3.5}$$

or

$$\frac{dg_2}{dt} = v_2\left(\frac{(g_1 f_1)}{k_1 + (g_1 f_1)}\right) \cdot \left(\frac{k_3}{k_3 + (g_3 f_3)}\right) - d_2 g_2. \tag{3.6}$$

and also two forms of equations for *ceh-16* expression:

$$\frac{dg_3}{dt} = v_{3a1}\left(\frac{(g_1 f_1)}{k_1 + (g_1 f_1)}\right) + v_{3a1}\left(\frac{(g_2 f_2)}{k_2 + (g_2 f_2)}\right) - d_3 g_3, \tag{3.7}$$

or

$$\frac{dg_3}{dt} = v_3 \left(\frac{(g_1 f_1)}{k_1 + (g_1 f_1)} \right) \cdot \left(\frac{(g_2 f_2)}{k_2 + (g_2 f_2)} \right) - d_3 g_3. \tag{3.8}$$

These equations represent the network depicted in Figure 3.5F, incorporating an incoherent feedforward network and negative feedback as integral components of the regulation of the three core genes.

3.2.6 Evaluating our model in a protein context

Our final test of the model was to test the response of protein reporters. We utilised ELT-1::EGFP, CEH-16::AID::nG, and EGL-18::nG (Table 2.1). We tested the same early L1 time point as we did for the smFISH study. We used RNAi to knock down the three essential genes and to make sure that the RNAi was functional we quantified the seam cell losses seen in (Figure 3.6A). We quantified the changes and then calculated the differences from the untreated animals using Mann Whitney U. As expected from our model we observed negligible changes in ELT-1 when treated with *egl-18* RNAi (p-value = 0.4). We observed a slight increase in ELT-1 in *ceh-16* RNAi treatments (p-value = 0.007) and a highly significant decrease of ELT-1 in the *elt-1* RNAi treatment (p-value = 7.7 x 10^{-12}). An unexpected result was the decrease in EGL-18 in the *ceh-16* treated animals. However, the *ceh-16* RNAi had a very weak response that wasn't comparable to that of the mutant. We did however observe a significant decrease in EGL-18 in the *elt-1* RNAi treated animals (p-value = 0.001) which is in line with our model. Finally in terms of CEH-16 we show significant losses of the protein in both *ceh-16* and *egl-18* treatment (p-values of 3.5 x 10^{-13} and 3.2 x 10^{-31} respectively). However, the fact that *egl-18* RNAi is more efficient than the *ceh-16* RNAi at knocking down CEH-16 could be indicative of very weak *ceh-16* RNAi or a very strong activation from EGL-18 to *ceh-16*. Another protein result that contrasts our findings is the apparent lack of activation of *ceh-16* by ELT-1, as the decrease in CEH-16 protein in *elt-1* RNAi treated animals is not significant (p-value = 0.15)

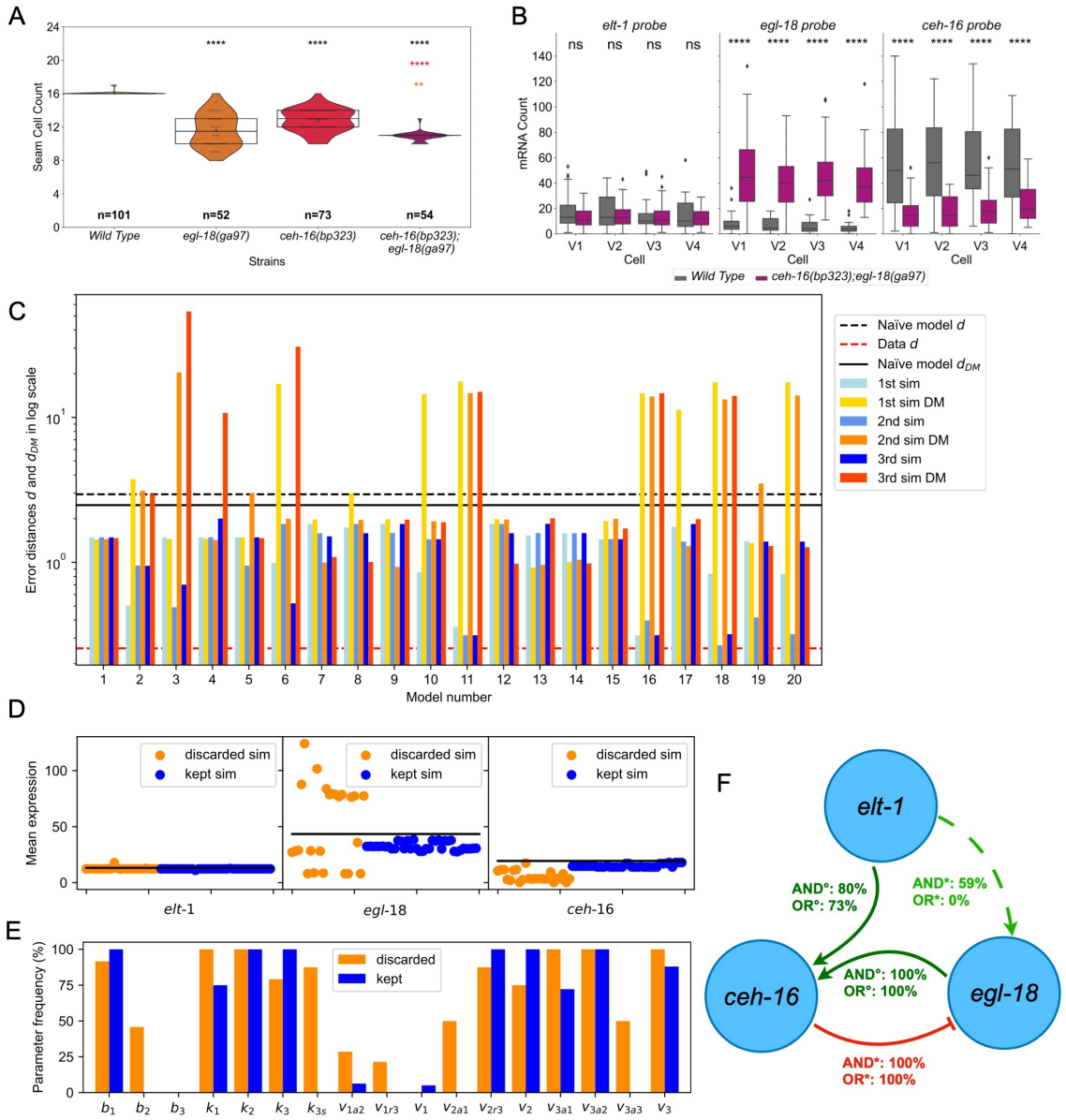


Figure 3.5: **Validating model predictions using the *ceh-16(bp323);egl-18(ga97)* double mutant: A)** Seam cell counts in the double *egl-18(ga97); ceh-16(bp323)* mutant in comparison to the wild type and the parental strains. The significance of the changes in cell number was calculated using the Z-test and denoted as shown in methods. The comparison between all the mutants and wild type is denoted in black asterisk, and the comparison between the double mutant and the parental strains is denoted with asterisks in the colour of the corresponding parental strain. We additionally calculated the difference in variability using the Levene's test between the double mutant and wild type, *ceh-16(bp323)* and *egl-18(ga97)* and the resulting p-values were 1.85×10^{-4}, 3.35×10^{-10}, and 2.37×10^{-16} respectively. **B)** Quantification of *elt-1*, *egl-18* and *ceh-16* expression in V1-V4 seam cells via smFISH in the wild type (n = 77, n = 106, n = 153) and *egl-18(ga97); ceh-16(bp323)* double mutant (n = 176, n = 208, n = 120). The significance of the changes in expression was calculated with respect to the wild type using the Mann-Whitney U test and denoted as described in methods. **C)** Distance measure d (on a logarithmic scale) of the median parameter sets in predicting the wild type and single-mutant behaviour vs the double mutant behaviour. The bars are superimposed with the data presented in Figure 4C, with the lowest distance measure (depending on the SLING run) shown in front. **D)** Predicted gene expression in the double mutant from SLING runs (using the median parameter set) is compared to data (black lines). SLING runs are grouped into 2 categories based on their double-mutant error: they are discarded when 2.5 $<$d and kept when $d \leq 2.5$. **E)** Parameter presence (in % of models) for the two categories of SLING runs is shown. **F)** Final model based on all kept SLING runs that are able to predict the double mutant data. The possible interacting actions on *egl-18* are indicated with the symbol *, while the activations on *ceh-16* are indicated with the symbol °. The percentage shown on each link is the number of times each connection is kept given an initial model that supposes an AND/OR relationship with the effect on the same gene by the other factor.

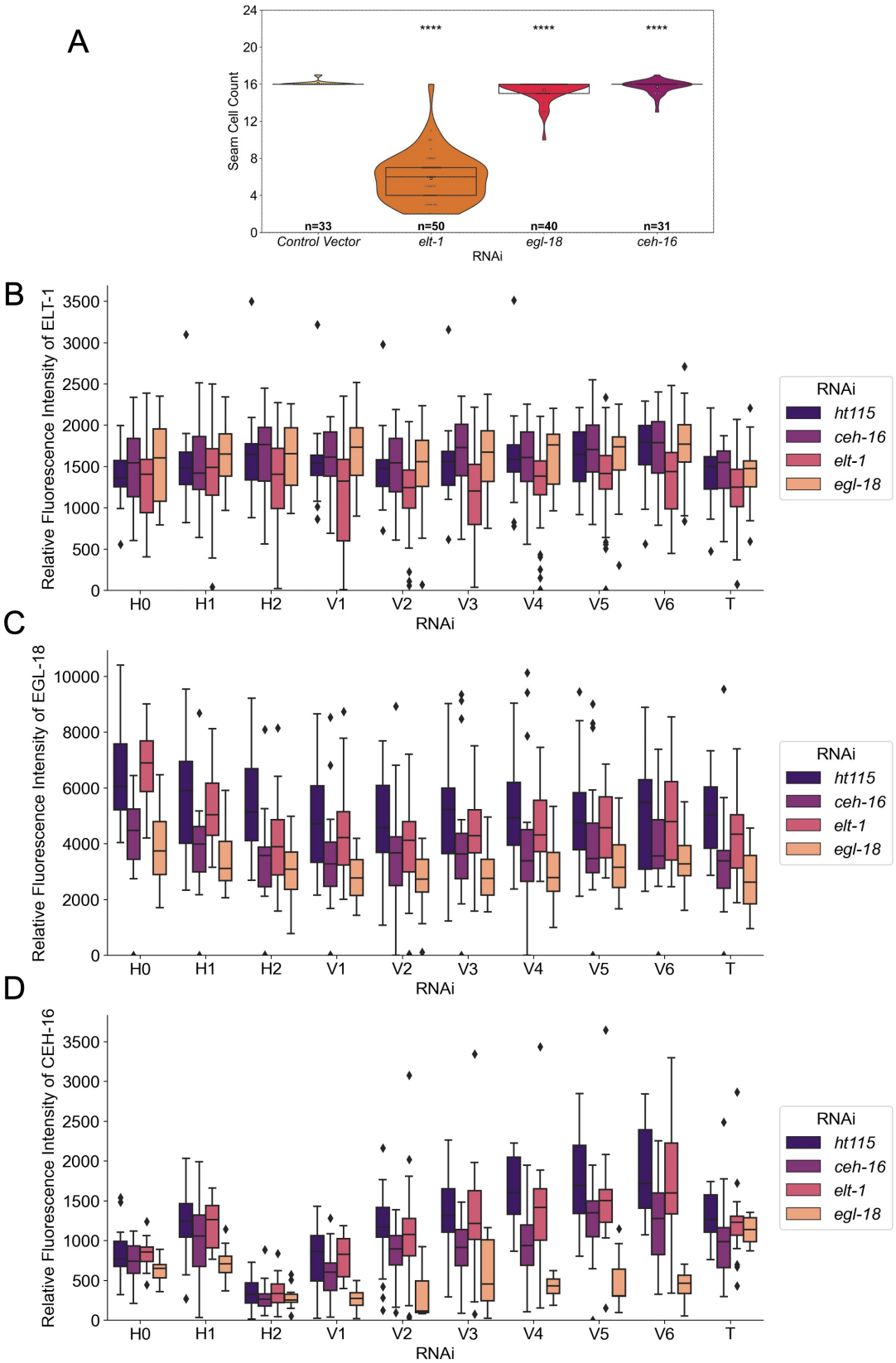


Figure 3.6: **Core seam cell genes' protein fusions' relative fluorescence in RNAi treated animals: A)** A graph showing seam cell counts in animals hatched from eggs bleached onto control vector, *elt-1*, *egl-18*, and *ceh-16* RNAi bacteria. **B-D** Fluorescence intensity of ELT-1, EGL-18 and CEH-16 respecively in each seam cell relative to the background.

3.3 Discussion

Understanding the genetic regulation of the core seam cell genes is crucial for utilizing seam cells as a model for stem cell biology. Our approach aimed to integrate experimental and computational biology to re-evaluate the GRN underlying seam cell patterning. We relied on smFISH data to quantify core gene expression in mutant and wild type backgrounds. Despite inherent noise in smFISH data, we successfully identified clear changes in expression patterns under mutant conditions, which served as the basis for extrapolating the network architecture using mathematical methods. While the MRA conversion of the global perturbation to a Jacobian matrix was designed around smaller perturbations, and thus can make erroneous estimates of the connection strengths at larger perturbations, the algorithm has previously been used together with full knockout data (Borg et al., 2023). As such we believed it necessary to complement our approach with other methods. We adapted the MRA algorithm (Kholodenko et al., 2002) to not rely on the quantity of gene expressed to infer the amount of functional protein. This allows for broader flexibility to work with hypomorphs and thus study essential genes in detail. We relied on hypomorphs as even seam cell specific knock-outs of *elt-1* are lethal (unpublised lab data) and we wanted to keep a consistent mutation type across all three genes. A future approach may include also quantifying the heterozygotes of these hypomorphs to have a more granular quantitative understanding of the interactions between the core seam cell genes. By quantifying the strengths of potential interactions between the genes using the adapted MRA, we were able to develop a series of ordinary differential equations (ODEs), which we modeled using the newly developed Bayesian model discovery approach, SLInG (Jørgensen et al., 2023). This facilitated the generation of a minimal model of the GRN while incorporating different regulatory logic. We utilised our newly generated model to formulate predictions on the behaviour of a double *ceh-16(bp323);egl-18(ga97)* mutant. This mutant served as a means to test and refine our model and we used it to show that we were able describe the most likely behaviour for our system given changes in the activity of the three core seam cell genes. Remarkably, the *ceh-16(bp323);egl-18(ga97)* mutant exhibited a highly robust loss of seam cell number, with only 11 seam cells present in

the majority of L4 larva. Additionally the seam cell positional pattern closely resembled that of newly hatched L1 larvae (Sulston and Horvitz, 1977). This indicates that the loss of the L2 duplication rather than a loss of seam cell fate from any of the seam cells. This differs from *egl-18(ga97)* and *ceh-16(bp323)* seam cell loss patterns that indicate that the *C. elegans* either hatches with less seam cells than normal or loses them over development. While most seam cell gene mutants display considerable variability, the robustness of the double mutant resembles that of the wild type. We delve deeper into investigating this phenotype in Chapter 5. Notably, this phenotype aligns with the newly proposed seam cell GRN, as the negative feedback loop would remain unbalanced in the absence of either *egl-18* or *ceh-16*, but the absence of both would prevent the deregulation of the other gene and thus downstream targets.

An element of our study that has left room for future directions is the presence of noise. Gene expression inherently exhibits high levels of noise, which can stem from both intrinsic factors, such as variability in biochemical reactions involved in gene regulation, and extrinsic factors, which result from diverse responses to environmental stimuli (Shahrezaei and Swain, 2008). Given the observed noise levels in our study and the significance of understanding noise in inferring GRNs (Munsky et al., 2009; Stumpf, 2021) one could look into further methods like two colour TF reporter quantification(Elowitz et al., 2002), dual smFISH (Kramer et al., 2021) or time lapse imaging (Dunlop et al., 2008) to decouple extrinsic and intrinsic noise. For our study, we have primarily focused on extrinsic noise and have adopted a deterministic approach using ordinary differential equations (ODEs), largely informed by mean expression data. While we have attempted a dual smFISH experiment, the data was too noisy to draw concrete conclusions as the dsRED dye was not bright enough when visualised. The experiment could be re-attempted with a different fluorescent dye, further away in the difraction spectrum from far-red. Additionally, we have assumed direct interactions, as supported by the binding of ELT-1 and CEH-16 in the upstream region of the *elt-1*, *egl-18*, and *ceh-16* gene loci in Luo et al. (2020); ENCODE Project Consortium (2012); Kudron et al. (2018). However, direct binding data for EGL-18 is currently unavailable.

The network described in literature portrayed ELT-1 as the activator of all seam cell genes,

including *ceh-16* and *egl-18*(Huang et al., 2009; Cassata et al., 2005; Thompson et al., 2016). In this study we have confirmed positive interactions from ELT-1 to *egl-18* and *ceh-16*, reinforcing the portrayal of *elt-1* as the master regulator of the seam cells (Spieth et al., 1991; Page et al., 1997b). We also confirmed the existence of an interaction from CEH-16 to *egl-18* as shown in (Cassata et al., 2005). Nevertheless while this embryonic *C. elegans* study relied on strong *ceh-16* RNAi and EGL-18:GFP fusions to show an activation from CEH-16 to *egl-18*, we observed a repressive interaction in the larval stages. This is in line with the function of its homolog engrailed which, despite having been observed to also carry out activatory functions, is largely described as a repressor (Markel et al., 2002; Alexandre and Vincent, 2003; Genestine et al., 2015; Huang et al., 2009). Several factors may contribute to this discrepancy. Firstly, the genetic perturbation used in the study by (Cassata et al., 2005) is RNAi-based rather than a hypomorph. This RNAi was demonstrated to be potent, leading to premature seam cell fusion and embryonic lethality (Cassata et al., 2005), thus a reduction in EGL-18 expression could result from a weaker seam cell phenotype rather than solely loss of expression. However, it is also plausible that an activation from CEH-16 to *egl-18* reflects the TF's differing roles across developmental stages. Additionally there may be a need to investigate post-translational regulation of EGL-18 protein expression, given that our own protein studies point to a decrease in EGL-18 in absence of *ceh-16*, despite a strong repression being seen at the same time point at mRNA level.

Our results from the protein studies show that the network has further complexities that need elucidating. Specifically some links, like the ELT-1 activation of *ceh-16* and the *egl-18* repression by CEH-16, require further investigation. We delve into this further in Chapter 4, where we conduct the analysis of the core gene network at different time points and with specific insight on the individual seam cells. Finally, we have uncovered a previously unreported interaction among the core genes, wherein we observe EGL-18 activating *ceh-16*. Our model suggests that this interaction is essential for the network's function.

Our newly developed seam cell GRN contains both positive and negative feedback loops as well as an incoherent feed-forward loop where ELT-1 activates CEH-16 which then represses EGL-18, which is also activated by ELT-1. Such a complex system could

contain many dynamic mechanisms contributing to robust cell fate maintenance. Numerous studies propose that robust cellular decision-making and developmental processes hinge upon nested feedback loops with nonlinear behaviors, such as bistability and threshold-based decisions (Balázsi et al., 2011; Moris et al., 2016; Sáez et al., 2022; Petratou et al., 2018). Our investigation adds a quantitative dimension to the exploration of the core epidermal stem cell GRN in *C. elegans*. We further delve into the GRN further by examining it in various spatial and temporal contexts (Chapter 4) and investigating a specific robust phenotype resulting from the mutation of two of the core seam cell genes (Chapter 5).

Chapter 4

Variations in the seam cell GRN across time and space

4.1 Introduction

Our initial model was highly focused on a small group of cells at one time point which we assumed to be steady state. In this chapter we evaluate the differences between the behaviour of the seam cells over time and consider the extent to which our model explains the changes in behaviour. Additionally, we explore differences in the behavior of various seam cells, aiming to understand whether variations in gene expression patterns lead to differences in phenotype and whether seam cell sensitivity to mutations can be predicted from gene expression at any of the time points. Finally, we try to disrupt specific positional patterns, that we have identified in response to a perturbation of *elt-1*, through the disruption of *Hox* and *Wnt* genes in order to identify whether the positional patterns given by the seam cell master regulator can be explained through either signalling network.

In order to achieve all this we combine a variety of experimental and mathematical techniques. We rely on the same hypomorph mutants that we showed in Chapter 3 to study the behaviour of the core genes over time. However, to obtain a more granular understanding of the resulting phenotype at a cell-to-cell level, we supplemented this approach with RNA interference (RNAi). RNAi by feeding is a pivotal technique that underscores the versatility

of *C. elegans* as a genetic model organism. Utilizing various dilutions of the RNAi bacteria, we probed the sensitivities of different seam cells to more moderate disruptions than those caused by hypomorph mutations.

We also integrated a series of mathematical approaches. We repeated the MRA approach that we showed in Chapter 3 to visualise gene interactions at different time points. We concluded that a minimal model may need to integrate additional factors to comprehensively elucidate the network's temporal evolution. We also show that while the 17 hour time point is representative of the behaviour of most of the genes, the gene expression of the core genes in early larval development is too dynamic to truly identify a steady state with snapshot data. Additionally, we employed clustering and statistical analyses to identify groups of cells exhibiting similar expression patterns and compared them to the phenotypic sensitivities of the seam cells to the various biological perturbations.

We identified specific seam cell sensitivities driven by *elt-1* and attempted to perturb them by disrupting both the *Wnt* pathway and the *Hox* genes. The *Wnt* pathway is essential for establishing anterior posterior patterns in *C. elegans*, and given we found clear differences between anterior and posterior cells, we thought it was an obvious candidate for a link to the core gene network.

The *Hox* gene family was first discovered in *Drosophila* (Bateson, 1894), and first studies in these genes revealed their dis-regulation could causes duplication of thoracic segments (Lewis, 1978). They have been demonstrated to regulate anterior/posterior cell patterning in various organisms (Di Bonito et al., 2013; Lanctôt et al., 2007; Krumlauf, 1994; Martinez and Amemiya, 2002). This includes *C. elegans*, where the *Hox* genes control cell proliferation, survival, and migration. Six *Hox* genes were identified in *C. elegans*:*lin-39, mab-5, egl-5, nob-1, php-3* and *ceh-13* (Van Auken et al., 2000; Maloof and Kenyon, 1998; Harris et al., 1996; Tihanyi et al., 2010). Among these, *ceh-13* and *lin-39* are predominantly implicated in anterior function, while *mab-5* functions in the mid-body, and *egl-5, nob-1*, and *php-3* act in the posterior region (Smith and Kratsios, 2024). There is strong evidence that seam cell patterning is regulated by the *Hox* genes (Aboobaker and Blaxter, 2003). For example, *egl-5* and *mab-5*, have been shown to be involved in the patterning of V5 and V6 (Kenyon, 1986;

Chisholm, 1991). As such when investigating positional patterns one must consider the role of *Hox* genes. Given the absence of key studies linking core seam cell genes to the *Hox* gene network, we decided to evaluate possible interactions of *Hox* genes with *elt-1* specific seam cell patterns.

In this chapter, we aim to understand the behaviour of the core seam cell genes at different time points and different states. We aim to understand pre-division states like the 10 and 26-hour post-symmetric division time points. We also aim to confirm the behaviour of the post-division states by looking at the 26-hour asymmetric division while aiming to evaluate whether it is a valid biological assumption to call the late L1 (17 hour) time point a steady state. We also aim to understand the positional changes between the 16 seam cells in the core seam cell genes and whether the positional differences in expression patterns and whether the expression patterns can be used to infer the sensitivities of the individual seam cells. We further expand our evaluation of the core GRN underlying stem cell patterning to incorporate differences observed between seam cells. We consider this both from an expression and a phenotypic stand point by combining our time based smFISH data with seam cell counts carried out at L4. Upon discovering a positional pattern in response to *elt-1* loss we evaluate whether seam cell specific sensitivities can be disrupted through the disruption of *Hox* or *Wnt* factors and we test whether *Hox* or *Wnt* play a larger role in the positional patterning of seam cells driven by *elt-1* expression.

4.2 Results

4.2.1 Constructing the seam cell gene network in the early L1 stage

While our results from the previous chapter illustrate the behaviour of the three core seam cell genes at an assumed steady state, it is crucial to understand these genes at other time points and conditions. The 17-hour time point was of particular interest as it lacks actively dividing cells, with all seam cells having undergone at least one asymmetric division and may be assumed to be a steady state. This was the earliest time point where none of the cells

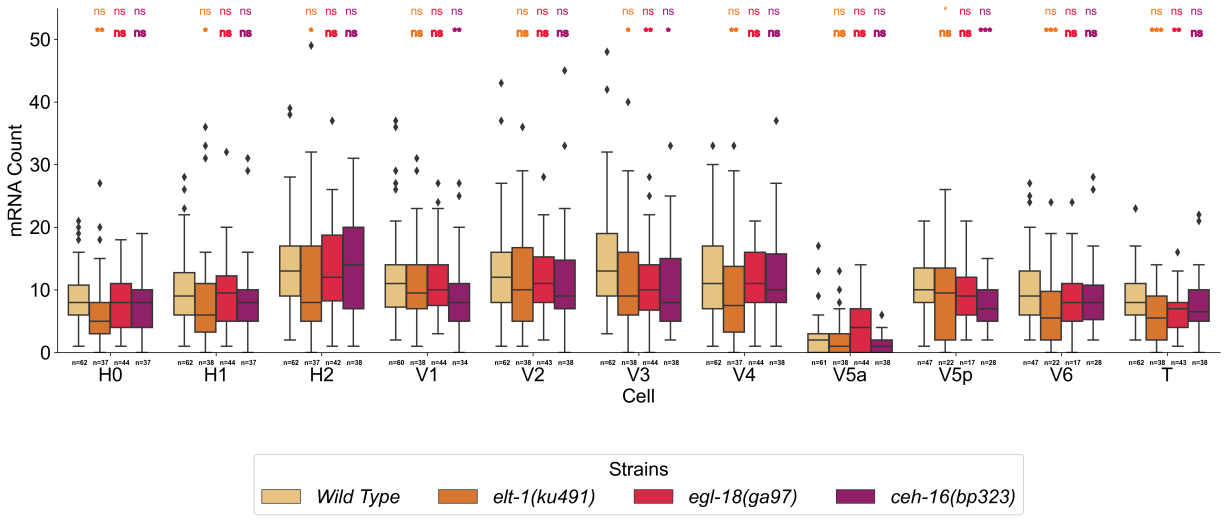


Figure 4.1: ***elt-1* expression 10 hours after bleaching in wild type, *ceh-16*, *egl-18* and *elt-1* mutants:** Quantification of smFISH signal in all seam cells in animals in early L1. Strains are colour-coded as indicated in the key underneath the panel. The significance of the changes between mutant and wild type backgrounds is presented using the system described in the methods, and the colour of the * matches the strain that is being compared to the wild type. The significance of the changes between mutant and wild type were calculated using Levene's test to compare variability, and are shown above, whereas below and in bold, we have the comparison calculated using the Mann-Whitney U test to determine the difference in the distributions.

were actively dividing.

At 10 hours after bleaching, which marks the first time point when the majority of animals have hatched from their eggs, V5 is typically already divided into V5a and V5p in most cases. The remaining seam cells are primed to divide shortly after. Therefore, we chose to investigate this time point to understand the genetic landscape of the core seam cell genes before their first asymmetric division. Based on our data at 17 hours, we anticipated that mutations in the core seam cell genes would have minimal effect on *elt-1* expression.

We observe minimal change between the wild type and the mutants. A slight decrease in expression is noted in the *elt-1(ku491)* mutant in the head cells and V3 and V4 (Figure 4.1). Additionally, *elt-1* expression in V3 appears to be the most susceptible to mutations, showing changes in across all mutants (Figure 4.1). However, given that this cell's expression is slightly higher than in the other cells, in the mutants the decrease still falls within the normal range for the other cells. Furthermore, the *ceh-16(bp323)* mutant exhibits a slight decrease in expression in V1, V3, and V5p. Conversely, in *egl-18(ga97)*, we observe no significant change

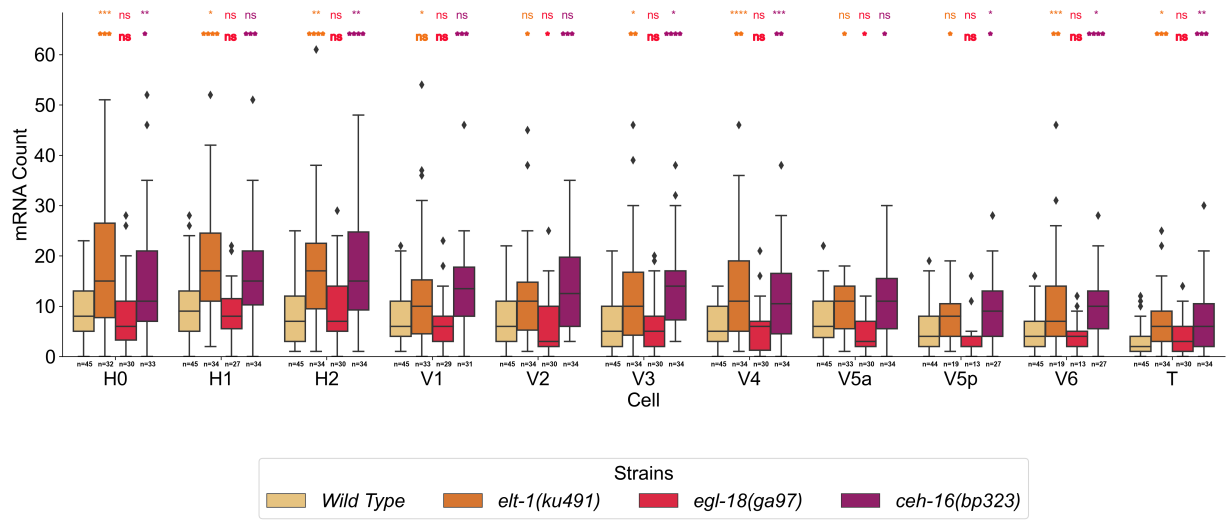


Figure 4.2: ***egl-18* expression 10 hours after bleaching in wild type, *ceh-16*, *egl-18* and *elt-1* mutants:** Quantification of smFISH signal in all seam cells in early L1. Strains are colour-coded as indicated in the key underneath the panel. The significance of the changes between mutant and wild type backgrounds is presented using the system described in the methods and the colour of the * matches the strain that is being compared to the wild type. The significance of the changes between mutant and wild type were calculated using Levene's test to compare variability, and are shown above, whereas below and in bold we have the comparison of calculated using the Mann Whitney U test to determine the difference in the distributions.

in expression outside of V3 (Figure 4.1). Variability remains largely unchanged, except for V5p in the *elt-1(ku491)* mutant (Figure 4.1).

The most striking pattern that we observe in the *elt-1* expression is the difference between the cells maintaining the seam cell fate and V5a (Figure 4.1). The mean V5a expression of *elt-1* is less than one-third of the mean of the other 9 cells we observed (Figure 4.1). This underscores the critical role of *elt-1* in seam cell maintenance.

In the wild type we also observe higher *egl-18* expression in the head cells than in the mid body. Based on the 17-hour data, we expected a gain in expression in absence of *ceh-16* (Figure 4.2). Indeed, we observe this in every cell, with *egl-18* levels increasing 1.5 times in the *ceh-16(bp323)* mutant relative to the wild type (Figure 4.2). We also note a similar level of increase of *egl-18* expression in the *elt-1(ku491)* mutant (Figure 4.2). Additionally, both mutants exhibit a slight increase in variability, albeit observed only in certain cells (Figure 4.2). In contrast, the *egl-18(ga97)* mutant shows no significant change in *egl-18* expression, except for a slight decrease in V5a, confirming our prediction of the absence of a self-regulatory loop

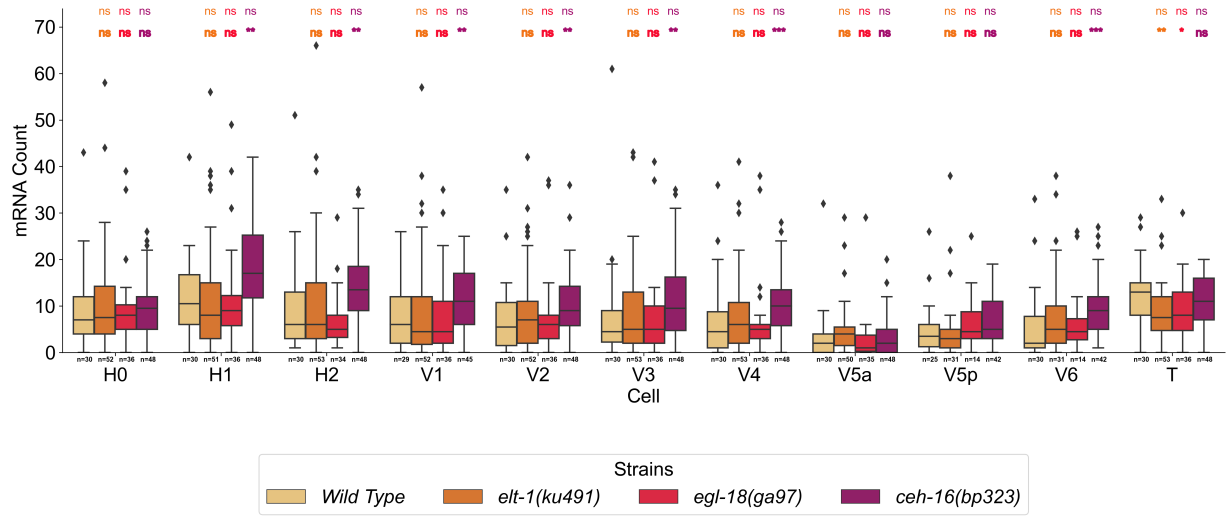


Figure 4.3: ***ceh-16* expression 10 hours after bleaching in wild type, *ceh-16*, *egl-18* and *elt-1* mutants:** Quantification of smFISH signal in all seam cells in early L1. Strains are colour-coded as indicated in the key underneath the panel.The significance of the changes between mutant and wild type backgrounds is presented using the system described in the methods and the colour of the * matches the strain that is being compared to the wild type. The significance of the changes between mutant and wild type were calculated using Levene's test to compare variability, and are shown above, whereas below and in bold we have the comparison of calculated using the Mann Whitney U test to determine the difference in the distributions.

(Figure 4.2).

ceh-16 levels are similar across all cells, however V5a expresses slightly less and H1a and T express slightly more *ceh-16* (Figure 4.3). Like with *elt-1*, *ceh-16* appears to have lower expression in the differentiating V5a (Figure 4.3). Furthermore *ceh-16* levels remain unchanged in both the *elt-1(ku491)* and the *egl-18(ga97)* mutants, unlike what we observed at 17 hours (Figure 4.3). The only cell in which we see a slight decrease in expression in these mutants is the T cell (Figure 4.3). In the *ceh-16(bp323)* mutant we note elevated *ceh-16* expression in all cells except V5a, V5p and T (Figure 4.3). We note particularly high *ceh-16* expression in the H1 cell (Figure 4.3).

We utilized the adapted MRA introduced in Chapter 3 to assess the interaction strengths at 10 hours. Upon pooling data from all cells, we observed three moderate interactions and three weak ones. The repression of *egl-18* by ELT-1 and CEH-16 both were calculated to have a strength of $\approx$0.5 (Figure 4.4). The interaction from CEH-16 to *egl-18* was previously observed at 17 hours (Figure 3.5F). However, the interaction from ELT-1 to *egl-18* was seen

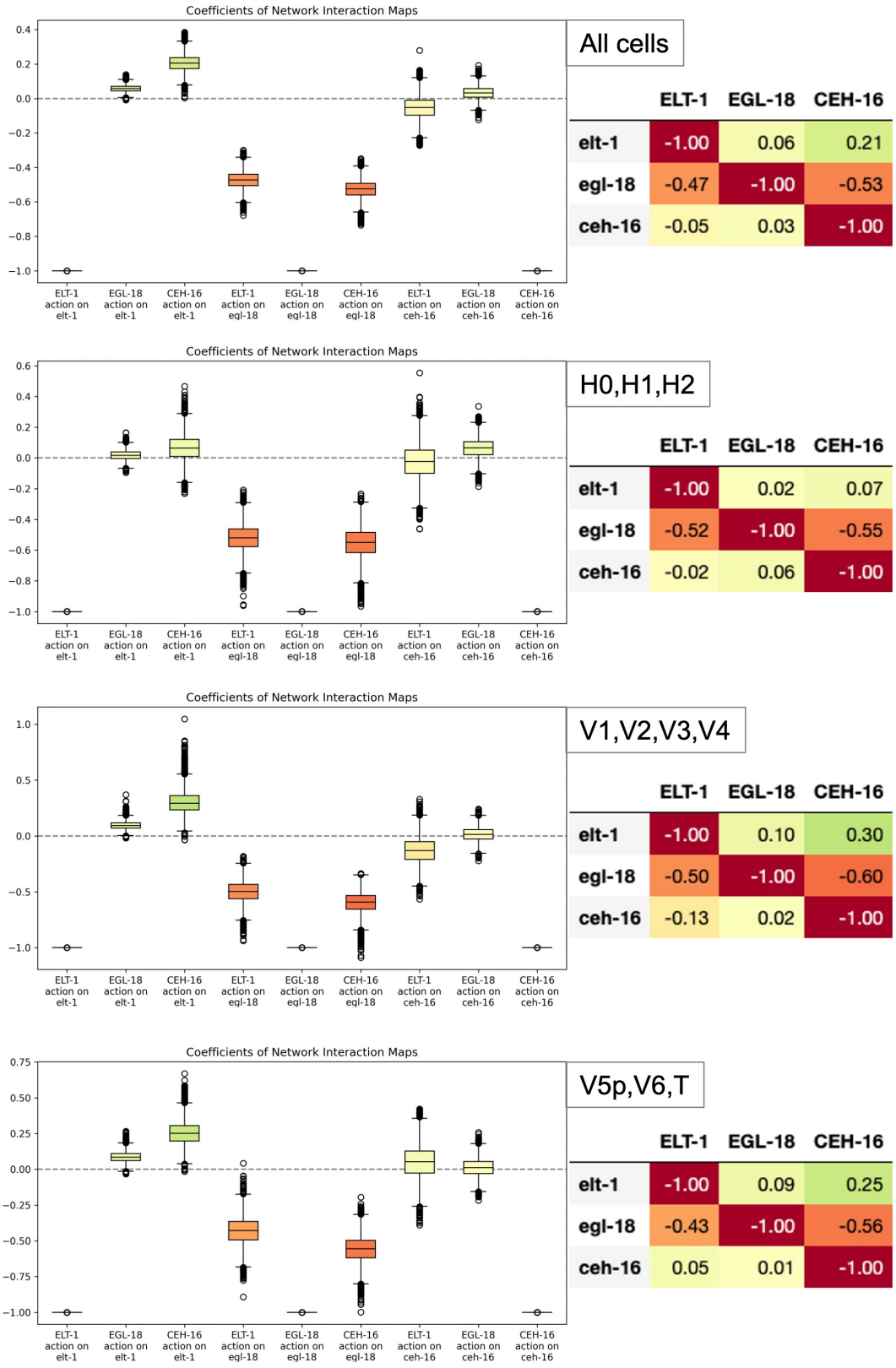


Figure 4.4: **Strengths of interactions between core genes in early L1 calculated using MRA:** Illustration of the MRA matrix when combining first, all cells, next just the head cells (H0, H1,H2, then the midbody cells (V1,V2,V3,V4) and finally the tail cells (V5p, V6, T). The colour coding correlates saturation to strength of interaction, and the colour to the polarity (green positive, red negative). The error bars show the range of response expected due to data variability and boostrapping.

as a positive interaction, conditional on the repression by CEH-16 (Figure 3.5F). Additionally, we observed a weak but positive interaction from CEH-16 to *elt-1* (Figure 4.4). Although weak, the error bars from bootstrapping indicate minimal overlap with 0. The remaining three interactions exhibited strengths below 0.1, and an error bar overlap with 0 (Figure 4.4). As such we considered them negligible.

In addition to analyzing the strengths of interactions in all pooled cells, we opted to categorize cells into V1-V4, H0-H2, and V5p, V6, and T to begin observing positional patterns (Figure 4.4).

In the head cells, we find that the activation of *elt-1* by CEH-16 becomes weaker, however the rest of the network remains similar to the pooled cells (Figure 4.4). Conversely, in V1-V4, the activation of *elt-1* by CEH-16 strengthened slightly, along with the CEH-16 repression of *egl-18*. All the other links remain the same as even though the repression of *ceh-16* by ELT-1 increases in strength, the errorbars are too large to consider this a viable connection (Figure 4.4). We note no large differences between V1-V4 and V5p-T (Figure 4.4).

Several disparities are evident in the output obtained at 17 hours. These include the positive interaction from CEH-16 to *elt-1*, the negative interaction from ELT-1 to *egl-18*, as well as a lack of a strong interaction from EGL-18 to *ceh-16*. We do note that the repression from CEH-16 to *egl-18* remains present and of similar strength to what we observe at 17 hours.

4.2.2 Constructing the seam cell gene network at the L2 symmetric seam cell division

The L2 symmetric and asymmetric cell divisions occur closely in succession, making it challenging to identify a time point where only one of these divisions is present. Therefore, we opted to concentrate on a time point where both types of divisions are easily discernible. At 26 hours, both symmetrically divided cells and asymmetrically divided quadruplets are observable. We begin by examining the symmetric daughters, where we refer to the anterior symmetric daughter by adding a to the cell name and the posterior cell by adding p to the cell name.

From our 10-hour (Figure 4.14) and 17 hour (Figure 7.1) data, the assumption that all seam cells are the same doesn't appear to be accurate. Therefore, in investigating the symmetric division, we aimed to examine potential differences in core gene expression between anterior and posterior cells, as well as disparities between the wild type and the mutants.

elt-1 expression appeared symmetric in all wild-type cells undergoing symmetric division (Figure 4.5). However, in asymmetrically dividing cells such as H2, V5, and T, *elt-1* expression exhibited asymmetry (Figure 4.5). This asymmetry persisted in the *elt-1(ku491)* mutant, except in the T cell (Figure 4.5). Conversely, *egl-18(ga97)* mutants displayed symmetric *elt-1* expression in all cells, including those where the anterior cell would normally activate the differentiation program (Figure 4.5). Conversely, *ceh-16(bp323)* mutants exhibited *asymmetric elt-1* expression in all cells, where the anterior cell expressed less *elt-1* than the posterior. This pattern was inverted in H1 (Figure 4.5).

In terms of expression level changes of *elt-1* in each cell, in both the H1 cells, we observe a decrease in expression accross all mutants (Figure 4.5). In H2, decreased expression was noted in *elt-1(ku491)* in H2p and in *ceh-16(bp323)* in both daughter cells, with the latter also showing reduced variability (Figure 4.5). In V1-V4, the most striking phenotype is the loss of *elt-1* expression in the *ceh-16(bp323)* mutant (Figure 4.5). This phenotype is particularly prominent in the anterior cell. V3 is the only V cell that is sensitive to the *elt-1(ku491)* mutation, with significant response in both cells (Figure 4.5). V3a appears to lose expression also in the absence of *egl-18*. However, the decrease in expression in both these mutants is very slight (Figure 4.5).

Overall the main effect we observe is the loss of *elt-1* expression in anterior symmetric daughters in the *ceh-16(bp323)* (Figure 4.5). We also observed lower *elt-1* expression in cells that were not fated to carry out an asymmetric seam cell division, or that were differentiating.

egl-18 expression remains consistent between anterior and posterior cells in the wild type, but also in the mutants (Figure 4.6). However, *egl-18* levels in the head cells show a positive response to the mutants. Particularly in the *elt-1(ku491)* and *ceh-16(bp323)* mutants, we observe *egl-18* expression levels more than double that of the wild type, and a smaller increase in expression in *egl-18(ga97)* (Figure 4.6). V1-V3 and V6 all respond very strongly

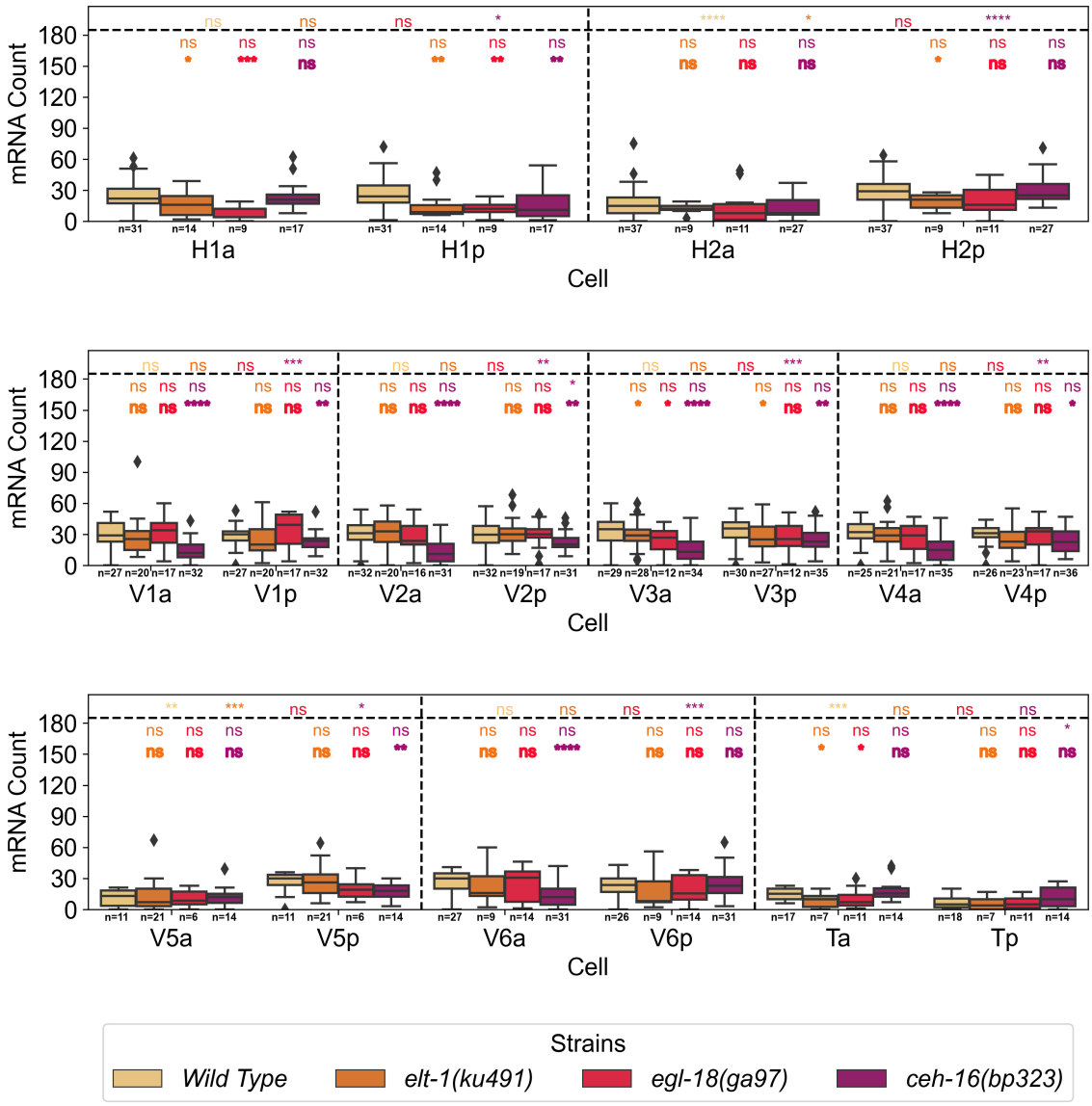


Figure 4.5: ***elt-1* expression 26 hours after bleaching, in the symmetric daughters in wild type, *ceh-16*, *egl-18* and *elt-1* mutants:** Quantification of smFISH signal in all seam cells. Strains are colour-coded as indicated in the key underneath the panel. The significance of the changes between mutant and wild type backgrounds is presented using the system described in the methods and the colour of the * matches the strain that is being compared to the wild type. The significance of the changes between mutant and wild type were calculated using Levene's test to compare variability and are shown above, whereas below and in bold, we have the comparison calculated using the Mann-Whitney U test to determine the difference in the distributions.

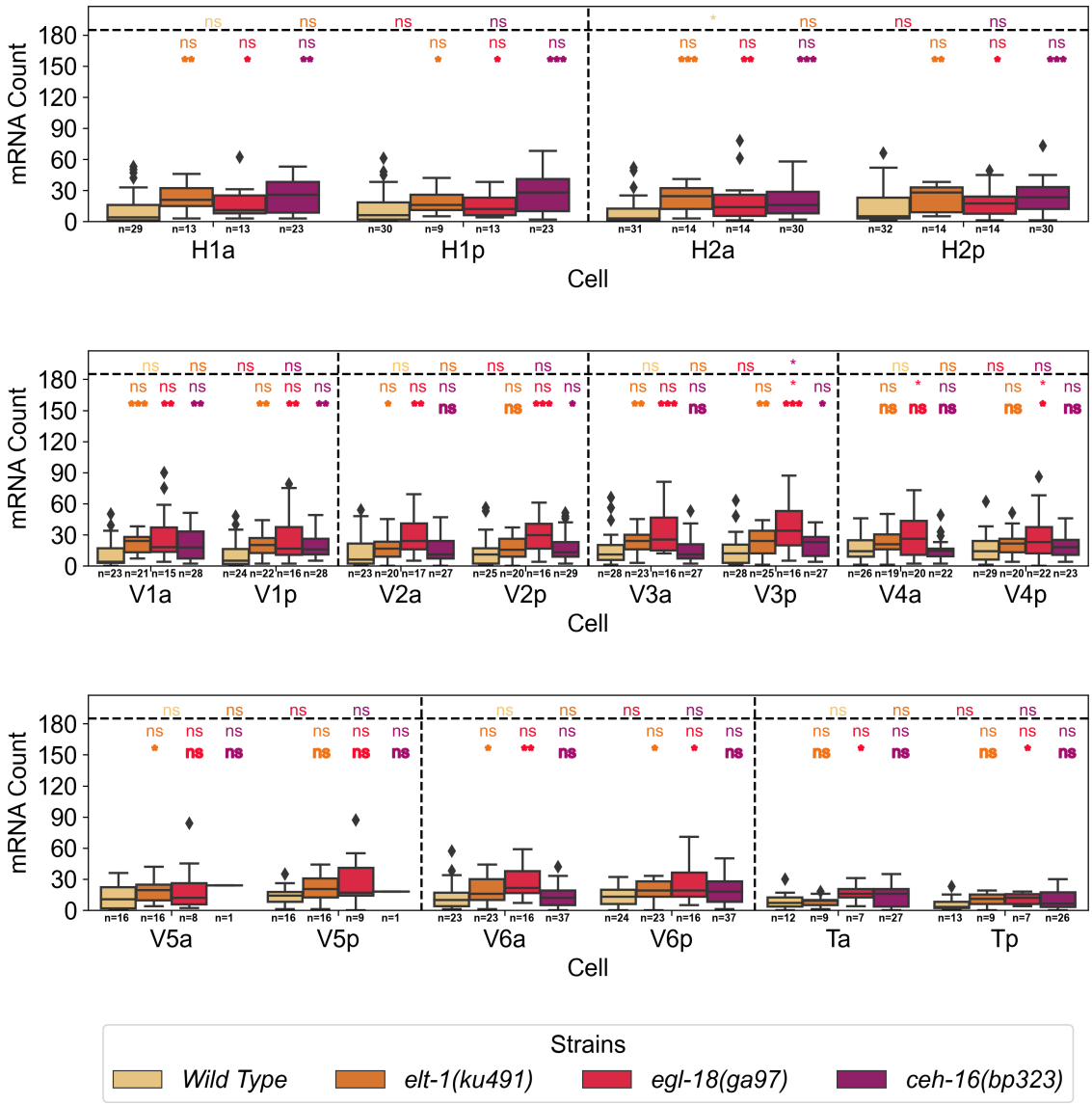


Figure 4.6: ***egl-18* expression 26 hours after bleaching, in the symmetric daughters in wild type, *ceh-16*, *egl-18* and *elt-1* mutants:** Quantification of smFISH signal in all seam cells. Strains are colour-coded as indicated in the key underneath the panel. The significance of the changes between mutant and wild type backgrounds is presented using the system described in the methods and the colour of the * matches the strain that is being compared to the wild type. The significance of the changes between mutant and wild type were calculated using Levene's test to compare variability and are shown above, whereas below and in bold, we have the comparison calculated using the Mann-Whitney U test to determine the difference in the distributions.

to a loss of *elt-1* and *egl-18* as evidenced by an increase in *egl-18* expression in all three cells in these mutants (Figure 4.6). *ceh-16(bp323)* only causes gains of *egl-18* expression in V1a, V1p, V2p and V3p (Figure 4.6). *egl-18* remains unchanged in V4, except for a slight gain in V4p in the *egl-18(ga97)* mutant (Figure 4.6). In V5, we observe gains in expression in the anterior cell in *elt-1(ku491)*, and in the posterior cell in *egl-18(ga97)*. In the T cell, we only note *egl-18* gains in the *egl-18(ga97)* mutant. We observe no decrease of *egl-18* in any of the mutants, and the variability of the expression remains relatively consistent with the wild type (Figure 4.6). Overall the sensitivity of *egl-18* after the symmetric division appears to be centred around increases in expression, particularly in absence of *elt-1* and *egl-18*. This is unusual as *egl-18* is associated with maintenance of seam cell fate, and all of these mutants produce a less stable seam cell phenotype (Gorrepati et al., 2013).

In the wild type *ceh-16* expression, we only note differences in expression between the anterior and posterior daughter in H2 (Figure 4.7). This difference is lost in *elt-1(ku491)* and *egl-18(ga97)* as the anterior cell greatly gains *ceh-16* expression (Figure 4.7). However, the asymmetry is strengthened in *ceh-16(bp323)* where the posterior cell sees the highest increase in *ceh-16* (Figure 4.7). None of the other cells show any asymmetry in the wild type. In the *ceh-16(bp323)* mutant, we note a slight gain in asymmetry in V1, V2 and V6 but none significant in V3 and V4 (Figure 4.7).

As with *egl-18* in the *ceh-16* mRNA counts we do not observe any losses in expression in the mutants but we do observe gains (Figure 4.7). The anterior head cells gain *ceh-16* expression in all three mutants, whereas the posterior cells show an increase in expression in *elt-1(ku491)* and *ceh-16(bp323)* (Figure 4.7). Additionally, we observe gains in expression in all the mutants in V1-V4, with the most significant increase observed in *elt-1(ku491)* and the smallest in *ceh-16(bp323)* (Figure 4.7). Similarly, in V5, we notice an increase in expression in the *elt-1* and *egl-18* mutants, although the change in expression is not statistically significant in *ceh-16(bp323)* (Figure 4.7). V6 displays increased *ceh-16* expression in *elt-1(ku491)*, and there is also elevated expression in the posterior cell in *ceh-16(bp323)* (Figure 4.7). Furthermore, in the T cell, we observe gains in *elt-1(ku491)*, along with a minor increase in expression in the anterior cell in the absence of *egl-18* (Figure 4.7).

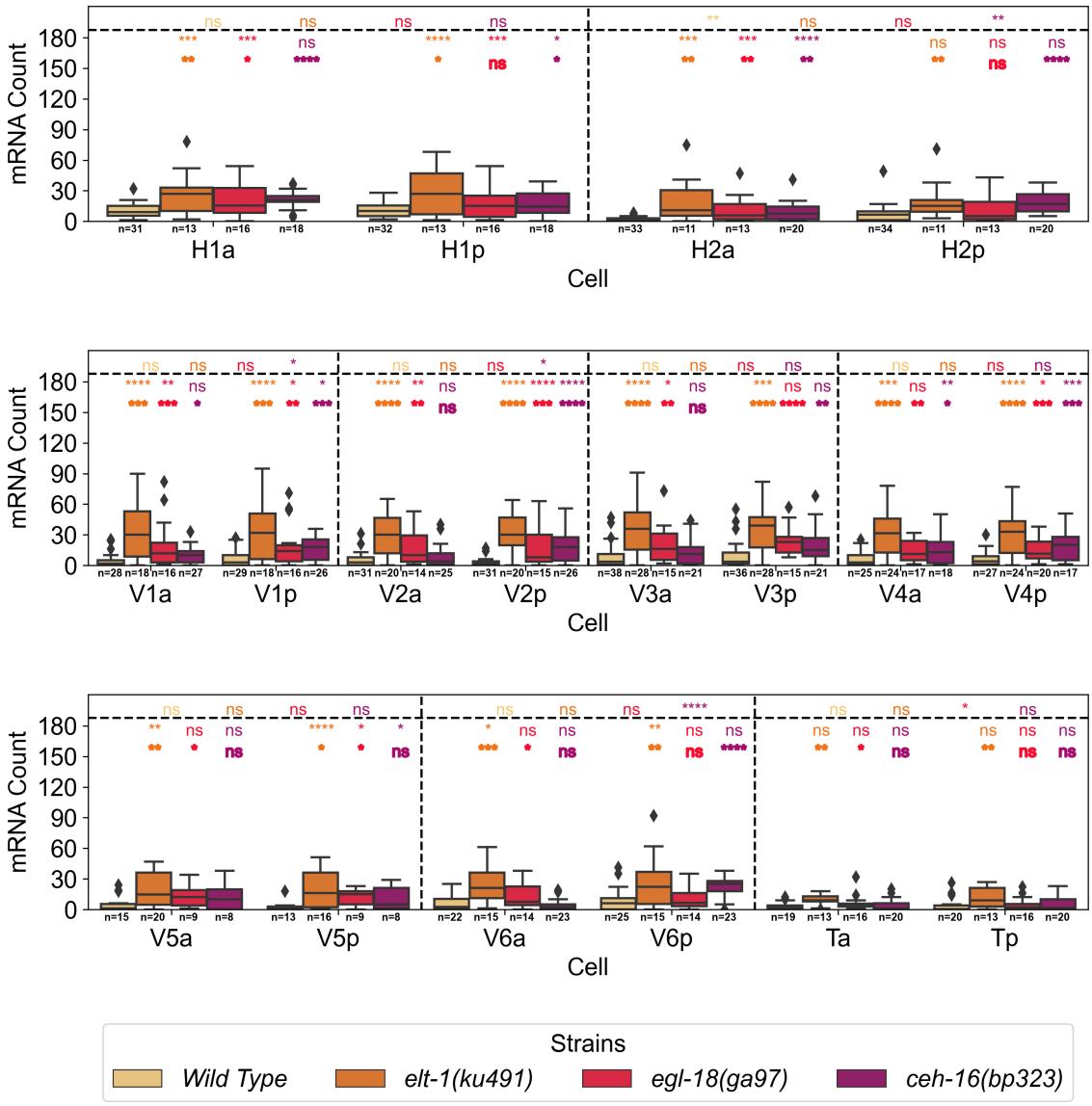


Figure 4.7: ***ceh-16* expression 26 hours after bleaching, in the symmetric daughters in wild type, *ceh-16*, *egl-18* and *elt-1* mutants:** Quantification of smFISH signal in all seam cells. Strains are colour-coded as indicated in the key underneath the panel. The significance of the changes between mutant and wild type backgrounds is presented using the system described in the methods and the colour of the * matches the strain that is being compared to the wild type. The significance of the changes between mutant and wild type were calculated using Levene's test to compare variability and are shown above, whereas below and in bold, we have the comparison calculated using the Mann-Whitney U test to determine the difference in the distributions.

The absence of *elt-1* not only increases the expression of *ceh-16* but also results in an increase in variability in all cells except the T cell (Figure 4.7). The gain in noise appears to occur together with an increase in expression levels. Overall *ceh-16* appears to be repressed by all three of the other core genes at this stage. Given that the same interractions appeared to be activatory at 17 hours, we hypothesise that mutations in the core seam cell gene network cause a delay in *ceh-16* expression.

In order to further unravel the interactions between the core seam cell genes, we utilized MRA to quantify the strengths of each transcriptional regulation and compared them to observations made at 17 hours. First, we examined pooled the expression data for all the cells together. At this time point, we noted strong repressive behavior from ELT-1 on both *egl-18* and *ceh-16* (Figure 4.8). While we did observe slight repression at 10 hours (Figure 4.4), it was considerably weaker than what we observed in the symmetric daughters (Figure 4.8). Additionally, the previous 17-hour time point shows ELT-1 in the role of an activator (Figure 3.5F). For the first time across our dataset, we also observed *egl-18* acting as a repressor of *ceh-16* (Figure 4.8). Furthermore, we observed for the first time an interaction between EGL-18 and *elt-1* (Figure 4.8). However, this interaction appears to be only moderately activatory, comparable to CEH-16's activation of *elt-1* at 10 hours (Figure 4.4). As in the 10 hour time point CEH-16 does appear to activate *elt-1* (Figure 4.8). The most robust link we observed is the repression from CEH-16 to *egl-18*, which persists in the merged data from all cells, albeit less strongly than in previous time points (Figure 4.8).

In the head cells, we detected cell-specific patterns. We observed weaker repression of *ceh-16* by both EGL-18 and ELT-1 (Figure 4.8). However, we also noted stronger repression of *egl-18* by both other core genes (Figure 4.8). Consequently, *egl-18* experiences slightly more repression, *ceh-16* slightly less, and *elt-1* encounters strong activation in the head cells (Figure 4.8).

Despite the anterior and posterior V cells being subject to an apparently symmetric division, we note at a gene expression level specific asymmetries. To address this, we segregated them for MRA to better comprehend their functional mechanisms (Figure 4.8).

The most noticeable change in the anterior V cells compared to other groups is the

absence of *egl-18* repression by CEH-16, marking the first observed change in this interaction across all time points (Figure 4.8). *egl-18* repression is also diminished by ELT-1. Conversely, *ceh-16* repression is stronger compared to head cells from both ELT-1 and EGL-18 (Figure 4.8). The activation of *elt-1* mirrors closely what is observed in the head cells (Figure 4.8).

Despite sharing functionality with the anterior V cells, the posterior V cells maintain high levels of *egl-18* repression by CEH-16 observed at other time points, alongside robust repression from ELT-1 (Figure 4.8). *ceh-16* regulation appears comparable to that in anterior cells, yet *elt-1* regulation by both other genes appears to induce considerable noise (Figure 4.8). Consequently, to visualize the data, we had to remove outliers. Despite the highly positive values assigned to the *elt-1* regulatory interactions, we cannot confirm a positive interaction due to very large error bars (Figure 4.8).

Overall the head cells experience strong activation of *elt-1* and strong repression of *egl-18* (Figure 4.8). They experience strong repression of *ceh-16*, however this is less striking than in the other cells (Figure 4.8). The anterior V cells lose the CEH-16 repression of *egl-18* which is also less repressed by ELT-1. Meanwhile *ceh-16* appears to experience strong repression from the other genes of the network (Figure 4.8). The difference from the anterior to the posterior V cells is primarily centred around the repression of *egl-18* which is much stronger in the posterior cells. The changes in *elt-1* regulation appear to be very noisy but positive, like in the anterior cells. Finally, there is no discernible difference in *ceh-16* regulation between the anterior and the posterior cells (Figure 4.8).

4.2.3 Constructing the seam cell gene network at the L2 asymmetric seam cell division

When discussing the daughters of the L2 asymmetric division we refer to the daughter of the anterior daughter of the anterior daughter of the symmetric L2 division with aa after the cell name, we refer to the posterior daughter of the anterior daughter of the L2 symmetric division as ap, we also refer to the anterior daughter of the posterior daughter of the symmetric L2 division as pa and finally the posterior daughter of the posterior daughter of the L2 symmetric

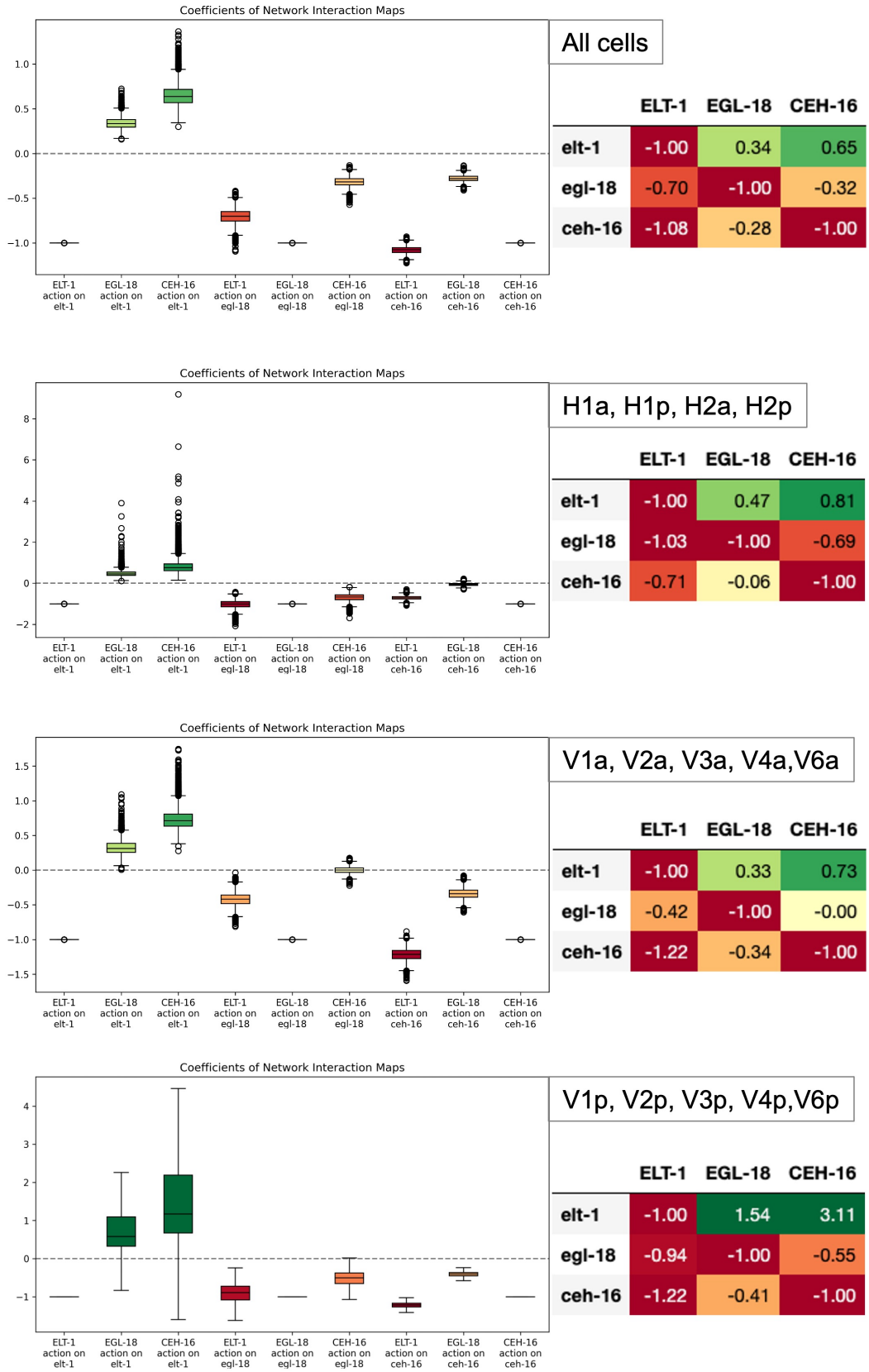


Figure 4.8: **Strengths of interactions between core genes in the symmetric L2 daughters calculated using MRA:** Illustration of the MRA matrix when combining first, all cells, next just the head cells (H1a, H1p,H2a,H2p) then the symmetric anterior daughters (V1a,V2a,V3a,V4a,V6a) and finally the symmetric posterior daughters (V1p,V2p,V3p,V4p,V6p). The colour coding correlates saturation to strength of interaction, and the colour to the polarity (green positive, red negative). The error bars show the range of response expected due to data variability and bootstrapping. Outliers were removed in this graph to maintain visibility of the interactions closer to 0.

division with pp after the cell name.

During the asymmetric L2 division, distinct differences in *elt-1* expression are observed between the differentiating cells and the non-differentiating cells in the wild type (Figure 4.9). However, these differences seem to diminish in mutant backgrounds (Figure 4.9). We observed a small yet significant decrease in *elt-1* expression compared to wild type in both *elt-1(ku491)* and *egl-18(ga97)* mutants. In contrast, *elt-1* expression in the *ceh-16(bp323)* mutant appears largely unchanged from wild type, except for the differentiating and non-differentiating cells exhibiting more equal expression levels (Figure 4.9). *egl-18* expression after the L2 asymmetric division appears to be equal between differentiating and non-differentiating cells (Figure 4.10). Its expression remains largely unchanged by the loss of either *elt-1* or *egl-18* from its wild-type expression (Figure 4.10). However, we observe more significant changes in both expression and symmetry in the *ceh-16(bp323)* mutant. In the absence of *ceh-16* the cells that maintain the seam cell fate appear to express significantly more *egl-18* than they do in the wild type (Figure 4.10). In the *ceh-16(bp323)* both the anterior asymmetrically divided cells (aa and ap) and the posterior ones (pa and pp) appear to gain asymmetry in *egl-18* expression, with the posterior cells expressing more *egl-18* (Figure 4.10). However, the pp cells appear to consistently have the strongest gains in *egl-18* (Figure 4.10).

At the end of the asymmetric L2 division it would appear that *ceh-16* expression is concentrated in the cell that maintains the seam cell fate, however due to gains in expression in the *elt-1(ku491)* and *egl-18(ga97)* mutants, this asymmetry is frequently lost (Figure 4.11). *ceh-16(bp323)* does also maintains an asymmetric expression of *ceh-16* (Figure 4.11). In all three of the mutant backgrounds we note gains in *ceh-16* expression, however this occurs mainly in V3 and V4 and less so in V1 and V2 (Figure 4.11). Finally, in running the MRA we realised that there was a large amount of variability in the gene expression patterns of our core genes at the asymmetric division stage (Figure 4.12). Particularly in the differentiating cells. However, we did manage to extract concrete trends in the eight non-differentiating cells. The MRA found seven very strong interactions and one weak one. These were activatory links from EGL-18 and CEH-16 to *elt-1* and repressive links from CEH-16 and ELT-1 to *egl-18* (Figure 4.12). We also observe ELT-1 repression of *ceh-16* (Figure 4.12). Overall in

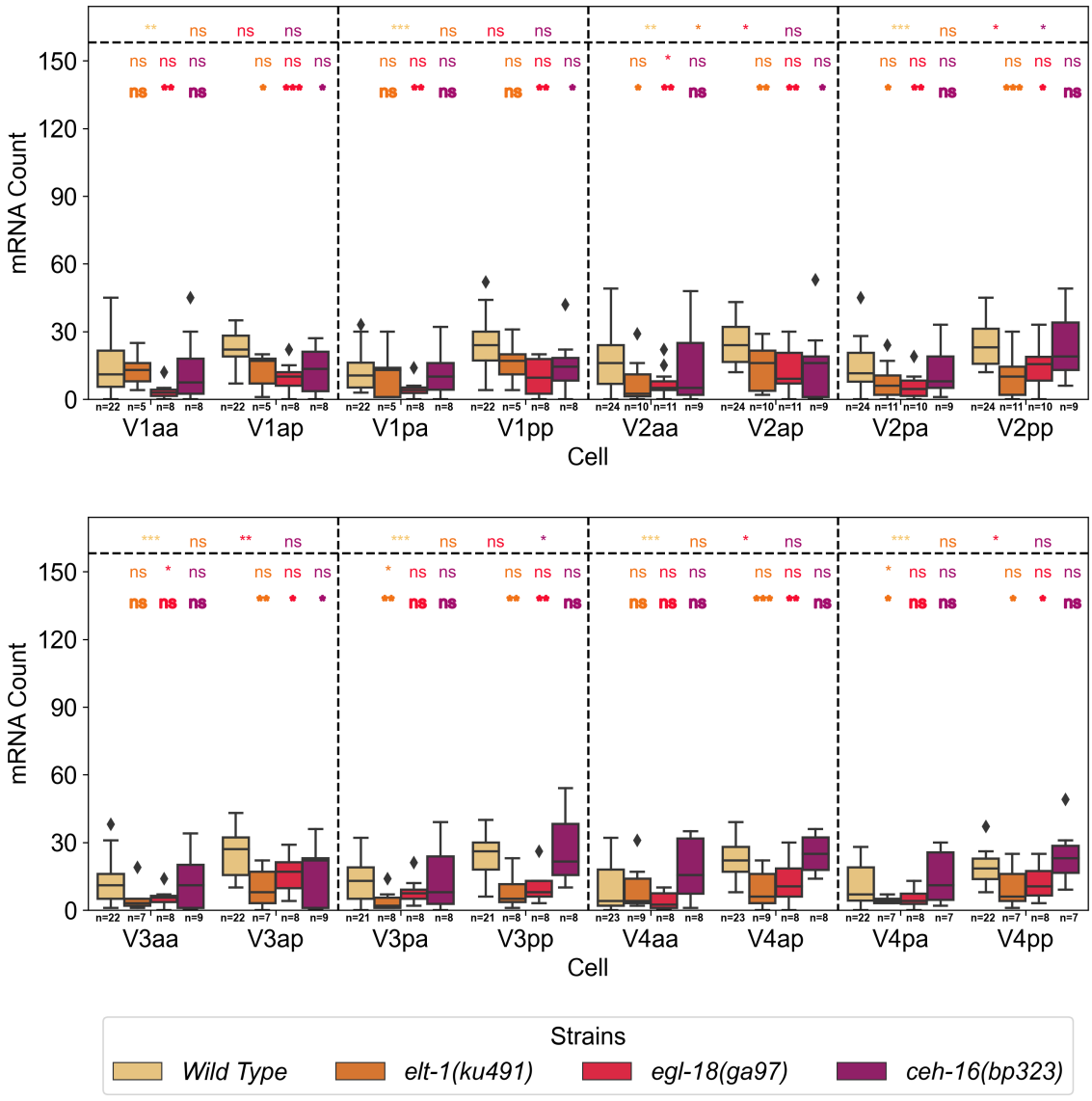


Figure 4.9: ***elt-1* expression 26 hours after bleaching, in the asymmetric daughters in wild type, *ceh-16*, *egl-18* and *elt-1* mutants:** Quantification of smFISH signal in all seam cells. Strains are colour-coded as indicated in the key underneath the panel. The significance of the changes between mutant and wild type backgrounds is presented using the system described in the methods and the colour of the * matches the strain that is being compared to the wild type. The significance of the changes between mutant and wild type were calculated using Levene's test to compare variability and are shown above, whereas below and in bold, we have the comparison calculated using the Mann-Whitney U test to determine the difference in the distributions.

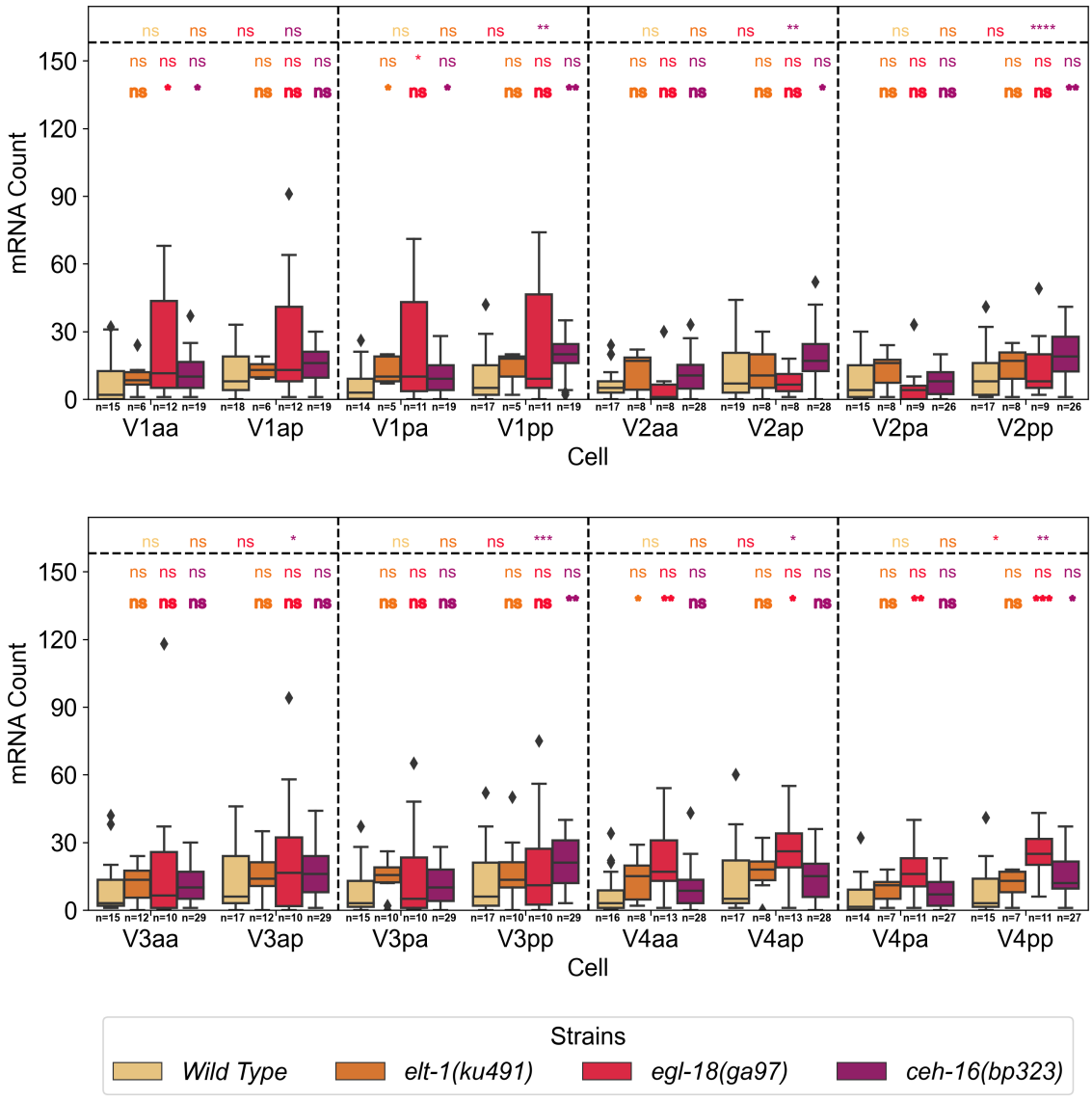


Figure 4.10: ***egl-18* expression 26 hours after bleaching, in the asymmetric daughters in wild type, *ceh-16*, *egl-18* and *elt-1* mutants:** Quantification of smFISH signal in all seam cells. Strains are colour-coded as indicated in the key underneath the panel. The significance of the changes between mutant and wild type backgrounds is presented using the system described in the methods and the colour of the * matches the strain that is being compared to the wild type. The significance of the changes between mutant and wild type were calculated using Levene's test to compare variability and are shown above, whereas below and in bold, we have the comparison calculated using the Mann-Whitney U test to determine the difference in the distributions.

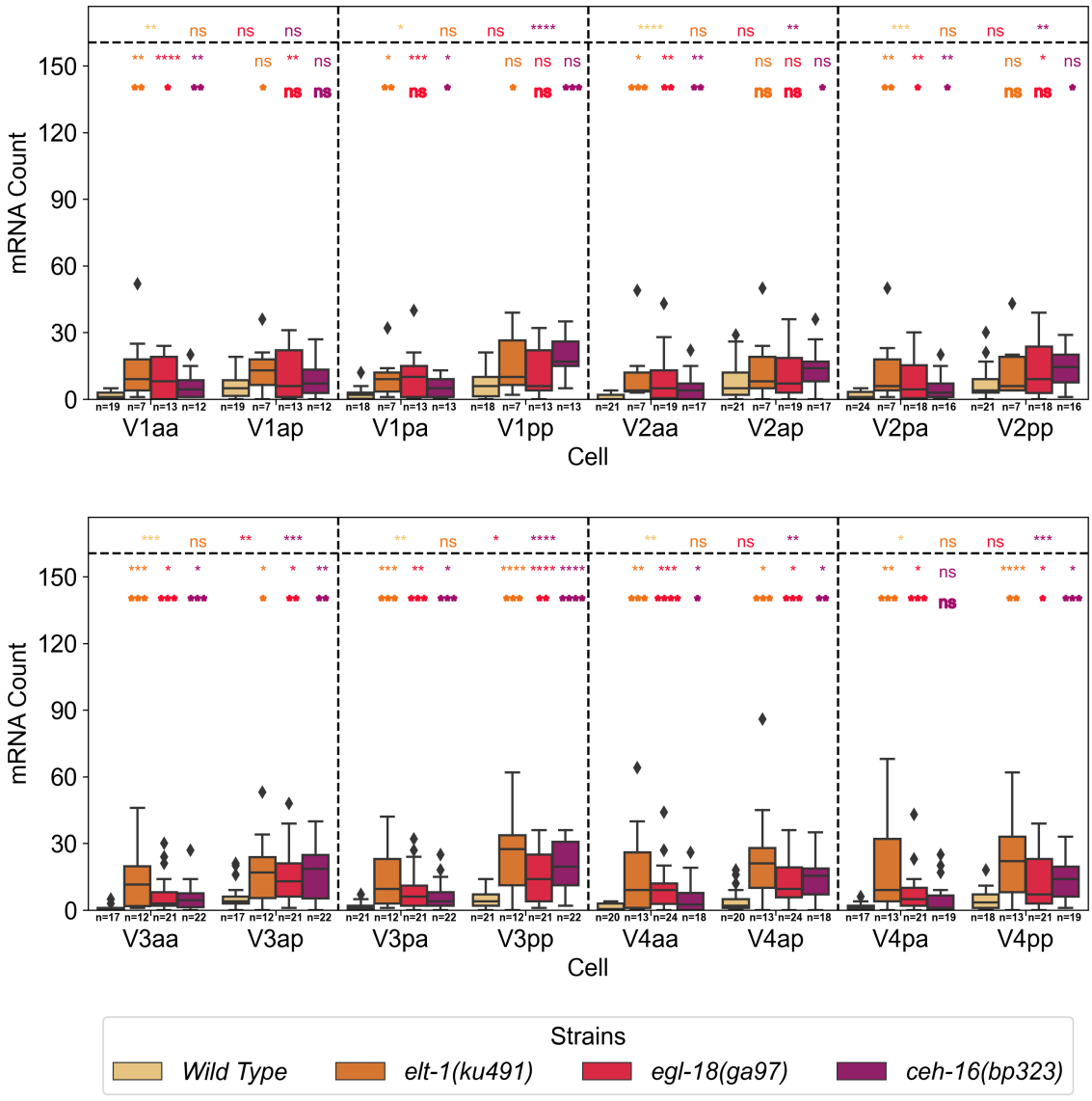


Figure 4.11: ***ceh-16* expression 26 hours after bleaching, in the asymmetric daughters in wild type, *ceh-16*, *egl-18* and *elt-1* mutants:** Quantification of smFISH signal in all seam cells. Strains are colour-coded as indicated in the key underneath the panel. The significance of the changes between mutant and wild type backgrounds is presented using the system described in the methods and the colour of the * matches the strain that is being compared to the wild type. The significance of the changes between mutant and wild type were calculated using Levene's test to compare variability and are shown above, whereas below and in bold, we have the comparison calculated using the Mann-Whitney U test to determine the difference in the distributions.

this state the network appears to be highly aimed at repression, leading us to believe that later developmental time points will experience an even more severe decrease in core gene expression in the seam cells.

4.2.4 The core seam cell genes follow an L2 centric expression pattern over time

To gain a comprehensive understanding of gene expression dynamics over time, we plotted our V1-V4 datasets from various time points, including 17 hours and 25 hours (limited data points collected alongside the 26-hour symmetric division). This enabled us to assess whether the 17-hour time point reflects steady-state conditions and whether specific gene expression information is elucidated only when considering multiple time points (Figure 4.13).

Our initial observation revealed striking similarities in the expression of both *elt-1* and *egl-18* between the 10-hour and 17-hour time points across wild type and mutant backgrounds. Notably, *egl-18* expression in *ceh-16(bp323)* mutants showed a substantial increase at 17 hours compared to 10 hours, despite significant variability. Conversely, wild type *ceh-16* exhibited a six-fold increase in mean expression at 17 hours compared to 10 hours, albeit with considerable variability (Figure 4.13). Another interesting observation we made about *ceh-16* is that while in the wild type we observe peak expression levels at 17 hours, in the mutants the peak occurs just before the symmetric division (Figure 4.13). This points to a delay in *ceh-16* expression rather than an outward repression. We also note that after the L2 asymmetric divisions pp and ap cells have higher gene expression in all mutants, than the aa and pa cells for all genes and all mutants. However, after the L2 symmetric division the a and p cells normally have similar behaviour, except in *ceh-16(bp323)* where the two diverge (Figure 4.13).

Another gene where peak gene expression changes in mutant backgrounds is *elt-1*. In the wild type, peak *elt-1* expression occurs at in the symmetric daughters. However, in all of the mutants the pre-division 25 hour time point shows expression that is either equivalent or higher than that of the symmetric daughters. This conversely would indicate an anticipation

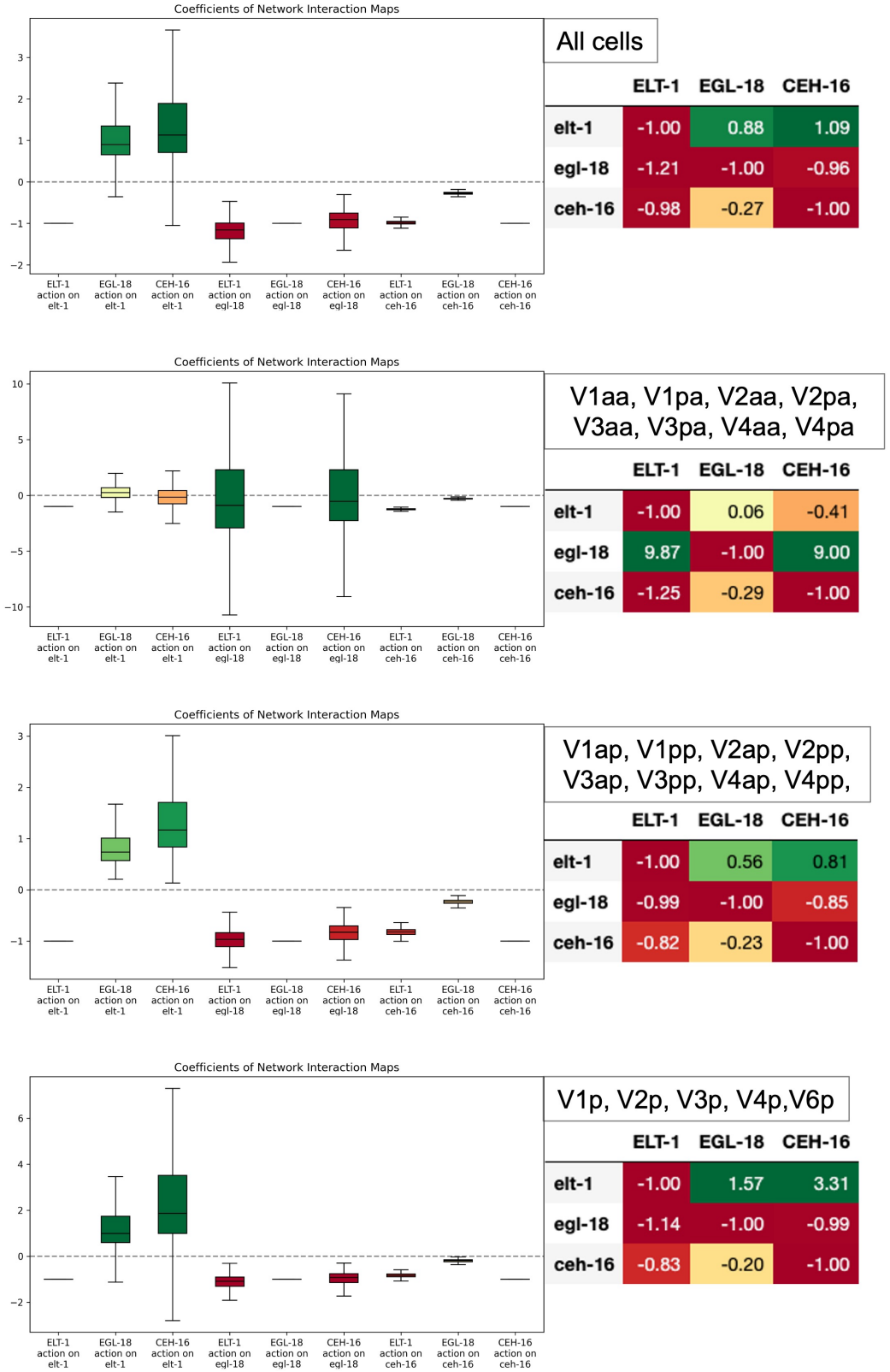


Figure 4.12: **Strengths of interactions between core genes in L2 asymmetric daughters calculated using MRA**: Illustration of the MRA matrix when combining first, all cells, next the differentiating cells (V1aa, V1pa, V2aa, V2pa, V3aa, V3pa, V4aa, V4pa then the non-differentiating cells (V1ap, V1pp, V2ap, V2pp, V3ap, V3pp, V4ap, V4pp) and finally the symmetric posterior daughters (V1pp, V2pp, V3pp, V4pp, V6pp). The colour coding correlates saturation to strength of interaction, and the colour to the polarity (green positive, red negative). The error bars show the range of response expected due to data variability and bootstrapping. Outliers were removed in this graph to maintain visibility of the interactions closer to 0

of *elt-1* expression (Figure 4.13).

Overall our assumption that the 17 hour time point is a good approximation of steady state behaviour is only a reasonable assumption for *elt-1* and *egl-18*. *ceh-16* expression appears more variable in early larval development and thus we believe it would only achieve a stable steady state at time point where it was inactive, as it is in late larval development when the cells acquire a terminal seam cell fate. We do note that there appear to be time specific changes in gene expression that snap-shot data may not have the resolution to assess in detail. Nevertheless, we do believe that investigating this network from a late L1 perspective and then reevaluating it at other time points allowed us to greatly progress in our understanding of seam cells.

4.2.5 Significant differences exist between gene expression in different seam cells at various time-points

Having increased the resolution of the behaviour of the core genes through time, we must also evaluate how position affects the different seam cells and whether that has any specific impact on their ability to maintain the seam cell fate. In order to do this we re-utilise our temporal data to gain further understanding on the changes of the behaviour of the individual cells over time.

Despite the variations in division patterns among seam cells, seam cells are frequently treated as equal. To scrutinize this assumption, we analyzed our smFISH data at each time point and conducted Mann-Whitney U tests to assess the similarity of expression patterns between individual cells. We used this to determine whether the cells can be deemed to have identical expression patterns or whether the differences between them were statistically significant. We then further separated the cells using clustering to see which cells were most likely to show mean expression similarity.

Already at the 10-hour time point, we observed highly significant differences in expression between seam cells in both the wild type and all three mutants (Figure 4.14A). These are however very similar to what we observe at 17 hours (Figure 7.1).

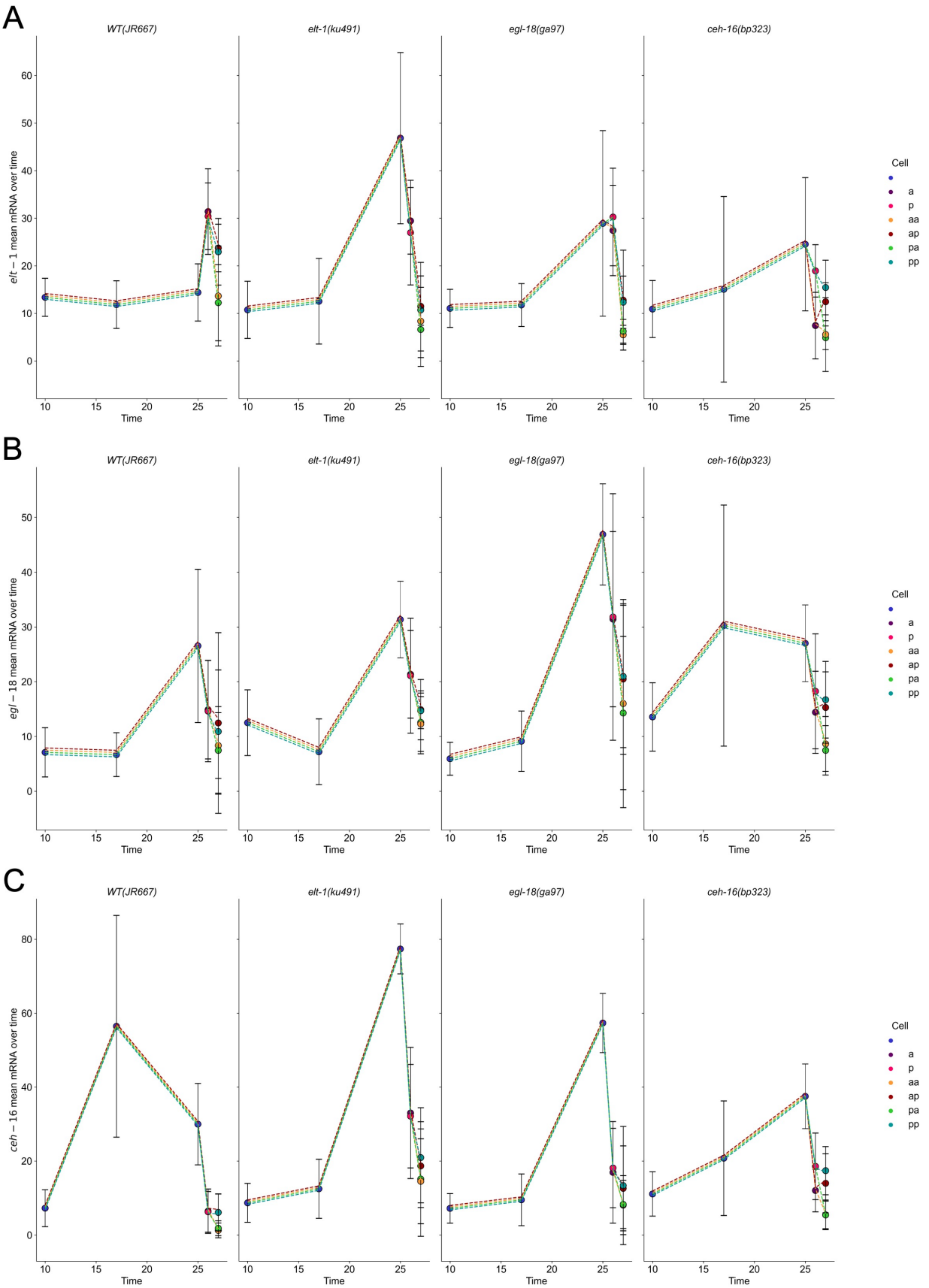


Figure 4.13: **Core gene expression throughout early larval development A)**The progression of *elt-1*, **B)** *egl-18* and **C)** *ceh-16* expression in V1-V4 over the course of the first two larval stages. The error bars denote 90% of the data to eliminate outliers and the points on the graph indicate mean expression levels. In blue we refer to cells in L1 and early L2 that have either not divided yet or are the posterior daughters of the L1 asymmetric divisions, a and p refer to the anterior and posterior symmetric daughters and aa,ap,pa and pp refer to the quadruplets formed after the L2 asymmetric division, in order of position.

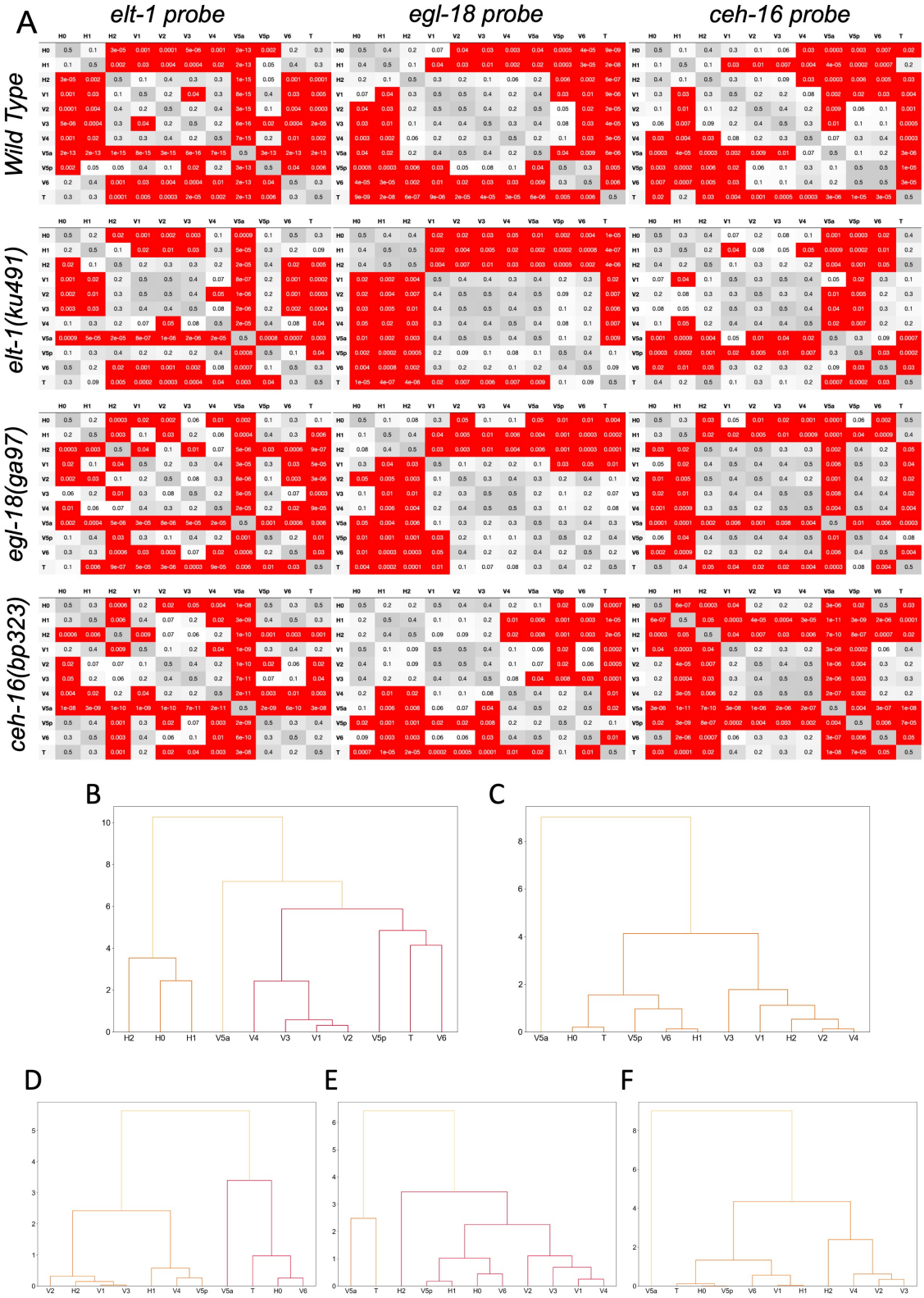


Figure 4.14: **At early L1 seam cell core gene expression shows similarities and differences between groups of seam cells: A)** A matrix of Mann-Whitney U p-values for each of the three genes in wild type and mutant backgrounds comparing each cell to each other cell to determine which cells' expression patterns can be considered identical. The red colour reflects highly significant values with p $<$0.05 and the gray colour reflects non-significant values with darker gray p-values being closer to 0.05.**B-F)** The clustering of the three genes' expression. The y-axis denotes the distances between the average means, with each graph grouping together the three gene expression patterns in different mutant backgrounds: **B)** wild type and all three mutants together, **C)** wild type alone, **D)** *elt-1(ku491)*, **E)** *egl-18(ga97)* and **F)** *ceh-16(bp323)*.

To better understand whether there are related groups among cells' expression patterns in different strains, we performed hierarchical clustering. This would allow us to correlate expression pattern trends with likelihood of seam cell loss. Initially, we clustered the expressions of all three genes together in all four strains (Figure 4.14 B), resulting in distinct clusters for head, midbody, and posterior cells, with the differentiating V5a cell forming a separate cluster. Next, we clustered the three genes in wild type animals (Figure 4.14 C)and then in *elt-1(ku491)*, *egl-18(ga97)*, and *ceh-16(bp323)* in Figure 4.14 D,E and F respectively.

In terms of *elt-1* expression patterns, we observed high similarity between H0 and H1 across all backgrounds, while H2 exhibited expression patterns similar to the V cells. Likewise, *egl-18* and *ceh-16* expression showed no significant difference between H2 and V1-V4 in the wild type. However, *egl-18* expression in (*elt-1(ku491)* and *egl-18(ga97)*), H2 exhibited expression patterns mainly resembling head cells and, in the latter case, V1, differing significantly from other seam cells. We also noted that the head cells showed similarity in expression to the T cell, particularly in *elt-1*. Additionally, head cells showed expression similarity to the T cell, particularly in *elt-1* expression. Clustering analysis grouped H2 with the V cells in wild type and mutant conditions (Figure 4.14C-F)

In all mutants, *elt-1* expression showed no statistical difference between V5p and V6. Notably, *egl-18* expression in the tail-most cells exhibited highest similarity between V5p and V6, although a small but significant difference was observed between V6 and the T cell in the *ceh-16(bp323)* mutant (Figure 4.14 A). Despite some variations, posterior seam cells' core gene expression tended to cluster together (Figure 4.14B).

V1-V4 were the cells that consistently showed no significant difference between them. The only condition where we noted any significant difference between these four cells was between V1 and V4 in *elt-1* expression in the *ceh-16(bp323)* mutant with a significance of 0.04. In terms of clustering, we noted that while in the pooled set, the wild type and the *egl-18(ga97)* clustering, the four cells were grouped into a closely-knit cluster (Figure 4.14 B, C, E), in *ceh-16(bp323)*, V1, and in *elt-1(ku491)*, V4 clustered more closely with H1 and V6 or V5p, respectively (Figure 4.14 D, F).

V5a displayed significant differences in both *elt-1* and *ceh-16* expression compared to

other cells, particularly pronounced in the *ceh-16(bp323)* and *egl-18(ga97)* mutants. Despite these differences, *egl-18* expression in V5a remained similar to other seam cells, especially under mutant conditions (Figure 4.14C-F).

V5a was also the cell that had the highest clustering distance from the other cells in the pooled data (Figure 4.14 B) and wild type (Figure 4.14 C), as well as all three mutants (Figure 4.14 D-F), with a slight similarity to the T cell in the *egl-18(ga97)*, indicating that despite the similarities in *egl-18* expression, the differentiating cell had the most diverging behavior of the core genes from all the seam cells.

In summary, V1-V4 displayed the strongest grouping, while V5a exhibited the most distinct expression patterns. Furthermore, the head cell group demonstrated less distinct clustering, with H2 and H1 tending to cluster with different V cells depending on conditions.

At 26 hours in the symmetric division, *elt-1* is expressed most similarly in H2p, V1a, V1p, V2a and V2p. H2a only shows similarity in behaviour with V5a and Tp, suggesting that that specific expression signature can be linked to differentiation. In *elt-1(ku491)*, H2a acquires similarity in expression to H1a and H1p as well as V6a, V6p and Ta. In all cases where cells resemble H2a in *elt-1* expression, we note lower *elt-1* mRNA counts than in the other cells (Figure 4.5). In the wild type we also note similarities between H1a, H1p and H2p,V1a,V1p,V2a, V2p, which is lost in *elt-1* and *egl-18* mutants, however the H1 similarity to V6 is maintained also in the mutants (Figure 4.15 A).While comparable in expression, V3 and V4 have slightly higher *elt-1* expression than the other V cells.

In wild type the V4 cells express *egl-18* at significantly higher levels than any of the eight anterior cells (H1a, H1p, H2a, H2p, V1a, V1p, V2a, V2p) as well as the T cells. Tp expresses lower *egl-18* than any of the posterior V cells (V3a, V3p, V4a, V4p, V5p, V6a, V6p). H2a exhibits a similar expression pattern to the T cell in this regard, as its expression is significantly lower than that of any of those cells as well. Notably, H2a's expression is lower than even V2p, which represents the cell with the lowest expression among the anterior V cells.

In the *elt-1(ku491)* mutant the differences in *egl-18* expression between the V cells are decreased. The only significant differences can be observed between V2 and V3 and even

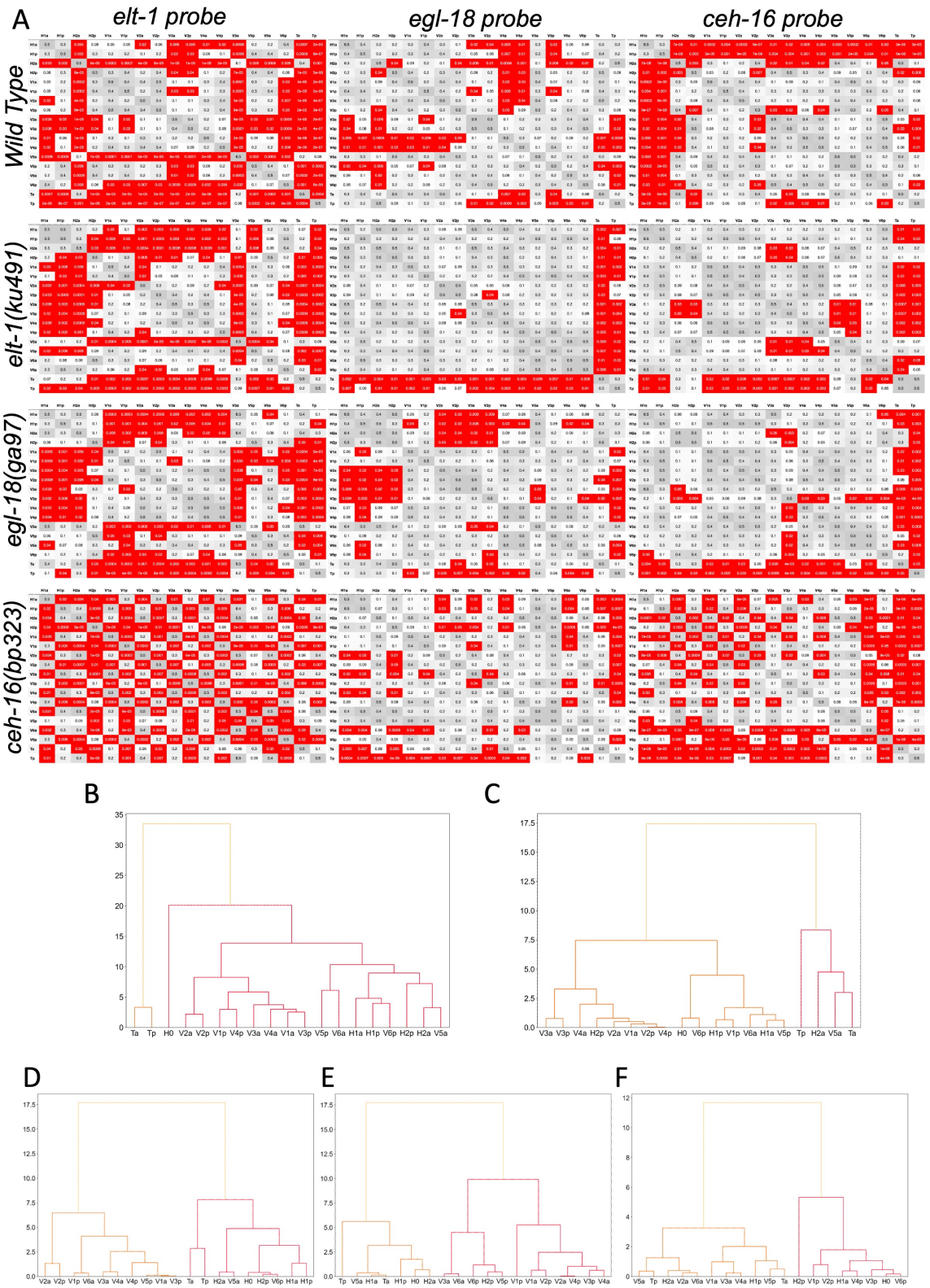


Figure 4.15: **At early L2 seam cell core gene expression shows similarities and differences between groups of seam cells: A)** A matrix of Mann-Whitney U p-values for each of the three genes in wild type and mutant backgrounds comparing each cell to each other cell to determine which cells' expression patterns can be considered identical. The red colour reflects highly significant values with p $<$0.05 and the gray colour reflects non-significant values with darker gray p-values being closer to 0.05 . **B-F)** The clustering of the three genes' expression. The y-axis denotes the distances between the average means, with each graph grouping together the three gene expression patterns in different mutant backgrounds: **B)** wild type and all three mutants together, **C)** wild type alone, **D)** *elt-1(ku491)*, **E)** *egl-18(ga97)* and **F)** *ceh-16(bp323)*.

those are relatively minor compared to what is observed in the wild type. However, the T cell shows significantly fewer *egl-18* counts than any of the other cells, including those that are differentiating (H2a and V5a). In the *egl-18(ga97)* mutant we observe a striking difference between the 4 head cells (H1a, H1p, H2a and H2p) from the middle 4 V cells (V2a, V2p, V3a, V3p). The head cells express less than half of the *egl-18* expressed in the V2 and V3. This decrease in expression in the head cells renders them more similar in expression pattern to the T cells. Additionally, V5a and V1p also strongly resemble the expression pattern found in head cells, suggesting a potential inclination towards differentiation in these cells.

ceh-16 in wild type behaves differently in H1a and H1p to all the other cells, with significantly higher expression levels observed. H2a, the differentiating cell, shows decreased expression compared to H2p, V3a, V3p, V4a, V4p, and V6p, but not compared to the anterior V cells (V1 and V2). V2p in particular appears to experience low *ceh-16* expression in the wild type. However, the pattern of seam cell *ceh-16* expression undergoes drastic changes in the *elt-1(ku491)* mutant. Due to the increase in *ceh-16* expression in absence of *elt-1*, we observe that despite the levels of *ceh-16* in the T cells being higher than most of what we observe in the other cells in the wild type, their expression levels are still lower than what we see in the other cells in this *elt-1(ku491)* mutant. Thus the T cells will express less *ceh-16* compared to other cells even in mutants that cause an increase in *ceh-16* expression. Apart from this, the remaining cells' *ceh-16* expression levels show high similarity, with the only significant inter-cell differences observed in V3a and V3p compared to H2a, H2p, V5a, and V5p. The *ceh-16* counts in V3a and V3p are significantly higher than in the four cells that did not divide symmetrically. This pattern is highly similar to what we observe in the *egl-18(ga97)* mutant. This pattern is highly reminiscent of what is observed in the *egl-18(ga97)* mutant. Here, the T cell again shows very little expression of *ceh-16*, while V3 cells, particularly V3p, exhibit the highest *ceh-16* expression among all the cells. This difference is not significant relative to the H1, V1, and V2 cells, which still have relatively high expression levels; however, the remaining cells all express *ceh-16* at significantly lower levels than V3p.

Finally, the most striking pattern we observed was in the *ceh-16(bp323)* mutant. Most prominently in *elt-1* and in *ceh-16* expression, but also to a lesser extent in *egl-18* expression,

we observe the asymmetrisation of the core genes in the symmetric daughters. This is particularly striking where each 'a' cell is significantly similar to the others and different from the 'p' cells. This pattern is reversed in H1 where H1a shows no significant difference from all the V p cells and shows highly significant differences to the a cells. The other exception to the rule are the T cells which have high similarity between them and both show similarities in expression pattern to the anterior cells but remain significantly different from the posterior cells. In *ceh-16* expression in *ceh-16(bp323)* we note this pattern in H2, V1, V2, and V6 and, while there are differences in V3a/V3p and V4a/V4p, these are not deemed significant. V4p does however show significant differences in expression pattern to H2a, V1a, V2a, and V3a. The other interesting pattern resulting from *ceh-16* expression in *ceh-16(bp323)* is a significant difference between V6a, V6p, Ta and Tp. In the case of V6a, Ta and Tp its due to particularly low expression, however in V6p we note particularly elevated *ceh-16* levels which causes very large asymmetry between anterior and posterior V6.

In order to summarise our findings and make predictions on the behaviour of the cells in mutant contexts we clustered our three gene's expression in both wild type and mutant backgrounds. The first approach involved clustering all four strains together (Figure 4.15 B). The most distinct cell was the T cell, primarily due to its low core gene expression. Following was H0, which formed its own distinct cluster. V1, V2, V3, and V4 clustered together, further supporting the validity of grouping these four cells to study their behavior collectively. Notably, among these four cells, V2 exhibited the highest difference from the other cells. H1, H2, V5, and V6 clustered in the fourth cluster.

In the clustering carried out solely in the wild type background we do note significant changes from 10 and 17 hours (Figure 4.15C). While most of the V1-V4 cells still clustered together, they also clustered with H2p, while V1p clustered with H1, H0, V6, and V5p. The third cluster comprised the two differentiating cells: H2a, V5a, along with the T cells. In the *elt-1(ku491)* mutant, we again observed V1-V4 clustering together, this time along with V5p and V6a. There is a second large cluster in *elt-1(ku491)* and this is composed of the T cells, the H1 cells, the two differentiating cells (H2a and V5a) and H0, H2p and V6p. In the *egl-18(ga97)* the V1-V4 cluster together again into a single cluster, with the exception of V3a

which clusters with V6, V5p and H2p. H0, H1, H2a, V5a and the T cells cluster separately. Finally, as described previously, the most significant change in seam cell similarity in gene expression at the L2 symmetric division time point was observed in *ceh-16(bp323)*. The clustering draws a clear line between the anterior daughters of a symmetric division and the posterior daughters. The clustering drew a clear distinction between the anterior daughters of a symmetric division and the posterior daughters. H2p clustered more distinctly, but V1p, V2p, H1a, V4p, V3p, H0, and V6p all clustered together. We'd expect these cells to have more robust phenotypes due to their being grouped with H0. The remaining cells are the anterior daughters that are frequently lost in *ceh-16(bp323)* due to the loss of the symmetric division. The final developmental stage that we studies was the late L2 stage following the asymmetric division. We excluded the anterior differentiating cells from our intracellular comparison as they exhibited clear differences from the non-differentiating cells (Figure 7.2). The number of animals studied were also fewer as cells were frequently lost in the mutant animals. At this stage the cells show extremely high similarity in expression (Figure 4.16 A). In the wild type the only large difference between the cells is between V4pp and the remaining 7 cells in *elt-1* expression. The only cell with which the difference is not significant is V4ap. V4pp exhibited lower expression than the other cells. This is lost in the *elt-1(ku491*), where for all probes the cells behave similarly (Figure 4.16 A). *egl-18* expression in *egl-18(ga97)* replicates the wild type pattern of V4 behaving differently to the other V cells, particularly V2ap and V2pp. In the *ceh-16(bp323) elt-1* shows a difference between V3pp, V4ap, V4pp and V1ap, V1pp, V2ap (Figure 4.16 A). The low significance values could be due to the low counts, however the trend points to high expression of *elt-1* in the posterior V cells and low in the anterior V cells in *ceh-16(bp323)*. There are also cell specific differences in *egl-18* expression in this mutant, with V4pp having lower expression than any of the other pp cells. The *ceh-16* mutant also shows differences in *ceh-16* expression with V1ap having the lowest expression of any non differentiating cell (Figure 4.16 A).

In order to formulate hypotheses on the behaviour of the individual cells in response to the mutants we relied again on clustering. The clustering of all the probes together with all the mutants showed the most distinct expression pattern in V4pp (Figure 4.16 B). The

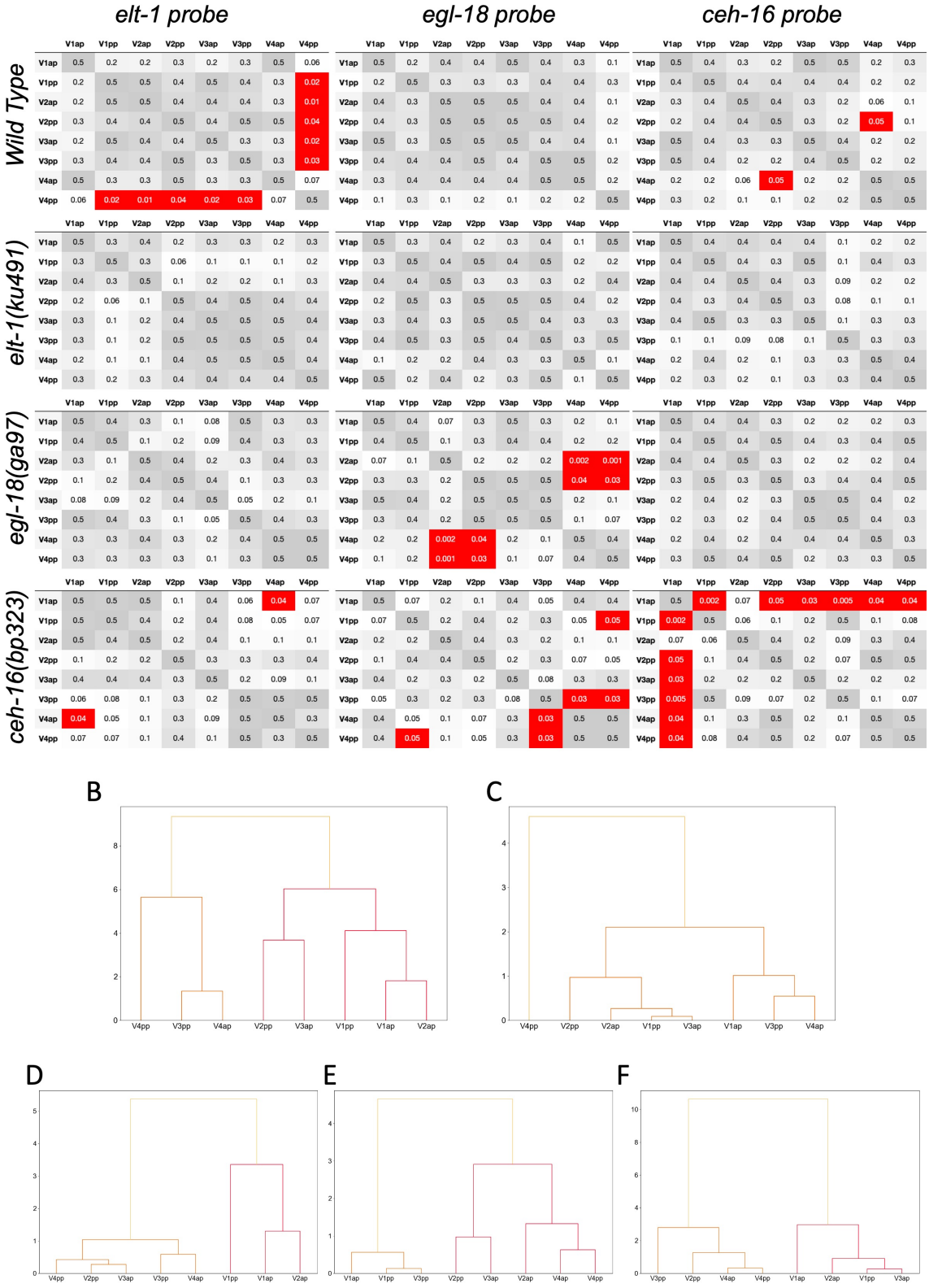


Figure 4.16: **At late L2 seam cell core gene expression shows similarities and differences between groups of seam cells: A)** A matrix of Mann-Whitney U p-values for each of the three genes in wild type and mutant backgrounds comparing each cell to each other cell to determine which cells' expression patterns can be considered identical. The red colour reflects highly significant values with p $<$0.05 and the gray colour reflects non-significant values with darker gray p-values being closer to 0.05. **B-F)** The clustering of the three genes' expression. The y-axis denotes the distances between the average means, with each graph grouping together the three gene expression patterns in different mutant backgrounds: **B)** wild type and all three mutants together, **C)** wild type alone, **D)** *elt-1(ku491)*, **E)** *egl-18(ga97)* and **F)** *ceh-16(bp323)*.

rest of the clustering pointed to two separate groups. One containing V1pp, both the V2s and V3ap, and the other V1ap, V3pp and V4ap. In the wild type, the clustering showed clear positional tendencies: the two V4 cells clustered with V3pp, while separately, the V1s clustered with V2ap (Figure 4.16 C). Additionally, the intermediate group of V2pp and V3ap also preferred to cluster with the V1s. The split between anterior and posterior V cells became more pronounced in the three mutants, particularly in *elt-1(ku491)* (Figure 4.16D) and *ceh-16(bp323)* (Figure 4.16F), but the boundary is shifted between them. In *elt-1(ku491)*, the V1 cells and V2 were grouped together. This group was characterized by high *elt-1*, low *egl-18*, and low *ceh-16* expression. The *ceh-16(bp323)* clustering appears to be largely driven by changes in *elt-1* which has heightened expression in the V4s and V3pp and V2pp. All the other genes appear to be equivalent across all cells. The *egl-18(ga97)* mutant clustered V1s together with V3pp, then the proximal V2pp with V3ap, and finally V2ap with the V4s.

Overall this highlights the validity of using V1-V4 as a group for most early stage experiments. However, in the late L2 stage we start observing differences between V1, V2 and V3, V4. As such it is important to maintain cell specific resolution when looking at later developmental time points in seam cells.

4.2.6 Understanding how expression differences in the core genes impact different seam cell sensitivity

Differences in susceptibility of a seam cell to be lost could be related to the differences in core gene expression between the seam cells not only in mutant but also in wild type backgrounds. For example, we'd expect cells that frequently cluster with differentiating cells to have a decreased ability to maintain seam cell fate. However, given that the groupings of the cells change over time it is important to understand which time point determines susceptibility to loss and which gene drives the changes. We already know from Huang et al. (2009) that symmetry in the L2 division is lost in *ceh-16(bp323)* and as such anterior cells acquire a non-seam fate. The illustration of the expression changes in the L2 symmetric division is clear evidence that core seam cell genes are linked directly with seam cell losses, as the

cells that are lost have significantly lower *elt-1* and *ceh-16* with little change in *egl-18*. We then need to evaluate whether a similar pattern of gene expression can be identified in the cells more commonly lost in *elt-1(ku491)* and *egl-18(ga97)* mutants.

To understand the phenotypic implications of the observed expression changes, we decided to evaluate the frequency of different cells' losses at different concentrations of the core genes. In order to do this we utilised the RNAi by feeding technique. We diluted bacteria containing RNAi vectors targeting specific genes with bacteria containing an empty RNAi vector to decrease the intensity of the treatment in order to allow us to observe which cells are most sensitive to changes in RNAi concentration. Unfortunately, in order for this to work the RNAi has to show a highly significant phenotype at 100% concentration, however not so strong that the cells became impossible to differentiate between each other. This meant that for this experiment we focused on *elt-1* RNAi exposure for one generation and *egl-18* RNAi treatment for two consecutive generations. *elt-1* RNAi treatment for two generations frequently resulted in a phenotype below 5 cells which was difficult for determining which cell is which as the cells were often not located near the markers we used for differentiating between seam cells. *egl-18* RNAi treatment of eggs showed a very weak phenotype after one generation of animals reached L4 even at 100% concentration. Likewise we couldn't conduct this experiment with *ceh-16* RNAi as the resulting phenotype was too weak to monitor its changes over serial dilutions. The percentages of RNAi indicate the proportion of bacteria of each strain mixed before plating. However, the slower growth rates of bacteria containing *elt-1* and *egl-18* RNAi vectors compared to empty vectors may result in a different final composition of the plate lawn.

elt-1 exhibits a substantial increase in variability even at 25% RNAi (Figure 4.17. However, at 25% *elt-1* RNAi, we observe the average seam cell number decrease only to $\approx$ 15. 50% and 75% *elt-1* RNAi result in larger decreases of seam cells with an average of $\approx$ 13 and 12 respectively. Finally, we observe a significant drop in the number of seam cells at 100% RNAi, with an average of 8 seam cells and some animals having as few as a single seam cell. The quantity of *elt-1* appears to be directly proportional to the number of seam cells present (Figure 4.17). Regarding the cells lost in *elt-1* RNAi treatments, we identify a highly specific

pattern. Despite H0 typically being considered the most robust cell, the least sensitive cell across all concentrations was T. As such, we will refer to it as a control, and cells whose losses are not statistically higher than those observed in T will be considered not significantly lost (Figure 4.17).

At 25% *elt-1* RNAi, despite the minimal change in seam cell counts, we observe highly significant losses in H1p, V1a, V1p, and V2a. In H1a, we also note a statistically significant difference from T, however this is less strong than the other 4 anterior cells. Meanwhile the posterior seam cells show no significant difference from the immutable T(Figure 4.17). This categorises the cells into sensitive (H0, H1p, H2, V1a, V1p, V2a,) and less sensitive (V3a, V3p, V4a, V4p, V5, V6a, V6p and T).V2p and H1a appear to be halfway between the sensitive group and the insensitive group, especially V2p (Figure 4.17). In terms of V2p losses, while not significantly different from T it does not show significant differences from the other cells except H1p and V6a which belong to different groups. H1a shows borderline significant differences from the sensitive group, and insignificant differences to the less sensitive group (Figure 4.17).

At 50% *elt-1* RNAi, we observe an expansion of the cell group sensitive to treatment, now encompassing V2p, V3a, V3p, V4a, and V4p. Subsequently, at 75%, a small percentage of losses (10%) are also noted in V6a and V5 (5%). At 50%, occasional losses are observed even in H0, with more pronounced losses at 75%. Despite the fact that H0 is only lost when H1a, H1p and frequently H2 are also lost this indicates the overall instability of the seam cells under the 75% *elt-1* RNAi treatment. However, the loss of H0 is not significant relative to T. At 50% *elt-1* RNAi concentration, although the posterior V cells (V3a, V3p, V4a, V4p) are significantly less robust compared to T, they still maintain significant differences from the other seam cells, particularly V4 relative to V1a, V1p, H2, and H1p. At 75%, the borderline behavior observed in V2p at 25% is now exhibited by V3a(Figure 4.17).

At 100% RNAi we see significant losses in all cells except T. However, once again, certain cells are lost with higher frequencies than others. Specifically, H1p, H2, V1a, V1p, and V2a exhibit very strong losses in most animals and do not show any statistical difference between them. Similarly, V3p, V4a, V4p, V5, and V6a display high similarity between them while

demonstrating less sensitivity. H1a and V6p appear not to belong truly to either group, and the same applies to H0 and T.In contrast, V2p and V3a show similarities to both groups. These findings suggest strong positional trends when *elt-1* RNAi is administered to wild-type animals (Figure 4.17).

The *elt-1(ku491)* mutant shows behaviour that seems to resemble the 25% RNAi treatment, showing the accuracy of our estimate of ELT-1 function in *elt-1(ku491)*, with two notable differences. In the mutant we see strong losses in the T cell. We did observe occasional losses in the T cell at 25% RNAi but this was much less frequent than what we observed in the mutant. Secondly, across all treatments, no significant gains are observed, except in the case of the *elt-1(ku491)* mutant, where we do note some gains in V3a and T. However, these occurrences are relatively infrequent, with frequencies of 9.4% and 12%, respectively, compared to the frequencies of losses. Additionally, slight gains are detected in V2a in the wild type (Figure 4.17).

In the context of *elt-1*, a pronounced difference between the anterior and posterior halves of the organism is evident. Specifically, H1p, H2, V1a, V1p, V2a, and occasionally V2p exhibit distinct behavior compared to V3a, V3p, V4a, V5, V6a, and V6p. Notably, H0 and T display the highest resilience among the cells. This observed pattern bears a striking resemblance to the expression pattern observed at L2 in *elt-1* expression in the wild type, *elt-1(ku491)*, and *egl-18(ga97)*. The impact of *egl-18* RNAi treatment is notably milder compared to *elt-1* RNAi. At 25% *egl-18* RNAi treatment, we observe no significant change in expression, with a mean of 16 similar to the wild type. However, the first discernible effect surfaces at 50% *egl-18* RNAi, where we observe an average of approximately 12 seam cells. Interestingly, this average remains consistent even at 100% *egl-18* RNAi.

In observing seam cell specific changes in *egl-18* RNAi, we once again observe that H0 and T are the least sensitive of the seam cells. We will consider a highly significant loss of seam cell one that has a p-value <0.01 relative to T. At 25% *egl-18* RNAi we note no significant losses in any of the cells. The only significant difference between T and other cells is the increased seam cell duplications in V2a, a phenomenon also observed in the wild type. However, at 50% *egl-18* RNAi, we once again observe a significant difference in sensitivity

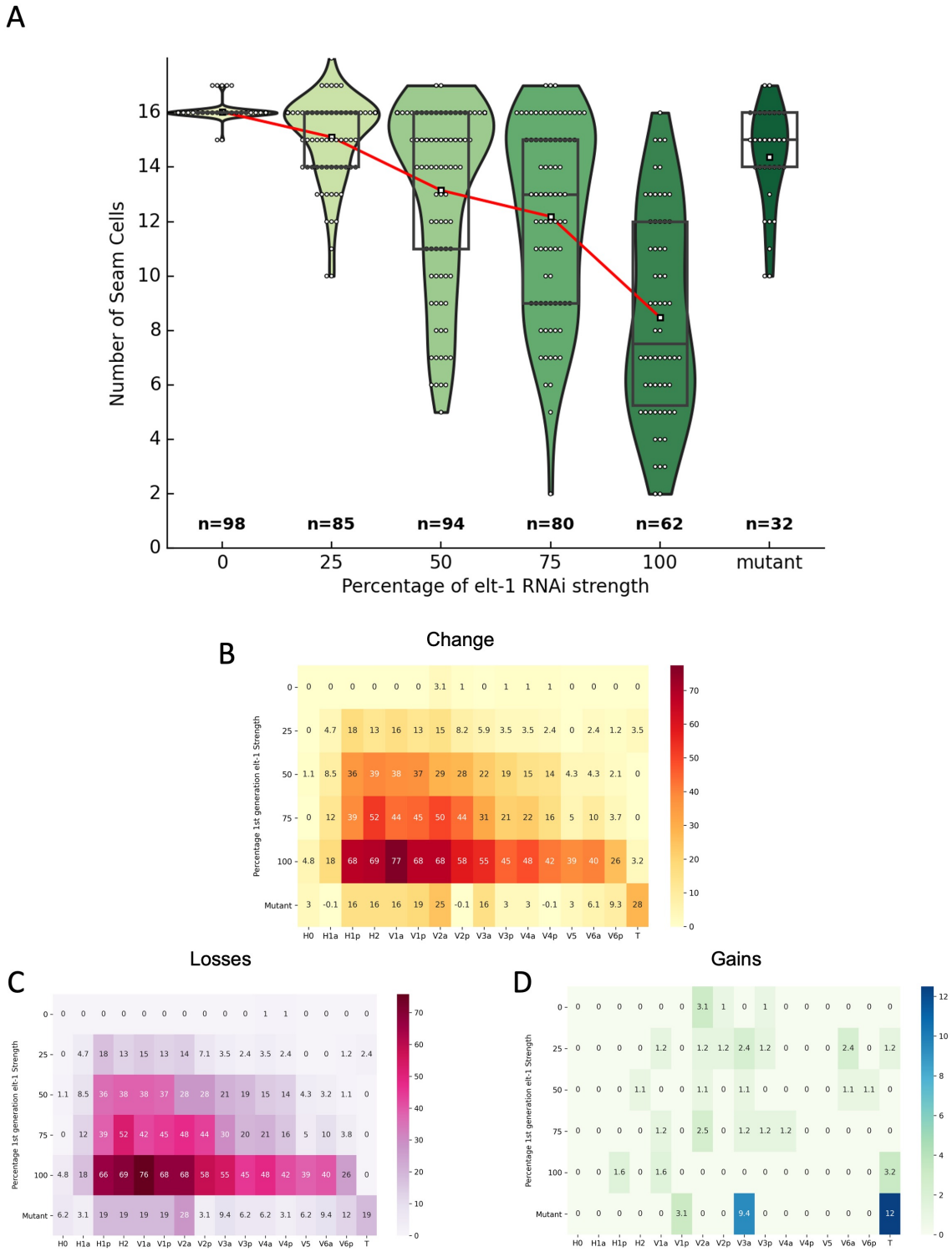


Figure 4.17: **Changes in seam cell sensitivities caused by different dilutions of *elt-1* RNAi bacteria: A)**The number of cells in each animal plotted against the strength of RNAi treatment and compared to the *elt-1(ku491)* mutant. The red line indicates the progressive decrease of the mean counts. The white spots indicate counts of individual animals. The box plot indicates the interquartile range of the seam cell counts. **B)** We show changes in individual cells. The n number for each treatment is the same as the one in the graph above as these quantifications were obtained from the same experiment. This graph is further separated into losses and gains **C)** and **D)** respectively.

between the anterior V cells (V1 and V2) and the posterior V cells (V3, V4, and V6). H1p and especially H2 behave very similarly to the anterior V cells, with H1p exhibiting even more severe losses than the anterior V cells. The losses in H0, V5, and T are not significantly different from those in the posterior V cells. H1a shows no significant difference to any cell except H1p; as such, this cell would be described as moderately sensitive. The most sensitive group of cells at 50% *egl-18* RNAi is H1p, H2, V1a, V1p, V2a, V2p, whereas the group most resistant to *egl-18* loss includes H0, V3a, V3p, V4a, V4p, V5, V6a, V6p, and T. At 75% RNAi the sensitivity caused an even bigger difference between H1p and the remaining cells with two thirds of the animals losing H1p. All cells between H2 and V2p are lost in at least 35% of the cases and H1a is shortly behind with losses in 29% of the animals while the remaining seam cells are never lost in more than 13% of the cases. Nevertheless despite this stark anterior-posterior differences at lower concentrations of *egl-18* RNAi, at 100% *egl-18* RNAi the cells become more equal in their loss patterns. H1p becomes lost in only 49% of the cases, whereas V3a, V3p, V4a, V4p, V6a, V6p, and T disappear in more than 13% of the animals and the only cells with less than 13% of losses are H0 (12%), V5 (7.8%) and T which remains unperturbed by the loss of *egl-18*.

Finally, upon comparing the average seam cell count between the different RNAi concentrations and the mutant, we observe that the mean is consistent for 75% *egl-18*, 100% *egl-18*, and *egl-18(ga97)*. We know that *egl-18(ga97)* is a strong loss-of-function mutant thus its phenotype can be regarded as equivalent to the complete absence of *egl-18*. RNAi administered through feeding is likely to have incomplete silencing within a population due to different feeding habits of different animals. As such both 75% and 100% RNAi should have higher levels of *egl-18* than the mutant. Nevertheless despite the increase in *egl-18* downregulation in the *egl-18* hypomorph we note that H1p has a decreasing frequency of losses from 75% to 100% RNAi and finally to the *egl-18(ga97)* mutant. This indicates that at lower levels of *egl-18*, while the number of cells does not decrease further (Figure 4.18A) they are lost more evenly (Figure 4.18 C). In the *egl-18(ga97)* mutant H1p is lost with approximately the same frequency as H2, V1a and V2a. We note the highest rate of losses in V6a, and we observe that V5 and V6p have very similar loss patterns to the anterior V cells and the head

cells, with the exception of H0. Additionally, we observe that in all the V cells, the anterior cell is lost with a higher frequency than the posterior, although the difference is not statistically significant(Figure 7.3).

Given the consistent pattern observed in *elt-1* RNAi, and the theoretical expectation that *elt-1*, as a master regulator, should affect all seam cells equally, we aimed to investigate if disrupting key pathways involved in maintaining positional information could alter this observed pattern. Thus, we targeted key *Hox* and *Wnt* genes.

To address *Hox* genes we focused on the posterior *Hox* gene *egl-5* through *egl-5(n945)*, and the anterior/midbody *Hox* gene *lin-39* through *lin-39(n1880)*. We also used the triple *Hox* gene mutant *lin-39(n1760);mab-5(e1239);egl-5(n945)* to look for overall signalling disruption. In order to target *Wnt* we focused on the *bar-1(ga80)* β-catenin mutant, the *egl-20(hu105)* *Wnt* ligand mutant and the frizzled mutant *lin-17(n671)*. We aimed to gain insight into whether either of these pathways could reverse the anterior sensitivity observed in *elt-1* RNAi down-regulation. We utilized *elt-1* RNAi at a 50% concentration to observe a strong positional pattern without compromising our ability to differentiate between cells if the phenotype proved to be compounded.

We initially evaluated our animals based on seam cell number (Figure 4.19). While all six untreated mutants exhibited slight differences in seam cell number compared to the wild type, only the *lin-39(n1760)* mutant showed significant change in distribution of counts. Upon treatment with the 50% *elt-1* RNAi, in the *Hox* mutants we observed a weak increase in seam cell number in *egl-5(n945)*. We also noticed a slight increase in seam cell counts in the triple *Hox* mutant. We observed no change *lin-39(n1760)* compared to the wild type when treated with 50% *elt-1* RNAi. When the *Wnt* mutants were treated with 50% *elt-1* RNAi we observed a significant increase in seam cell defects in all of them except *bar-1*. The *bar-1(ga60)* mutant was highly variable however the resulting change in seam cell defects was not significant. Nevertheless both *egl-20(hu105)* and *lin-17(n671)* both showed strong seam cell losses and with fewer animals having higher seam cell counts. This strongly suggests that Wnt, rather than Hox, may play a more significant role in maintaining the posterior seam cells.

Most of the *Hox* genes showed a similar pattern to what we observed in the wild type when

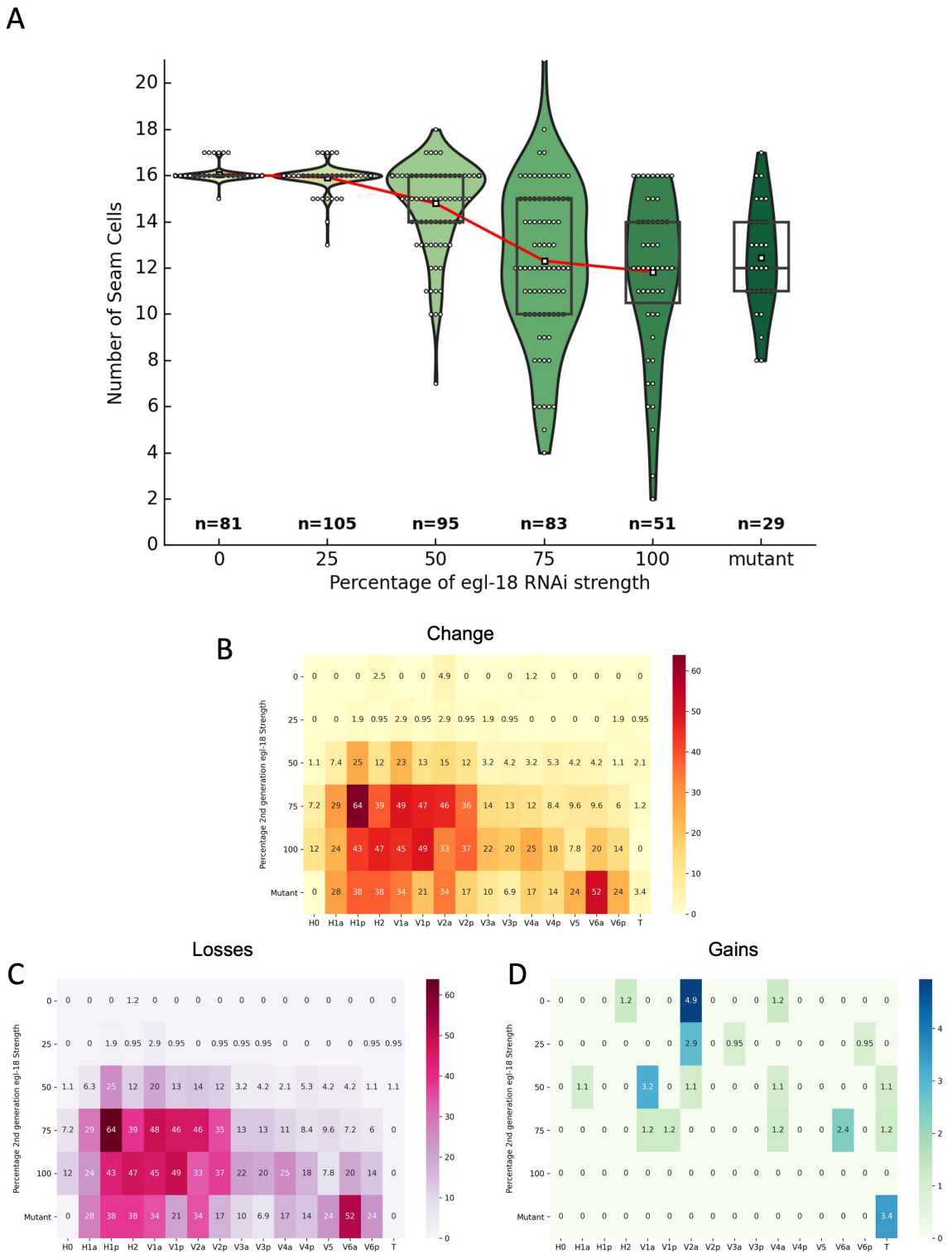


Figure 4.18: **Changes in seam cell sensitivities caused by different dilutions of *egl-18* RNAi bacteria: A)**The number of cells in each animal plotted against the strength of RNAi treatment and compared to the *egl-18(ga97)* mutant. The red line indicates the progressive decrease of the mean counts. The white spots indicate counts of individual animals. The box plot indicates the interquartile range of the seam cell counts. **B)** We show changes in individual cells. The n number for each treatment is the same as the one in the graph above as these quantifications were obtained from the same experiment. This graph is further separated into losses and gains **C)** and **D)** respectively.

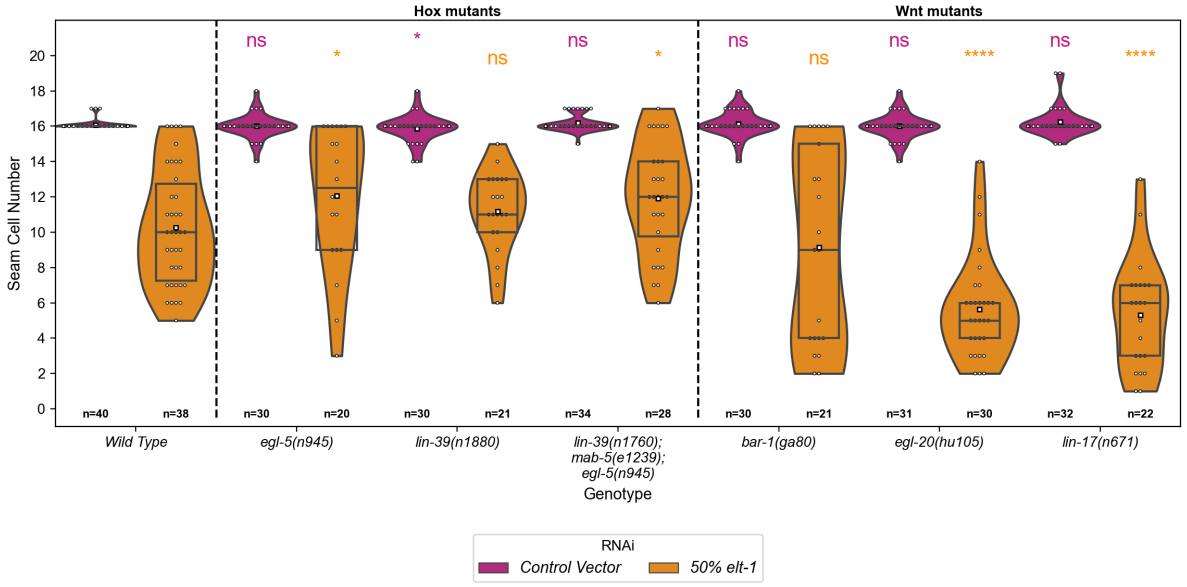


Figure 4.19: **Seam cell counts in *Hox* and *Wnt* mutants upon treatment with 50% *elt-1*:** We plot the number of cells in each animal when treated with control vector RNAi and 50% *elt-1* RNAi, in seven different strains. The strains are divided into wild type, *Hox* mutants *(egl-5(n945), lin-39(n1880) and lin-39(n1760);mab-5(e1239);egl-5(n945))* and *Wnt* mutants *(bar-1(ga80), egl-20(hu105) and lin-17(n671))*. The white spots indicate counts of individual animals. The box plot indicates the interquartile range of the seam cell counts. The significance relative to the wild type is denoted with * following the method discussed in the method section, with the colours matching the treatment in the legend.

treated with 50% *elt-1* RNAi (Figure 4.19). However, we did note slight differences. In the *egl-5* mutant, the only cell showing any change from the wild type 50% *elt-1* treatment is H2, which appears to have a slight decrease in the number of losses (p-values in supplementary figure 7.4). In the *lin-39* mutant, we again did not observe a particularly strong effect; however, we noted that V3a, V3p, and V4a appeared to decrease in sensitivity. In the triple *Hox* mutant, we observed this decrease in sensitivity of V3 but not in V4a. However, similar to *egl-5(n945)*, we noted an increase in H2 sensitivity. Nevertheless in none of the *Hox* mutants do we see any decrease in H1p losses, and while the mid body cells to regain some robustness this doesn't fully counterbalance the anterior/posterior inequality in losses that we see in the 50% *elt-1* treatment.

bar-1(ga80) did not show a significant change in seam cell number most likely due to a large increase in variability between animals. Nevertheless we did observe a significant increase in V6 sensitivity. In the other two *Wnt* mutants, where an increase in overall seam cell losses was observed, we noted that these losses were concentrated in the posterior of

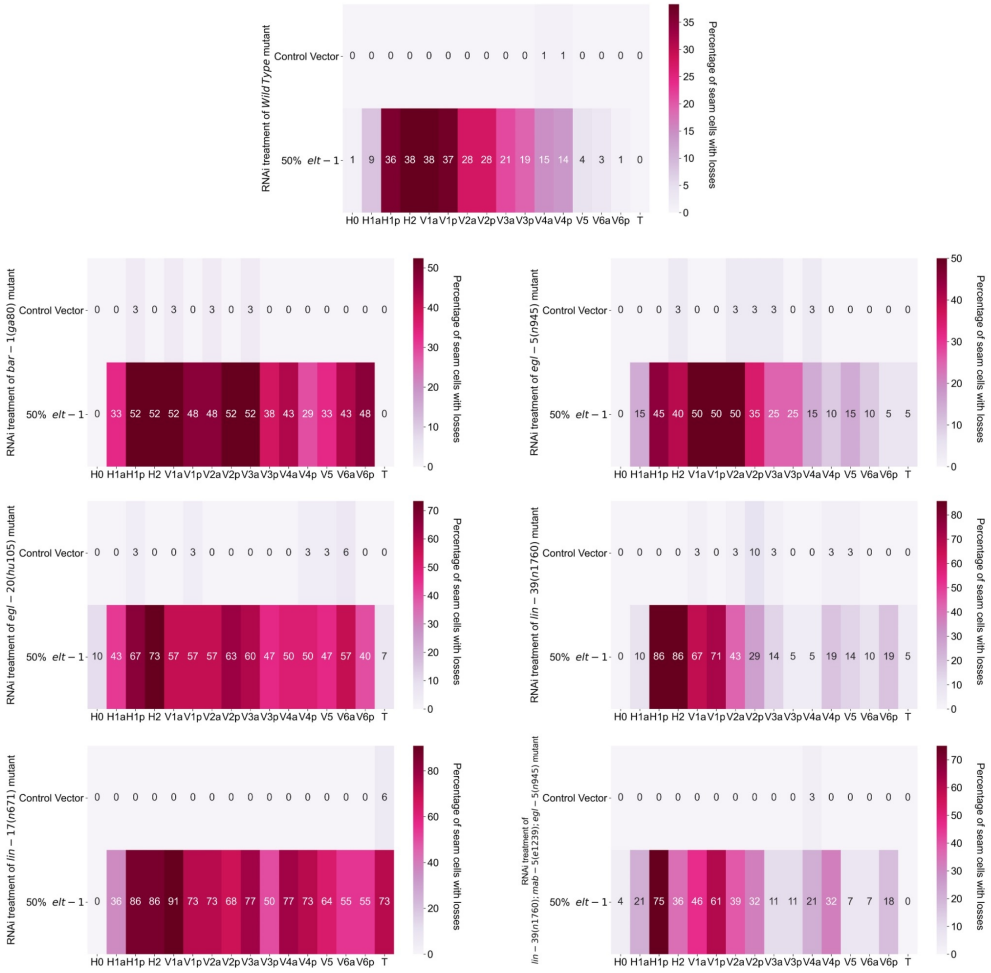


Figure 4.20: **The *elt-1*-specific seam cell sensitivities are disrupted in *Wnt* but not *Hox* mutants:** Percentages of animals with each seam cell lost and gained in wild type, *Hox* mutants *(egl-5(n945), lin-39(n1880) and lin-39(n1760);mab-5(e1239);egl-5(n945))* and *Wnt* mutants *(bar-1(ga80), egl-20(hu105) and lin-17(n671))*. The percentage is as a factor of the n values shown in Figure 4.19 as the data is a part of the same experiment. This graph showcases losses in red, with the intensity of the colour showing the relative frequency of the change.

the animal. In the *egl-20(hu105)* mutant, significant increases in sensitivity occurred in V4p, V5, V6a, V6p, as well as H1a. Meanwhile, in the *lin-17(n671)* mutant, we noted increases in sensitivity in the entire posterior from V4a to T, as well as V3a and H1p. Nevertheless, this posterior-specific increase in losses in both *lin-17* and *egl-20* mutants was paired with a general increase in seam cell loss. Therefore, we propose *bar-1* as the leading candidate for disrupting the *elt-1*-specific seam cell loss pattern.

4.3 Discussion

4.3.1 The layout of the core seam cell GRN changes over time

Overall, both time and position significantly impact the arrangement of the core gene regulatory network (GRN) underlying seam cell patterning. Interestingly many of the core connections change over time. We find that at 10 hours ELT-1 only regulates *egl-18* and even then only as a repressor. This is a novel perspective on *elt-1* that is largely portrayed as an activator in Gorrepati et al. (2013); Thompson et al. (2016) and in Chapter 3. However, the two roles of this gene are not mutually exclusive and the entire GRN may involve different regulatory mechanisms at different time points.

One element that is consistent between 10 and 17 hours, and across all time points is that CEH-16 represses *egl-18*. This again is opposed to the portrayal of CEH-16 as an activator in Cassata et al. (2005) in embryonic regulation of *egl-18*. Again the two opposing roles of CEH-16 are not mutually exclusive but show the different roles that these proteins can have in different contexts.

The concentration of core seam cell genes appears to increase before the L2 symmetric division. Specifically in the *ceh-16(bp323)* mutant, known for losing symmetric divisions (Huang et al., 2009), there is a decrease in expression of all three core genes before this division. After the L2 division we observe both a severe decrease in core gene expression under all conditions and a conversion of most of the interactions within the network to repressive ones (Figure 4.21).

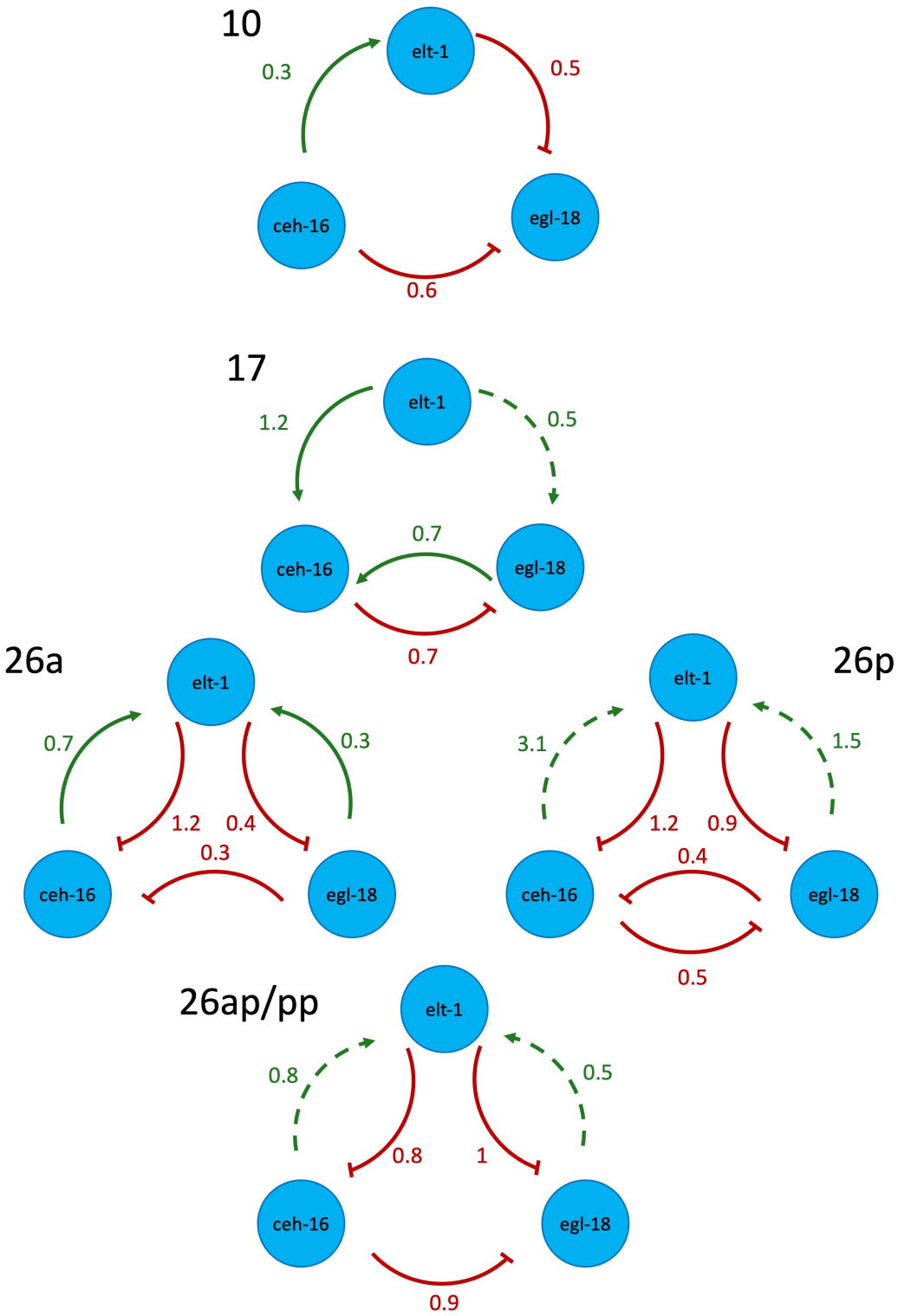


Figure 4.21: **A diagram illustrating quantitative GRNs for the core seam cell genes in early larval stages:** The five main networks of the regulation of V1-V4, at 10 hours, 17 hours, 26 hours symmetric (separated into anterior(a), and posterior(p) symmetric daughter) and 26 hours asymmetric where we only focus on the non-differentiating cells (ap and pp). The colour and ending of the arrow indicate the polarity of the interaction (red blunt-ended arrow for repression, and green arrow for activation) and the number refers to the strength of the interaction predicted by the MRA.

The core seam cell genes start acquring repressive roles at the end of L2, this could imply that over the course of development the seam cells decrease their expression of the core genes. This allows them to terminally differentiate at the end of L4. Our data indicates that the L2 symmetric division requires the highest concentration of core genes.

The peak of core gene could be because there needs to be a sufficient accumulation of the core gene mRNA in the nucleus to allow both daughters to maintain a seam cell fate. We observe that low concentration of the three core genes, particularly *elt-1* is a symptom of differentiating cells. This goes against the concept of a steady state as it shows that the core seam cell genes are highly dynamic in early larval development. We believe that while *elt-1* and *egl-18* can be considered to be at steady state in the 17 hour time-point, *ceh-16* may experience too many transcription rate changes in early larval development for there to be an easily identifiable steady state. Furthermore all three of the genes appear to exprience highly dynamic transcription patters as is shown by the low expression of *elt-1* in V5a. Given that the cell would have only divided at most in the previous hour and levels of *elt-1* are already significantly lower than in the posterior cell, either the mRNA is asymmetrically segregated during the division, or the degradation rate of the mRNA is extremely high. While our assumption of a 17 hour steady state is fair based on the biological behaviour of the system, reflecting on the expression patterns we believe that *ceh-16* may be more dynamic than originally expected at that time point. Additionally all three of the mutants appear to affect the dynamics of the system, with *elt-1* experiencing premature peak expression in all three mutants, and *ceh-16* delayed peak expression. In order to study the dynamics of this system, one would most likely have to rely on live imaging of gene expression. At present this is not achievable through smFISH. However, one can utilise novel imaging techniques like the MS2 repeats fused to MCP/GFP (Tutucci et al., 2018; Kinney et al., 2023; Lee et al., 2019; Tocchini and Mango, 2024), paired with microfluidics-based imaging techniques (Keil et al., 2017). This could shed a more detailed light on the dynamics of the system and provide the possibility of using modelling approaches that incorporate time as a variable.

An important thing to consider is the likelihood of these interactions being direct or indirect. While a direct repressive function of *ceh-16* is highly likely due to its homology with the known

repressor *engrailed*, *elt-1* is portrayed more frequently as an activator. As such a potential mechanism for the increase in repressive action from ELT-1 to the other two genes in the network, could involve the activation of a repressive gene. A gene that whose upstream region contains ELT-1 chip-seq peaks (Luo et al., 2020; ENCODE Project Consortium, 2012), and that has been shown to repress *egl-18* (Katsanos and Barkoulas, 2022), is *nhr-25*. While we have no evidence of it downregulating *ceh-16* its role as an intermediary between ELT-1 and *egl-18* makes it a promising candidate. Another candidate for an intermediary for L2 based repression of the core seam cell genes could be *unc-37* whose homolog *groucho* is described frequently as a repressor (Cinnamon and Paroush, 2008; Chambers et al., 2017), particularly of the homolog of *ceh-16*, *engrailed* (Celis and Ruiz-Gómez, 1995). UNC-37 has also been described as a repressor (Winnier et al., 1999).

The ultimate proof of concept of a stem cell regulatory network would be the ability to revert to a more pluripotent state. In the case of *C. elegans* this could require reprogramming L4 seam cells back to an early L1 state as shown in Okita et al. (2007). Based on our results this would require raising the levels of all three genes, in particular *elt-1*.

Here we also note similarities between the *C. elegans* stem cell gene network and the human stem cell gene network. Both appear to require extremely high levels of the core factors, boosted through internal feedback loops in order to maintain a high level of potency. In both cases the progression of the network to a more differentiation-prone state requires the downregulation of one specific factor. In *C. elegans* this would be *elt-1* whereas in humans it would be NANOG.

Another important consideration that we should account for is the limitation of using MRA. While the MRA has the capacity of predicting a steady state response, it makes no estimation of long term behaviours. For example the activatory/repressive interaction between *egl-18* and *ceh-16* could lead to oscillatory behaviour. This could only be proven through live imaging as singular snapshots wouldnt be able to capture any oscillations. There have been new methods that have been developed with the purpose of applying an MRA to a dynamic system (Sarmah et al., 2022), however these require a regular time course dataset following a singular cell over time. Our system followed the subsequent daughters of cells and as such

the gene behaviour wasn't based around changes within a single cell.

4.3.2 The frequencies of seam cell losses and gains are correlated to core gene expression patterns

We observed important positional patterns in both expression levels and seam cell losses and gains. We use our smFISH data that the seam cells have non-identical expression patterns, and over the course of development the clusters of similarity change. At 10 hours the head (H0, H1, H2), midbody (V1, V2, V3, V4) and tail (V5, V6, T) distinction remains as the leading pattern in similarities and differences between cells. The V5a cell which goes on to differentiate into the hypodermis has a highly distinct expression pattern, marked by low *elt-1* and *ceh-16* with an average of $<$5 counts in both.

In the symmetric division time point the distinction between the head, midbody and tail, becomes slightly less clear. The pattern remains to some extent in *elt-1* expression, however is lost in both *egl-18* and *ceh-16* expression. Once again differentiating cells (H2a and V5a) have low *elt-1* and *ceh-16*, an expression signature shared by the anterior symmetric daughters (V1a, V2a, V3a, V4a, V6a) in the *ceh-16(bp323)* mutant. *egl-18* expression does nor appear to be as clearly related to cells prone to loss of seam cell fate. In the asymmetric division, we observe differences between V1, V2 and V3, V4 with V3 and V4 having higher *elt-1* and *ceh-16* expression.

Our findings correlate the absence of *elt-1* and *ceh-16* to seam cell losses and thus reinforce the portrayal of *elt-1* as an essential seam cell gene (Koh and Rothman, 2001; Smith et al., 2005; Brabin et al., 2011; Page et al., 1997a), and strengthen the role of *ceh-16* in seam cell maintenance (Huang et al., 2009; Cassata et al., 2005; Hughes et al., 2013) and not only in seam cell symmetry, as cells with low expression of *ceh-16* are more prone to being lost. Additionally our data from early larval stages indicates that even early differences in expression can make a cell more or less prone to being lost. V3 is the cell least susceptible to losses and has elevated *elt-1* expression at all time points. While this difference is not significant relative to all cells it shows significant differences to the cells with least *elt-1*

expression. We again highlight *elt-1* as a key gene in seam cells using RNAi as we show that the loss of seam cell number is directly proportional to the loss of *elt-1*. However, in order to establish the causality between the absence of specific genes in individual seam cells and gene expression one would have to develop core gene knock-outs specific to individual seam cells.

We also suggest that *egl-18*-based seam cell loss plateaus at an average of 12.5 and that knockdowns of *egl-18* stronger than what we see with the 75% RNAi lose cells more evenly, rather than losing more cells. This could be caused by EGL-18 opposing either a *Hox* gene function or a *Wnt* gradient. Given that *egl-18* has previously been shown to be regulated by *Wnt* (Gorrepati et al., 2013) and that *egl-18* and *ceh-16* appear to also regulate *Wnt* effectors, as we show in Chapter 5, and given that *Wnt* has been shown to have an anterior-posterior gradient, it is very likely that this effect is due to an interaction with *Wnt* (Harterink et al., 2011; Pani and Goldstein, 2018).

4.3.3 The positional effect of *elt-1* is modulated by Wnt

Given the clear seam cell pattern caused by decreasing amounts of *elt-1* we decided to evaluate whether this could be caused by *Hox* or Wnt. Our aim was to find a phenotype that would make H1p, H2, V1a, V1p, and V2a more robust or that would cause more significant losses in the posterior seam cells (V3a, V3p, V4a, V4p, V5, V6, T). We tested three *Hox* gene mutants and three *Wnt* effector mutants. For *Hox* we focused on an anterior/midbody gene *(lin-39)*, a posterior gene *(egl-5)* and a triple *Hox* mutant *(lin-39, mab-5 and egl-5)* (Chisholm and Hsiao, 2012). For *Wnt* we looked at the β-catenin *bar-1*, the *Wnt* ligand *egl-20* and the frizzled homolog *lin-17*.

These pathways are closely interlinked as various *Wnt* genes have been shown to regulate *Hox*, for example EGL-20 regulating *mab-5* (Harris et al., 1996), and *Wnt* component PRY-1 regulating *lin-39, mab-5* and *egl-5* (Maloof et al., 1999). However, when we treat the *Hox* mutants with 50% *elt-1* RNAi we note only a weak increase in seam cell losses and positional sensitivities. Meanwhile in the *Wnt* mutants we note severe changes in both number of seam cells and frequency of individual seam cells lost. In *bar-1(ga80)*, we note no significant

change in seam cell number, however we do observe an increase in V6 losses. The other two *Wnt* mutants experience both a loss in seam cell number and a sensitisation of the posterior of the animal to *elt-1* RNAi treatment. This indicates that *Wnt* is essential for the *elt-1* uneven sensitivity. This could be caused by an increased severity of the *elt-1* phenotype as more concentrated *elt-1* RNAi treatment does also lose the posterior cells. As such *bar-1* makes a slightly better candidate than the other two genes as its absence causes changes in frequency of losses of the posterior cells without a significant effect in seam cell number.

Wnt signalling is also involved in stem cell regulation in humans. It is one of the signals that are linked to pluripotency maintenance in stem cells (Chickarmane et al., 2006). While in humans cells with high levels of *Wnt* are linked to a failure of self-renewal (Tao et al., 2015), our findings show that the posterior cells which tend to be exposed to higher levels of *Wnt* signalling are less prone to erroneous differentiation (Harterink et al., 2011). This indicates that like in humans high levels *Wnt* are correlated to the maintenance of the stem cell fate.

Chapter 5

Understanding the robust seam cell number phenotype caused by the combined loss of *ceh-16* and *egl-18*

5.1 Introduction

C. elegans carries out all of its cell divisions in a highly robust manner. Robustness was first defined by Waddington (1957, 1961) and is used to describe phenotypes that remain invariant in the face of internal (genetic) or external (environmental) perturbation (Félix and Wagner, 2008). Most studies have focused on robustness in single cell organisms like yeast (Levy and Siegal, 2008) due to ease of use. However, recently this became a topic of interest in the *C. elegans* community. Due to the accuracy and precision with which the cells in *C. elegans* divide, changes in robustness of phenotype are easily quantifiable through variability in terminal seam cell number (Katsanos et al., 2017; Hintze et al., 2021; Koneru et al., 2021b). Studies of robustness in *C. elegans* have closely linked disruptions of robust phenotypes to the *Wnt* pathway. However, most seam cell gene mutants that were studied through the lens of robustness showed significant gains in variability. We introduce the mutant *ceh-16(bp323);egl-18(ga97)* which, despite producing a consistently erroneous seam cell number, shows minimal loss of robustness.

In the wild type during, L2 V1-4,V6 and H1 seam cells undergo two divisions, one of which is a self-renewal division, while the other results in one seam cell and one differentiating cell. Remarkably, this symmetric division occurs despite an asymmetry in *Wnt* signaling between the resulting cells. *Wnt* signalling is essential for selecting which cell differentiates and which becomes a seam cell. The gene *pop-1* is one of the key players in the maintenance of asymmetry in divisions and has to be modulated using RNT-1/BRO-1 in order to allow symmetric divisions to occur (van der Horst et al., 2019).

Despite all the knowledge regarding symmetry and asymmetry in seam cell divisions, there are very few studies linking the genes involved in seam cell divisions to the core seam cell genes. The phenotype we identified in Chapter 3 in the *ceh-16(bp323);egl-18(ga97)* double mutant was characterised by a regular 11 seam cell phenotype. This appeared to be a more extreme version of the phenotype resulting from *ceh-16(bp323)* as described in Huang et al. (2009). In order to study this mutant we generated a seam cell specific knockout of *ceh-16* using the Loxp Cre system. This mutant will be referred to henceforth as *ceh-16*$_{LoxP}$.

The objectives of this chapter include evaluating the robustness of the *ceh-16(bp323);egl-18(ga97)* double mutant, analyzing its role in regulating symmetric division, and assessing if the phenotype solely stems from the loss of *ceh-16* through the comparison of its behaviour to the *ceh-16*$_{LoxP}$ mutant. We investigate the resistance of this phenotype to internal (genetic) and external (environmental) perturbations, identifying specific developmental stages and locations where cell count errors occur, and elucidating the underlying mechanisms behind the consistent loss of seam cells.

5.2 Results

5.2.1 Generating and characterising a seam cell specific *ceh-16* mutant

To discern the specific phenotype of the double mutant from that potentially caused solely by an increased loss of *ceh-16*, we needed to address the challenge posed by the lethality associated with complete loss of *ceh-16*. To overcome this obstacle, we generated a seam

cell-specific *ceh-16* knockout utilizing the Cre/LoxP system. Originally adapted from the P1 bacteriophage for eukaryotic use Sauer (1987); Sauer and Henderson (1988) the Cre/LoxP system functions through the recognition of specific short repeat sequences flanking a target genetic sequence, known as LoxP sites, by a recombinase, resulting in excision of the target sequence. This process leaves behind one copy of the target sequence as a 'scar'. In *C. elegans*, the first Cre-LoxP system was employed for Pn.P cell-specific deletion of *apr-1*, aiming to circumvent embryonic lethality associated with such deletions (Hoier et al., 2000). Conditional knockouts using a Cre/LoxP system can be achieved by expressing the Cre with different promoters, such as heat-shock or tissue specific promoters. For our *ceh-16*$_{LoxP}$ construct we relied on a seam cell-specific promoter in order to avoid *ceh-16* based defects in other tissues. The newly generated strain had an average seam cell count of $\approx$ 10.4 (n = 40) with a maximum of 13 seam cells and a minimum of 8. This range was slightly wider than what we observed with the *ceh-16(bp323);egl-18(ga97)* mutant (maximum of 12, minimum of 10), however the double mutant mean of 10.9 (n = 31) was highly similar (Figure 5.1 and 5.2). One of the preliminary experiments we did was to test the *ceh-16*$_{LoxP}$ on *egl-18* RNAi and compare it to the double mutant. The difference between the two strains was significant. While *ceh-16(bp323);egl-18(ga97)* showed no response to the RNAi (n = 28) the *ceh-16*$_{LoxP}$ average seam cell count dropped to an average of 5.5 (n = 33) (Figure 5.3). This is an indication that the phenotype observed in *ceh-16(bp323);egl-18(ga97)* is not solely due to a strong loss of *ceh-16*, but that a complete loss of *ceh-16* may cause a similar phenotype. Additionally the seam cell specific *ceh-16* knock-out showed a similar seam cell number to what we observed in the double mutant, however there was more variability in the seam cell specific *ceh-16* knock-out compared to the double mutant. In 19% of the *ceh-16*$_{LoxP}$ animals the H1 cells on each lateral side were missing before the L2 symmetric division. Additionally in 78% of the *ceh-16*$_{LoxP}$ animals one of the H1 cells was missing before the L2 division. Only in 3% did both the lateral sides of the animal have a H1 cell (n = 32).

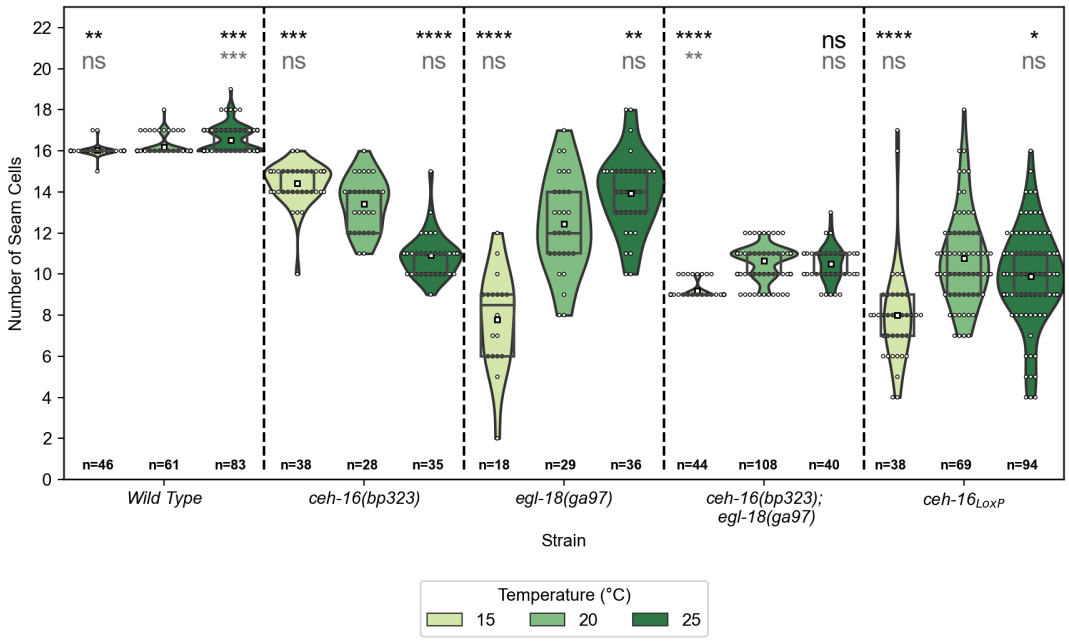


Figure 5.1: **Impact of temperature on *ceh-16(bp323);egl-18(ga97)* and *ceh-16*$_{LoxP}$:** A graph showing the seam cell number of wild type, *ceh-16(bp323)*, *egl-18(ga97)*, *ceh-16(bp323);egl-18(ga97)* and *ceh-16*$_{LoxP}$ animals grown at 15, 20 and 25℃. The significance is shown as described in methods. The p-values shown at the top in black were calculated comparing the 20℃ animals to the 15℃ or 25℃ using Mann Whitney U test to compare the means. The p-values shown below in grey are also a comparison with the 15℃ or 25℃ to the 20℃ animals but these are compared using Levene's test, that compares the variances of the animals.

5.2.2 The *ceh-16(bp232);egl-18(ga97)* is more phenotypically resistant to both environmental and genetic perturbations than *ceh-16*$_{LoxP}$

Our initial experiments aimed to determine whether the double mutant and the *ceh-16*$_{LoxP}$ exhibited resistance to environmental perturbations, such as changes in temperature. The standard growth temperature of all *C. elegans* strains in this study was 20℃. Wild type animals observed at the L4 stage have been shown to exhibit an increase in seam cell duplications, particularly in V6a at 25℃ (Hintze et al., 2020). Our data shows that the decrease in temperature to 15℃ results in fewer erroneous seam cell duplications relative to the 20℃, albeit occasionally resulting in seam cell loss (Figure 5.1), thus while there is a significant decrease in seam cell number from $\approx$ 16.1 to 16.0 there is no substantial change in variability at the lower temperature. It is important to consider that animals kept at 15℃ are slow growing and are not as healthy as those kept at 20℃. As is in line with the literature we

observe an increase in cell duplications at 25℃ which leads also to an increase in variability. The *ceh-16(bp323)* strain has been shown to be a temperature sensitive mutant (Huang et al., 2009) and has been shown to lose the symmetric division in the majority of the cells at 25℃. Indeed, we observed a higher occurrence of seam cell losses at 25℃ and a reduction in seam cell losses at 15℃. However, the variability of seam cell counts for the mutant remained consistent across all three temperature conditions.

The *egl-18(ga97)* mutant has previously shown a reduction in errors at higher temperatures (Koneru et al., 2021a). We confirm this finding as we see a statistically significant increase in seam cell number at 25℃. However, we also observe a significant decrease in seam cell number at 15℃ which has not previously been reported (Figure 5.1). However, despite the changes in seam cell number in both *ceh-16(bp323)* and *egl-18(ga97)* we do not observe any changes in variability at either 15 or 25℃ that would be statistically significant when tested using a Levene's test (Figure 5.1).

Notably in the *ceh-16(bp323);egl-18(ga97)* we do note a decrease in variability at 15℃. This is accompanied by a loss in seam cell number to ≈ 9. We observe no significant change in either seam cell number or variability of seam cell counts at 25℃ (Figure 5.1). Given the inherent temperature sensitivity of the *ceh-16(bp323)* mutant we also evaluated the temperature sensitivity of a *ceh-16*$_{LoxP}$ strain (Figure 5.1). As with the double mutant and the *egl-18(ga97)* mutant we note a decrease in seam cell number at 15℃ compared to 20℃, however the seam cell counts between 20℃ and 25℃ remain comparable with a slight decrease in seam cell counts at the higher temperature (Figure 5.1). The variability for all conditions of the *ceh-16*$_{LoxP}$ strain remains high. Given our findings in Chapter 3, in order to dissect the temperature effect on the two new mutants we focused more closely on the pattern of the cell losses.

Throughout our seam cell counts we noticed that that temperature affects cells unequally (Figure 5.2). Consistent with previous findings, wild type animals exhibit the highest sensitivity to errors in the V6a cell (Hintze et al., 2020) which we also observed (Figure 5.2). This sensitivity manifests through seam cell duplications. Additionally, we observed gains in V1a that were statistically more frequent than in other cells, with p-values lower than 0.03 when

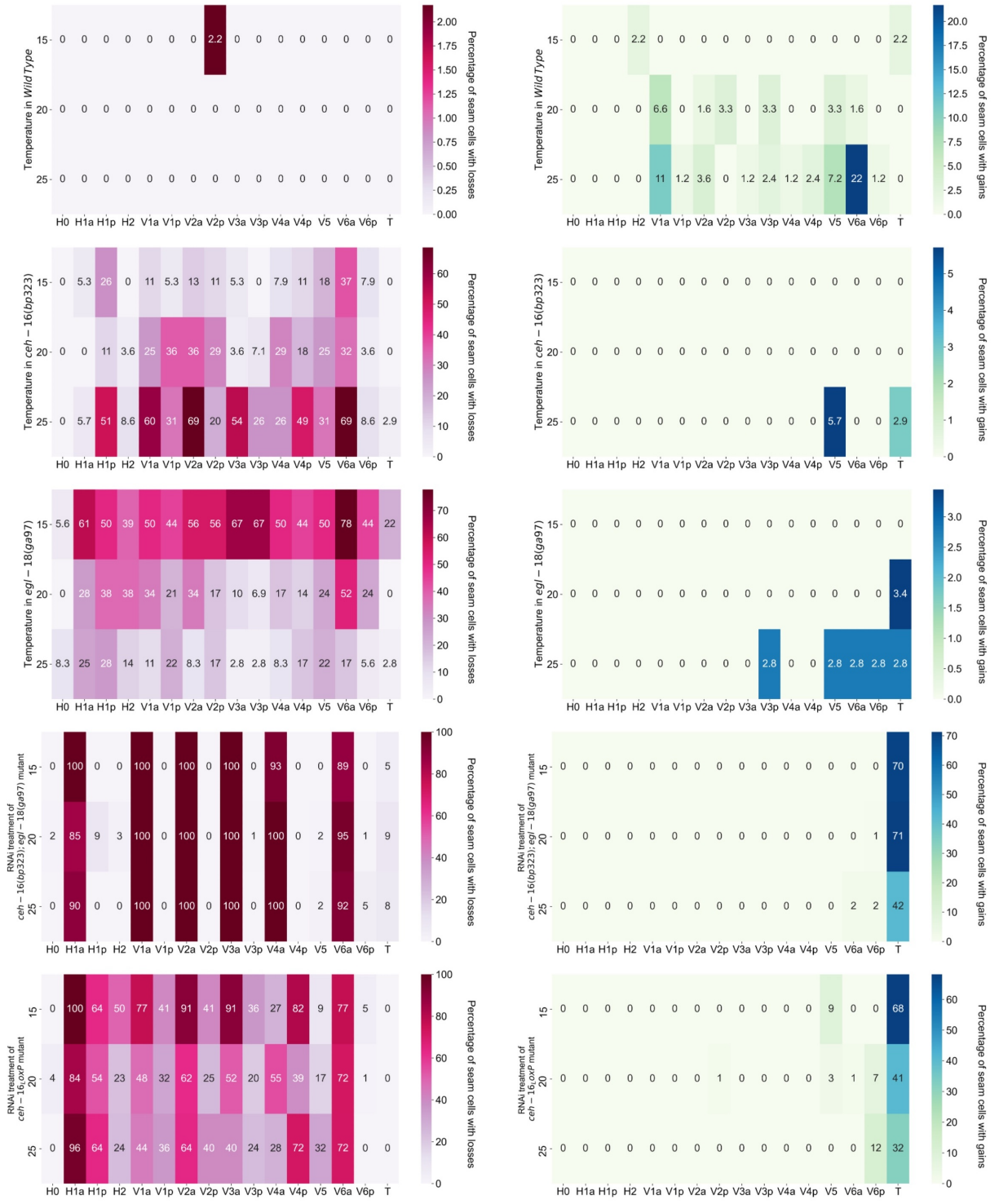


Figure 5.2: **Both the *ceh-16(bp323);egl-18(ga97)* mutant and the *ceh-16*$_{LoxP}$ show frequent losses in anterior cells, and gains in tail cells:** The percentage of animals with specific cells lost (left) or gained (right) at different temperature conditions. The intensity of the colour corresponds to how high the percentage is relative to all conditions in the same mutant. The strains analysed are (In order from top to bottom): *Wild Type*, *ceh-16(bp323)*, *egl-18(ga97)*, *ceh-16(bp323);egl-18(ga97)* and *ceh-16*$_{LoxP}$. The n numbers are the same as shown in Figure 5.1 and the p-values for the inter-cell comparisons are shown in Figure 7.7

.

compared to all cells except V2a, V5, and V6a (Figure S7.7). In the *ceh-16(bp323)* mutant at 20℃, we noted high similarity among all V cells, except for V3 cells, which appeared to be more resistant to this mutation. At 25℃, we observed a higher sensitivity in the anterior V cells compared to the posterior V cells, particularly in V6, where V6a showed the highest significance of losses relative to the other cells. V4a was a notable exception, as it lost fewer seam cells than its posterior cell. Likewise, H1a is more resistant at 25℃ in the *ceh-16(bp323)* mutant than H1p. In fact, H1p, H2, and V6p showed no significant differences in their losses compared to the notoriously stable H0. The frequent loss of anterior V cells in the *ceh-16(bp323)* mutant is characteristic of a loss of symmetry, consistent with the findings of Huang et al. (2009). We also observed gains in V5 and T cells in the *ceh-16(bp323)* mutant at 25℃, attributed to a failure of neuronal differentiation, aligning with the observations of Huang et al. (2009). Finally, in the *ceh-16(bp323)* mutant at 15℃, the majority of its losses were attributed to H1p and V6a. Despite fewer losses occurring at this temperature, it is interesting that these cells maintained equal or even higher sensitivity than expected at higher temperatures (Figure 5.2). Other than those two cells the majority of the seam cells maintain a frequency of losses at $\approx$ 10%. H0, H2, V3p and T cells were never recorded to be lost and are significantly more robust than the other cells in *ceh-16(bp323)* at 15℃.

In the *egl-18(ga97)* mutant at 20℃, we observe frequent losses in the majority of the V cells (Figure 5.2). However, the two V3 cells are once again notable exceptions. The highest frequency of losses was observed in V6a, V2a, V1a, H2, H1p and H1a. H0 and T exhibit negligible losses although the T cell features a significant number of gains. The 15℃ treatment of *egl-18(ga97)* greatly homogenises the distribution of losses, as there are no significant differences between any of the V cells except V6a, which is lost slightly more frequently than the others. H1 and H2 cells also follow the same pattern. H0 and the T cell remain relatively unperturbed. At 25℃ we see 10-20% of losses in most cells. Once again the V3 cells exhibit fewer losses. The most spatially striking pattern in *egl-18(ga97)* at 25℃ is in the seam cell duplications. These occur exclusively in V3p, V5, V6a, V6p and T showing a strong posterior sensitivity to the loss of *egl-18* at high temperatures.

The *ceh-16(bp323);egl-18(ga97)* phenotype is highly robust. Not only are the cells lost

consistently the same ones but also there is very little variation of the loss pattern even at unfavourable temperatures (Figure 5.2). The only variability we see is whether or not H1a, V4a and V6a are lost, which happened in $>85\%$ of the animals, and whether or not a T or V6 cell are gained. The T cell gain happened in $\approx 70\%$ of the cases at 15℃ and 20℃, but dropped to 42$\%$ at 25℃. We also noted a $<10\%$ occurrence of V6 gains and H0, H1p, H2, V3p, V5 losses. These losses amount to 24$\%$ of the animals at 20℃ but drops to 15$\%$ and 5$\%$ in 25℃, and 15℃ incubated animals respectively. V1a, V2a and V3a were lost in every single animal that was counted, and we only note one occurrence of any of their posterior cells being lost (V3p at 20℃) (Figure 5.2). This result is strong evidence that *ceh-16(bp323);egl-18(ga97)* is resistant to environmental perturbation and that the seam cell losses and gains in it follow a highly specific pattern which is virtually unchanged under environmental perturbation.

Finally, when comparing the loss pattern of the *ceh-16(bp323);egl-18(ga97)* to that of *ceh-16*$_{LoxP}$, we note significant differences. Firstly, the seam cell loss pattern is a lot less binary in the latter, the differences between anterior and posterior cells are less pronounced. H1p is lost with a very high frequency in *ceh-16*$_{LoxP}$, while it is never lost in the double mutant. Despite the strain being highly resistant to external perturbation, there is a lot more variability as to which cells are lost.

To assess the robustness of both the *ceh-16(bp323);egl-18(ga97)* and the *ceh-16*$_{LoxP}$, against genetic perturbations, we carried out an RNAi screen of the core seam cell genes, described in Chapter 3 as *elt-1*, *egl-18* and *ceh-16*. We found that RNAi treatment of these three genes has a significant effect on both the wild type and mutant worms with losses in seam cell number and an increases in variability between worms (Figure 5.3). In order to more completely also target *egl-18* we wanted to evaluate the effects of removing *elt-6* as well given *elt-6* and *egl-18* have been shown to be at least partially redundant (Gorrepati et al., 2013).

In the *ceh-16(bp323)* mutant we see a seam cell loss even in absence of RNAi with the average seam cell count being $\approx$ 13.5 (Figure 5.3). *elt-6* RNAi treatment causes no further defects. However, both *egl-18* and *ceh-16* RNAi treatment do cause a significant decrease in seam cell number. The *elt-1* RNAi treatment of *ceh-16(bp323)* results in a highly significant

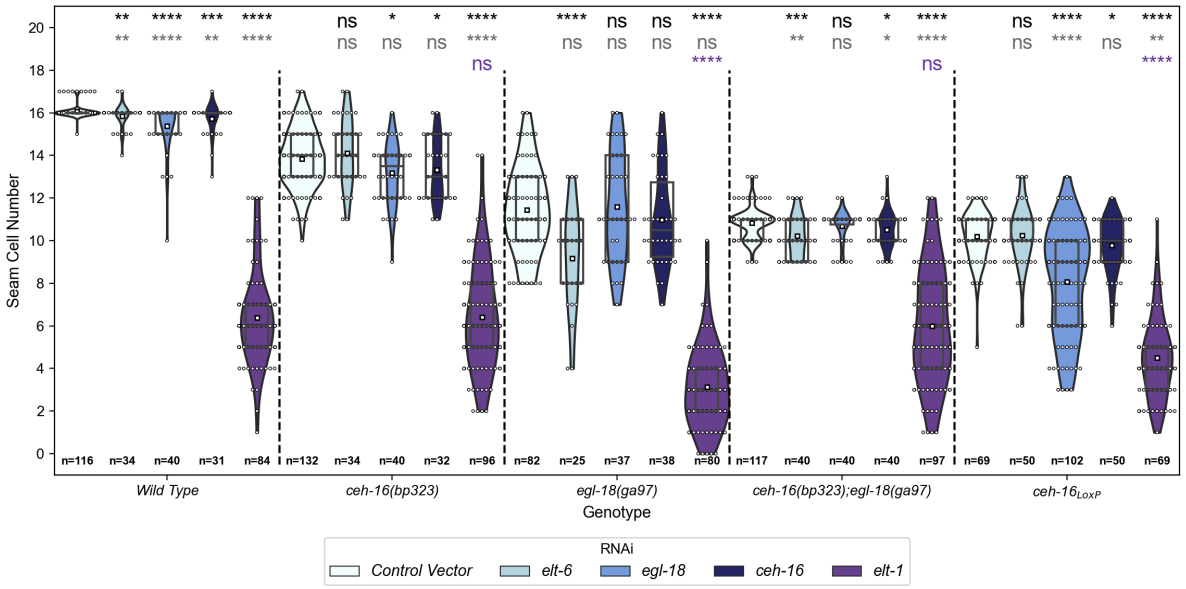


Figure 5.3: **The core gene RNAi on *egl-18* and *ceh-16* mutants showcases the robustness of *ceh-16(bp323);egl-18(ga97)* in the face of genetic perturbations:** A graph showing the seam cell number of wild type, *ceh-16(bp323)*, *egl-18(ga97)*, *ceh-16(bp323);egl-18(ga97)* and *ceh-16*$_{LoxP}$ animals grown on *elt-6*, *egl-18*, *ceh-16* and *elt-1* RNAi. The p-values are denoted as discussed in methods. The p-values shown at the top in black were calculated comparing the RNAi treated animals to the same strain grown on control vector RNAi using Mann Whitney U test to compare the means. The p-values shown below in grey are also a comparison with the RNAi treatments to the control vector animals but these are compared using Levene's test, which compares the variances of the animals. Finally, in purple we show the significance of the changes in the *elt-1* RNAi treatment compared to the same treatment of the wild type using Mann Whitney U test.

loss of seam cells, however the loss is equivalent to the same treatment of the wild type indicating that *ceh-16* is likely downstream of *elt-1*. *elt-1* is also the only RNAi treatment which results in a in increase in variability compared to the control *ceh-16(bp323)* (Figure 5.3).

Untreated *egl-18(ga97)* mutants exhibit a lower seam cell count compared to *ceh-16(bp323)*. The average seam cell count is $\approx$ 12 with the control vector (Figure 5.3). Consistent with findings by Gorrepati et al. (2013), treatment of *egl-18(ga97)* with *elt-6* RNAi leads to a significant reduction in seam cell number, resulting in an average count of $\approx$ 10 (Figure 5.3). We did not observe any significant effects from either the *egl-18* or *ceh-16* RNAi treatments. This outcome aligns with expectations, considering that *egl-18(ga97)* represents a complete loss-of-function mutant and the effect of *ceh-16* RNAi is minimal. However, it is worth noting that treatment with *elt-1* RNAi resulted in a more severe loss of seam cells compared to other mutants, with an average count of $\approx$ 3.5, representing just over half the count observed in any other strain except *ceh-16*$_{LoxP}$. However, due to the high variability of *egl-18(ga97)* itself, the changes in seam cell number caused by the various RNAi treatments do not increase this mutant's variability (Figure 5.3).

After analyzing the parental strains, our next step is to assess the behavior of the double mutant (Figure 5.3). The *ceh-16(bp323);egl-18(ga97)* double mutant exhibits a seam cell number range between 9 and 13, with only 5% of the animals displaying either 9 or 13 seam cells, while the remaining 95% fall within the range of 10 to 12 cells. Notably, 75% of the animals have 11 seam cells (Figure 5.3).

elt-6 RNAi treatment results in a modest reduction of seam cells relative to the untreated double mutant, with the average decreasing from 11 to 10. However, there is a large change in the distribution of the counts with animals having 9, 10, and 11 seam cells are observed with approximately equal frequency. The difference appears to be largely centred around whether the T cell is duplicated or not or lost in animals with 9 seam cells (Figure 5.3). The animals at L4 appear to have the same cells that they hatched with, however while in the control RNAi we frequently see a T cell duplication, when treated with *elt-6* RNAi the T cell is frequently lost. The difference between the double mutant treated with the control vector, and

the *elt-6* RNAi is significant, and is most likely due to the change in frequency of T cell losses and duplications (Figure 5.3).

Treating the *ceh-16(bp323);egl-18(ga97)* mutant with *egl-18* RNAi yielded no significant effect, consistent with observations from the single *egl-18(ga97)* mutant (Figure 5.3). As the latter is a complete loss-of-function mutant, it is already incapable of further loss of *egl-18* function. Conversely, *ceh-16* RNAi treatment induced a minor effect, resulting in a slight increase in the frequency of animals with 10 cells. This subtle change in seam cell number is similar to what we observe in the *ceh-16(bp323)* mutant. The final treatment on the *ceh-16(bp323);egl-18(ga97)* mutant was *elt-1* RNAi, resulting in an average seam cell count of $\approx$ 6.5. While this count matches that of the wild type and *ceh-16(bp323)*, it significantly exceeds that of its parental strain, *egl-18(ga97)*. This suggests that the absence of *ceh-16* partially mitigates the impact of *elt-1* on *C. elegans*' seam cells (Figure 5.3).

The last strain we perturbed using RNAi was *ceh-16*$_{LoxP}$ (Figure 5.3). As previously discussed in this chapter, this strain is a seam cell-specific knockout of *ceh-16*, exhibiting an average seam cell count similar to that of the double mutant (Figure 5.3). Like the double mutant it appears resistant to temperature-based perturbation. It does however, have a different seam cell loss pattern to the *ceh-16(bp323);egl-18(ga97)* mutant. Therefore, we proceeded to compare the response of these two mutants to RNAi-based perturbations targeting essential seam cell genes (Figure 5.3). In the case of *ceh-16*$_{LoxP}$, contrary to the double mutant, we see no difference between the untreated mutant and the *elt-6* RNAi treatment. However, consistent with our initial analysis of the mutant, it demonstrated heightened sensitivity to *egl-18* RNAi treatment (Figure 5.3). The response of *ceh-16*$_{LoxP}$ to the loss of *egl-18* was highly variable with L4 larvae exhibiting seam cell counts ranging from as high as 13 to as low as 3. From the untreated mutant the seam cell count drops from $\approx$ 10 to $\approx$ 8. We also noted a minor effect of *ceh-16* RNAi on the LoxP mutant, possibly attributable to the general epidermal effects of *ceh-16*, given that the RNAi is not seam cell-specific. Finally, the *elt-1* RNAi treatment resulted in an average seam cell count of $\approx$ 4 which is slightly higher than that of the *egl-18(ga97)* mutant ($\approx$ 3.5), however also slightly lower than that of the other three strains which is $\approx$ 6.5 (Figure 5.3).

Overall, this experiment highlights the robustness of the *ceh-16(bp323);egl-18(ga97)* strain to genetic perturbations compared to its parental strains. Even with significant perturbations its seam cell count range doesn't change outside of the 9-13 range that the untreated strain presents. The only exception is *elt-1* RNAi treatment which shows a significant loss in all strains. However, in the *ceh-16(bp323);egl-18(ga97)* strain, this loss is less pronounced compared to its parental strain *egl-18(ga97)*. Furthermore, the *elt-1* RNAi treatment elicits a similar phenotype in wild type, *ceh-16(bp323)* and *ceh-16(bp323);egl-18(ga97)* strains, suggesting that *elt-1* may act upstream of *ceh-16*.

The comparison between the behavior of double mutant animals and the *ceh-16*$_{LoxP}$ strain reveals significant differences. Unlike the double mutant, the *ceh-16*$_{LoxP}$ strain does not respond to *elt-6* RNAi and demonstrates a severe defect upon treatment with *egl-18* RNAi. Additionally the complete loss of *ceh-16* makes the seam cells more sensitive to the *elt-1* RNAi treatment. This finding suggests that the observed effect in the double mutant cannot be solely attributed to the decrease in *ceh-16* activation by EGL-18 and thus the two strains are different. In order to delve into understanding a robust phenotype one should focus on *ceh-16(bp323);egl-18(ga97)*.

5.2.3 The double mutant converts the L2 symmetric division to an asymmetric division

To dissect the double mutant and the *ceh-16*$_{LoxP}$ phenotypes we had to understand both the temporal and mechanistic aspects of the errors. Positionally, as illustrated in Figure 5.1, we observed that the only seam cells lost in *ceh-16(bp323);egl-18(ga97)* mutant were the anterior V cells and H1a, and while *ceh-16*$_{LoxP}$ followed a similar pattern, this was less stringent. We observed the two mutants at all four of the larval stages we found that at every stage we could see divisions. We tracked divisions by observing whether we could see instances of sets of two divided cells. This would occur just before the anterior cell would acquire the hypodermal fate. In the L2 this meant that we expected to see some animals with sets of two cells (the artefact of the symmetric division) and, for a brief period of time, sets of four cells,

referred to here as quadruplets. We'd expect quadruplets to form in every asymmetric division following the L2 symmetric. The animals hatched with all 10 cells which divided normally until L2. However, despite normal division activity, we never observed the formation of quadruplets at the L2 stage when observing 26 animals. This observation strongly indicated a shift from symmetric to asymmetric division.

Our next hypothesis was that the normally symmetric division became one of the asymmetric divisions where the anterior cell fused with the hyp7 hypodermis. This fusion would require the expression of the *eff-1* fusogen. We confirmed that the divisions we saw at L2 were asymmetric using *eff-1* smFISH, which showed extremely high expression levels of *eff-1* in the anterior cell in all of the seam cells meant to divide symmetrically in in the double mutant (Figure 5.4). We observed that 26 out of 32 animals showed most of the V cells following the same eff-1 expression pattern as the double mutant. In *ceh-16*$_{LoxP}$ we saw *eff-1* expression also, albeit less frequently, in the posterior cells (1 out of 32 animals had two cells with only posterior expression and 5 out of 32 animals had at least one V cell expressing eff-1 in both cells) (Figure 5.4). In the wild type *eff-1* is absent during the symmetric division (Figure 5.4). Additionally having followed both the *ceh-16(bp323);egl-18(ga97)* and the *ceh-16*$_{LoxP}$ mutant strains over the course of development, alongside the wild type, we noticed no discrepancies in developmental timing. This is despite the fact that *ceh-16(bp323)* has a slight developmental delay.

In order to evaluate the response of *ceh-16(bp323);egl-18(ga97)* and *ceh-16*$_{LoxP}$ to a perturbation of key pathways involved in symmetry, we decided to carry out an RNAi screen on the double mutant, the seam cell specific *ceh-16* mutant, and their parental strains, targeting several genes implicated in symmetry regulation. These include *rnt-1* (*pop-1* repressor thought modulate *Wnt* signalling during the symmetric division), *lit-1* (a kinase involved in *pop-1* export in the anterior cells of an asymmetric division), *efl-3* (a potential regulator of *lit-1* and a gene known to cause increases in seam cell number), *eff-1* (the hyp7 fusogen, expressed in the anterior cells of both *ceh-16(bp323);egl-18(ga97)* and *ceh-16*$_{LoxP}$ during the first L2 division), *pop-1* (one of the main components of the *Wnt* signalling pathway involved in conveying asymmetry in seam cells, its absence causes seam cell hyperplasia), and *apr-1*

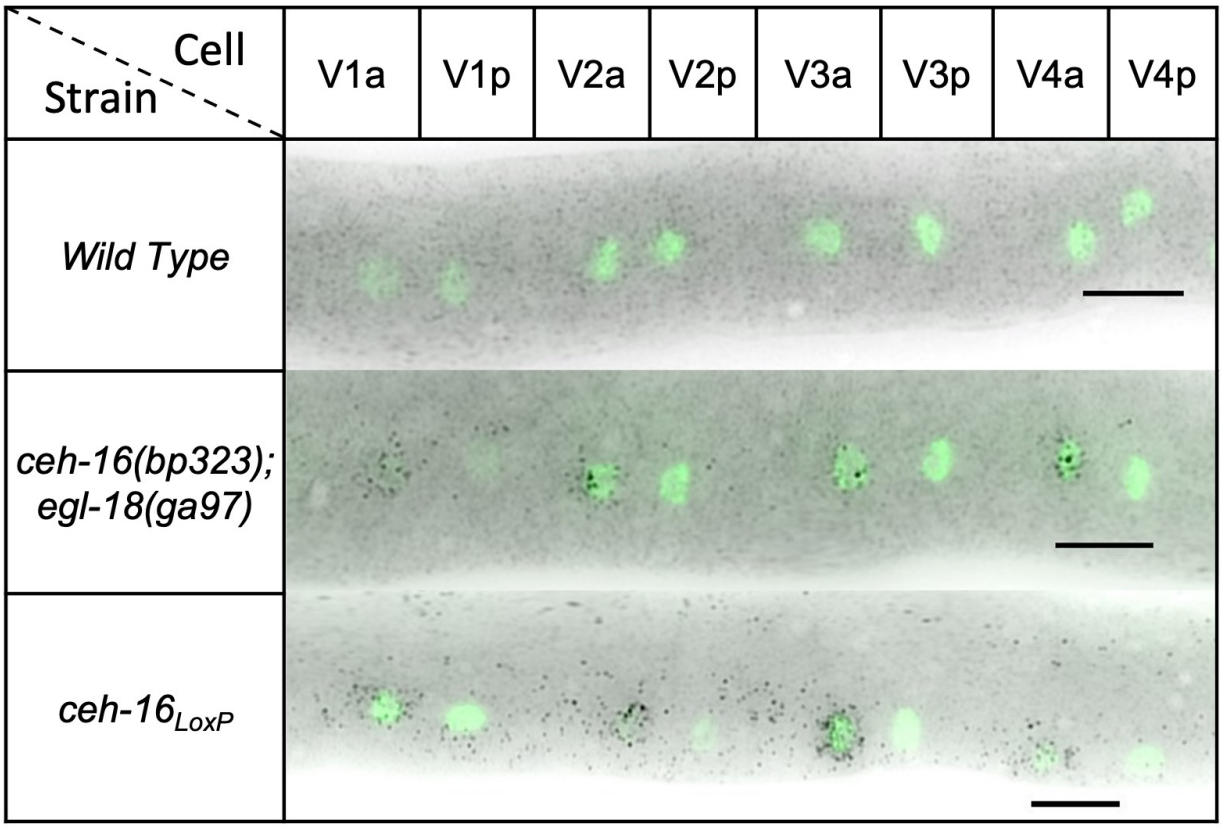


Figure 5.4: ***eff-1* distribuition in the double mutant and *ceh-16*$_{LoxP}$ 26 hours after bleaching:** smFISH image showing *eff-1* expression. *eff-1* mRNA is bound to a cy5 fluorescent probe shown as black spots. We show differences in *eff-1* expression between the wild type, *ceh-16(bp323);egl-18(ga97)*, and *ceh-16*$_{LoxP}$ strains at the symmetric division time point in L2. The figure shows the difference in the expression of this gene between the anterior and the posterior cell for each of V1-V4 shown using SCM::GFP. The scale bar in the bottom right corner indicates a distance of 10μm

(mutations of which cause the conversion of the L4 asymmetric division to a symmetric division; this gene's mutant has been shown to rescue *ceh-16(bp323)* (Huang et al., 2009)). Among these, *rnt-1* showed no significant effect in any strain. This could be as a result of the RNAi treatment being particularly weak rather than a resistance to the loss of *rnt-1*.

lit-1 RNAi has been previously shown to promote differentiation and cause a decrease in seam cell number (Harandi and Ambros, 2015). Our own *lit-1* RNAi treatment resulted in significant seam cell count reductions in all strains except the *ceh-16(bp323);egl-18(ga97)* mutant. We also observed cases of lethality with some of the mutants (Figure 5.6), a phenotype that has not been previously observed. The most severe lethality occurs in the *egl-18(ga97)* and the *ceh-16*$_{LoxP}$ strains when treated with *lit-1* RNAi. *lit-1* treatment of these two mutants also causes the greatest reductions in seam cell counts. *lit-1* RNAi treatment of *ceh-16(bp323)* has a significant decrease in seam cell number but doesn't result in lethality. The *ceh-16(bp323);egl-18(ga97)* mutation does become lethal in 28% of the cases when paired with *lit-1* RNAi. However, there is no significant change in seam cell counts, other than an increase in variability (Figures 5.5 and 5.6).

efl-3 RNAi treatment has previously been shown to cause an increase in seam cell number (Katsanos and Barkoulas, 2022). We confirm this as we observed a significant increase in both variability and seam cell number to an average of ≈ 17 in the wild type. However, in all the *efl-3* RNAi treated mutants except the *ceh-16(bp323);egl-18(ga97)* double mutant we observe a decrease in seam cell number. In particular, *egl-18(ga97)* experienced drastic seam cell loss, with some nematodes having only 3 seam cells, accompanied by substantial variability increase. In *ceh-16(bp323);egl-18(ga97)*, *efl-3* RNAi treatment increases the variability of the seam cell counts but doesn't affect the mean. This once again shows that *ceh-16(bp323);egl-18(ga97)* and *ceh-16*$_{LoxP}$ have different responses, as the latter experiences a loss of seam cell number without a change in variability when treated with *efl-3*.

The *eff-1* treatment, like the *rnt-1* treatment, shows limited effect. Specifically, it induces a significant increase in seam cell number in the wild type, and increases the variability of the counts between worms in the *ceh-16(bp323);egl-18(ga97)*, as illustrated in Figure 5.5. However, apart from these effects, no other significant alterations were observed (Figure 5.5).

The *pop-1* RNAi treatment induces a significant increase in seam cell number in the wild type, with seam cell counts averaging $\approx$ 32 seam cells. However, this effect is severely impaired in all of the mutants with *ceh-16(bp323)* mutants only having $\approx$ 23 seam cells per lateral side. The *egl-18(ga97)* mutant even showing a decrease in seam cell number when treated with *pop-1* RNAi compared to the untreated *egl-18(ga97)*. In *ceh-16(bp323);egl-18(ga97)* there is a gain of seam cells relative to the untreated strain, with an the average increasing to $\approx$ 13 seam cells (Figure 5.5). Likewise, the *ceh-16*$_{LoxP}$ strain shows minimal change in the mean seam cell number, increasing only from approximately 11 to 12 with *pop-1* RNAi treatment. However, a significant increase in variability is noted, with seam cell counts ranging from as low as 3 to as high as 17 (Figure 5.5).

With *apr-1* RNAi treatment, a gain in seam cell number is observed, albeit less pronounced in the wild type compared to *pop-1*, with an average of $\approx$ 20 seam cells (Figure 5.5). However, nearly all mutants exhibit an increase in seam cell number in the absence of *apr-1*. In the *ceh-16(bp323)* mutant, this gain is slightly higher than in the wild type, averaging 20.5 seam cells. The *egl-18(ga97)* showed no significant gain in seam cell number when treated with *apr-1* RNAi. However, unlike the *pop-1* treatment of this mutant the average seam cell number did not decrease. Some individual *egl-18(ga97)* animals exhibited an increase in seam cell number beyond that of the untreated mutant, although still lower than the *apr-1* treated wild type, albeit with significantly more variability (Figure 5.5). The most striking phenotype of the *apr-1* RNAi treatment is observed in the *ceh-16(bp323);egl-18(ga97)* double mutant, where seam cell counts surpass not only those of both parental strains subjected to the same RNAi, but also the wild type *apr-1* RNAi response. The *ceh-16*$_{LoxP}$ mutation showed an increase in seam cell number when treated with *apr-1* RNAi, however this was not as large of an increase as we observed in the *ceh-16(bp323);egl-18(ga97)*. Indicating that the complete loss of *ceh-16* and RNAi of *apr-1* isn't enough to generate the 33 cell phenotype we observe in the *apr-1* treatment of the double mutant. This again highlights the differences between the two mutants, despite exhibiting similar phenotypes. It gives a strong indication that the mechanism of the loss of symmetry is different between the two strains.

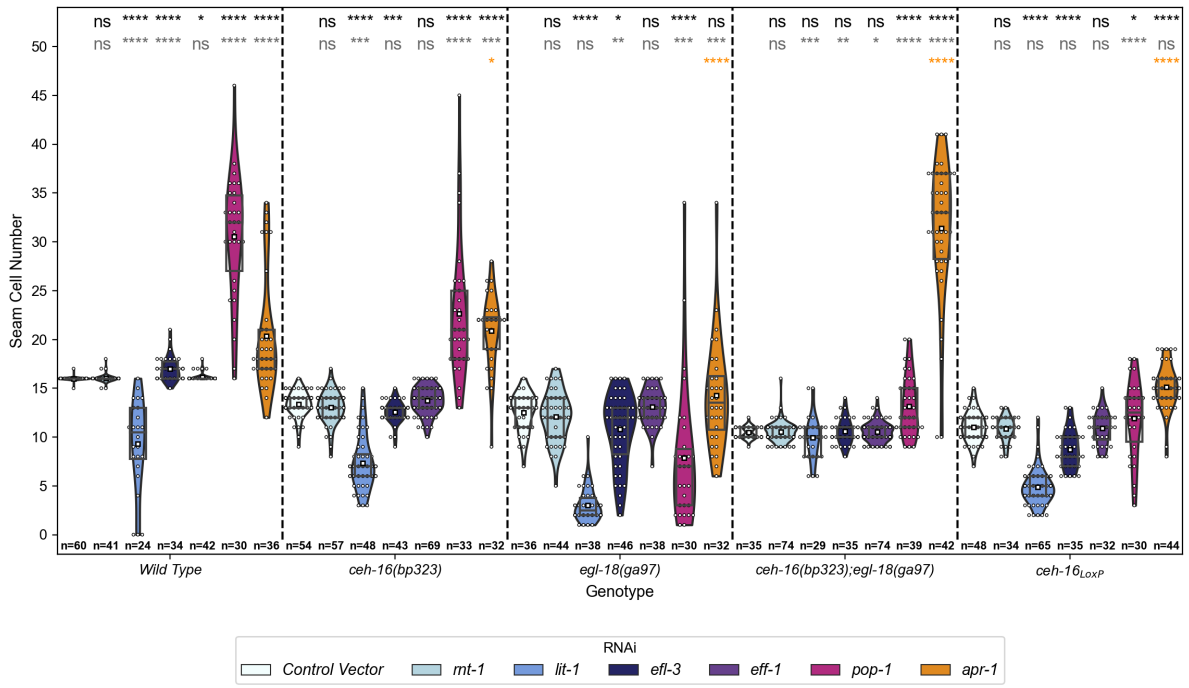


Figure 5.5: **Impact of symmetry regulators on the loss of symmetry in the double mutant and parental strains**: A graph showing the seam cell number of wild type, *ceh-16(bp323)*, *egl-18(ga97)*, *ceh-16(bp323);egl-18(ga97)* and *ceh-16*$_{LoxP}$ animals grown on *efl-3*, *rnt-1*, *lit-1* , *pop-1*, *eff-1* and *apr-1* RNAi. The p-values are denoted as described in methods. The p-values shown at the top in black were calculated comparing the RNAi treated animals to the same strain grown on control vector RNAi using Mann Whitney U test to compare the means. The p-values shown below in grey are also a comparison with the RNAi treatments to the control vector animals but these are compared using Levene's test, which compares the variances of the animals. In orange we show Mann Whitney U comparison of all *apr-1* treatments to the the wild type animals subjected to the same treatment

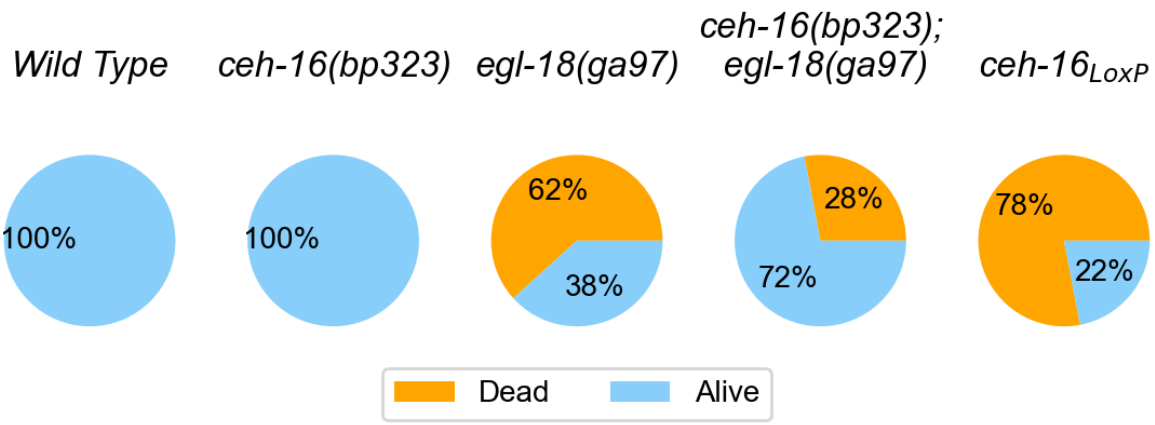


Figure 5.6: **Survival of *lit-1* RNAi**: Pie charts showing the percentage of animals that were alive and mobile at the end of L4 when treated with *lit-1* RNAi. The strains tested were wild type, *ceh-16(bp323)*, *egl-18(ga97)*, *ceh-16(bp323);egl-18(ga97)* and *ceh-16*$_{LoxP}$

In order to evaluate whether the increase in seam cell number was caused by a rescue of the L2 division or simply further symmetric divisions at L3 and L4 we had to look more closely at L2. For an animal to exceed 28 seam cells without engaging any of the cells that don't normally divide symmetrically in a symmetric division, one would require all of the cells carrying out at least 2 symmetric divisions. However, in order for the seam cell number to remain below 52 under the same conditions, all of the cells can divide a maximum of 3 times. That means that in order for more than half the cells to have between 28 and 43 seam cells we'd expect the majority of the cells to divide 2 times symmetrically and no more than 3 of them to divide 3 times. Despite the fact that in this case the RNAi treatment was applied to two generations before quantifying the seam cells, RNAi by feeding is always limited by the fact that the animals need to start eating the bacteria in order for the RNAi to take effect. This in all likelihood excludes the possibility of the L1 division being affected. Given more than half of the *ceh-16(bp323);egl-18(ga97)* mutant animals treated with *apr-1* RNAi showed a phenotype of more than 28 seam cells but less than 43 there are three possible scenarios that can occur during this treatment. First of all the remaining four divisions could have cells randomly dividing symmetrically and asymmetrically. Secondly one could have the L2 symmetric division happening as normal, and the L4 division resulting in two seam cells as occurs in the *apr-1* mutants. The final option is that the first L2 division also remains unaffected, whether this is because of the type of treatment applied or because it is insensitive

to symmetrisation, and the asymmetric L2 is only partially affected, however by the time one reaches L3 and L4 all the divisions are symmetrised. Our observations of the treated mutants however showed no change in the L2 division indicating that the effect of the RNAi was on the L3 and L4 division. Despite all this information we decided to delve deeper in order to further understand the mechanism.

5.2.4 Investigating possible mechanisms of symmetry loss through mutations in both *ceh-16* and *egl-18*

Having established the conversion of symmetric division to asymmetric division in the *ceh-16(bp323);egl-18(ga97)* and *ceh-16*$_{LoxP}$ mutants, it is essential to understand the mechanism underlying this switch. One of the key proteins involved in generating asymmetry between seam cells during division is *pop-1*, prompting us to initially examine changes in its gene expression. Our analysis of *pop-1* focused on two aspects: first, the differences between anterior and posterior cells' expression, and second, the variation in expression levels compared to wild type. We first looked at *pop-1* and *rnt-1* (Figure 5.7). *pop-1* is considered essential for maintaining asymmetry as its high levels of expression are thought to induce differentiation, whereas *rnt-1* is thought to downregulate *pop-1* in order to permit symmetric divisions (van der Horst et al., 2019). As such in the case of our two mutants the expected levels of these two genes would be increased levels of *pop-1* particularly in the anterior cell and lower levels of *rnt-1* to cause a symmetric division to fail. We notice that although the symmetric division results in similar behaviors in both cells later in development, the expression of *pop-1* appears slightly asymmetric in several cells. Specifically, we observe slightly higher expression in the anterior cell in H2, V1 and V2 while in V3, the posterior cell exhibits slightly higher expression where the p-value comparing the anterior and posterior cell was <0.05.

The most significant finding is the disparity in expression levels between wild type and the two mutants. We observe a marked reduction in *pop-1* expression at the time of symmetric division in the *ceh-16(bp323);egl-18(ga97)* mutant (Figure 5.7A). While some cells, such

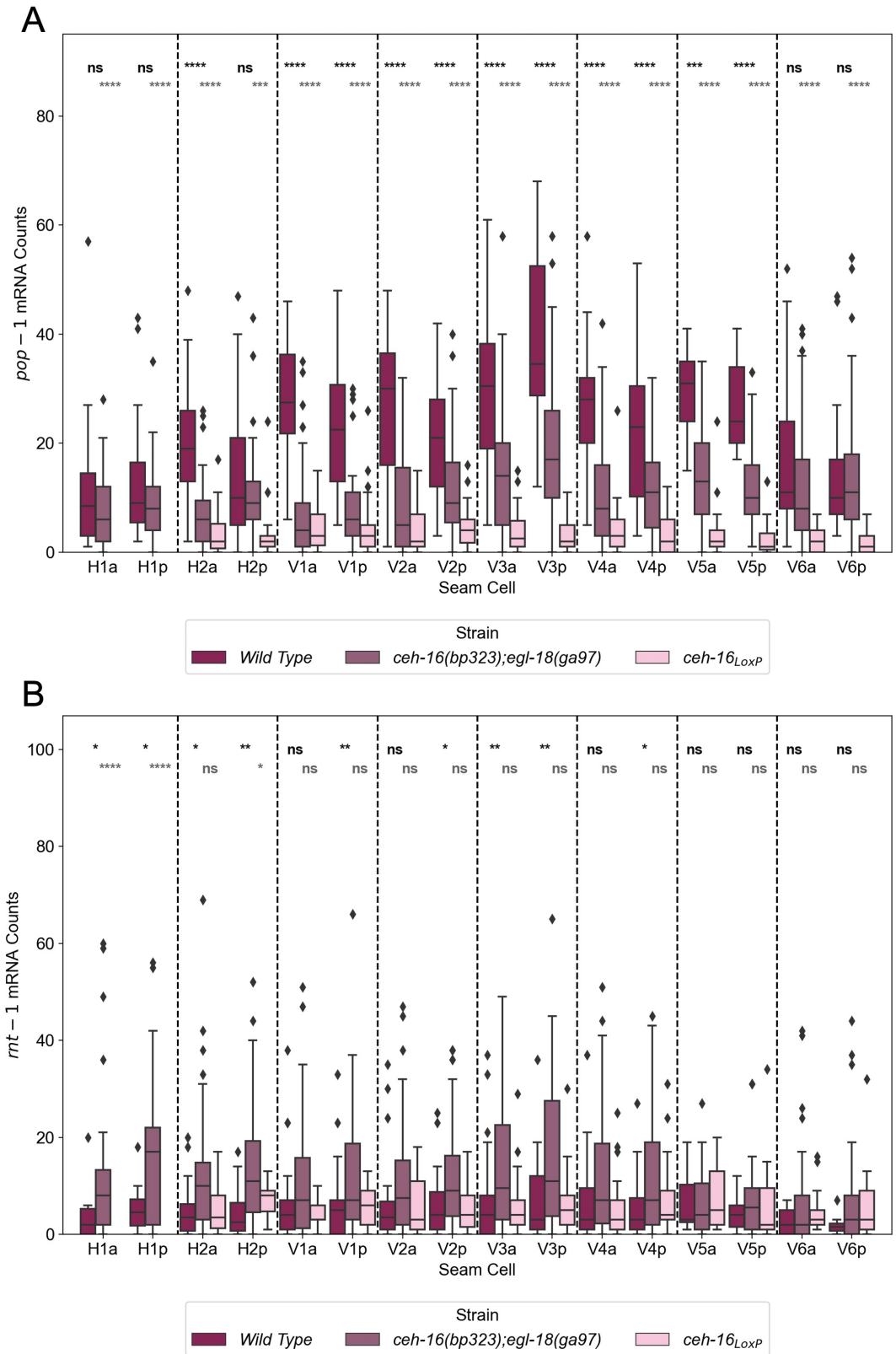


Figure 5.7: **Change in *pop-1* and *rnt-1* levels between wild type and *ceh-16(bp323);egl-18(ga97)* double mutant and *ceh-16*$_{LoxP}$ at the L2 symmetric division: A)** A graph showing the number of *pop-1* mRNA spots and **B)** *rnt-1* mRNA spots in wild type, and *ceh-16(bp323);egl-18(ga97)* and *ceh-16*$_{LoxP}$. The p-values were calculated using Mann-Whitney U denoted as described in methods. The stars shown in black show the significance of the difference between wild type and *ceh-16(bp323);egl-18(ga97)* and in gray *ceh-16*$_{LoxP}$.

as H1a, H1p, H2p, V6a, and V6p, behave similarly to wild type, V1-5 are highly affected. Additionally we see an inversion of asymmetry in H2, V1, and V2, as in these three cells in the mutant the posterior cell has higher *pop-1* expression than the anterior cell (Figure 5.7A).

For the *ceh-16*$_{LoxP}$we were unable to quantify *pop-1* expression in the H1 cell as it is largely absent in this mutant. Nevertheless, even more so than in the *ceh-16(bp323);egl-18(ga97)* strain, we observe a significant decrease in *pop-1* expression. In V1 and V2, we note a greater decrease in expression in the anterior cell compared to the posterior cell, with some animals exhibiting non-detectable levels of *pop-1* expression in one or more cells. While the extent of *pop-1* loss varies across cells compared to wild type, every cell experiences this change in the *ceh-16*$_{LoxP}$ strain (Figure 5.7A).

rnt-1 expression levels showed no discernible difference between anterior and posterior cells of the same lineage in both wild type and the *ceh-16(bp323);egl-18(ga97)* and *ceh-16*$_{LoxP}$ mutants. Additionally there was minimal change in mean between the wild type and the *ceh-16*$_{LoxP}$ mutant (Figure 5.7B). We did note a small increase in *rnt-1* mRNA counts in the *ceh-16(bp323);egl-18(ga97)* double mutant, particularly in the posterior cells of V1, V2 and V4 and in H1, H2 and V3. The increase in *rnt-1* levels could explain the low levels of *pop-1*(Figure 5.7B).

Our next candidates were *efl-3*, *lit-1* and *apr-1*. We quantified *efl-3* as its loss resulted in seam cell duplications in wild type but failed to do so in the mutants. Additionally it has been shown (unpublished lab data) that *efl-3* regulates *lit-1* which is involved in POP-1 export. If *pop-1* levels are low but *lit-1* isn't functional or is highly repressed by *efl-3* then the POP-1 protein would accumulate in the nucleus still resulting in the loss of seam cell fate mainly in the anterior. We also quantify *apr-1* levels as we saw large gains in seam cell number in both the double mutant and the seam cell specific *ceh-16* mutant in absence of *apr-1*, indicating that the loss of symmetry in both these mutants occurs via *apr-1*.

In the wild type expression pattern of *efl-3*, we observed a subtle anterior-to-posterior difference in wild-type expression, particularly noticeable in cells V4, V5, and V6, although overall expression appears relatively consistent along the nematode's body (Figure 5.8A). This anterior posterior pattern is not observed in either of *ceh-16(bp323);egl-18(ga97)* and *ceh-*

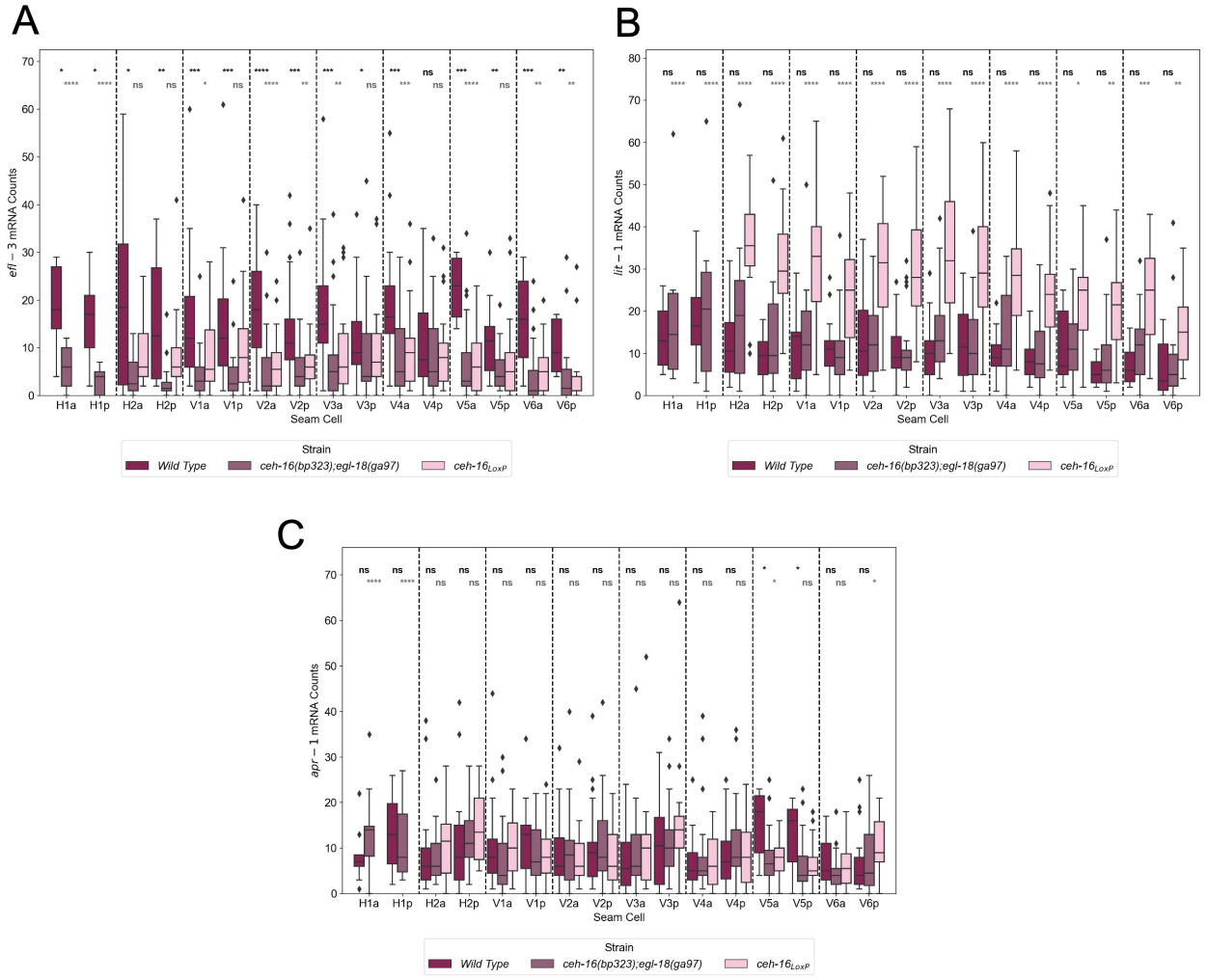


Figure 5.8: **Change in *efl-3*, *lit-1* and *apr-1* levels between wild type, *ceh-16(bp323);egl-18(ga97)* double mutant and *ceh-16*$_{LoxP}$ at the L2 symmetric division**: **A)** *efl-3* expression levels **B)** *lit-1* expression levels and **C)** *apr-1* expression levels in *wild type, ceh-16(bp323);egl-18(ga97)* and *ceh-16*$_{LoxP}$ in different cells. The p-values were calculated using Mann-Whitney U, denoted as discussed in methods. The asterisks are in black for *ceh-16(bp323);egl-18(ga97)*, and in grey for *ceh-16*$_{LoxP}$.

16$_{LoxP}$ (Figure 5.8A). There is however a significant loss of *efl-3* in these mutants particularly in the anterior cells. Particularly in the *ceh-16*$_{LoxP}$ mutant the posterior V1,V3, V4 and V5 cells' expression is comparable to that of the wild type, whereas the anterior cells show significant losses in expression (Figure 5.8A).

In the wild type animal, *lit-1* expression does not exhibit a clear anterior-to-posterior pattern, with all cells displaying very similar expression levels (Figure 5.8B). The only cells with slightly lower expression are V5p and both the V6 cells. The *ceh-16(bp323);egl-18(ga97)* strain shows no significant difference in *lit-1* expression levels compared to the wild type. However, a significant increase in *lit-1* levels is observed in the *ceh-16*$_{LoxP}$ strain. We note that *lit-1* expression becomes slightly asymmetric in H2, V1 and V6 cells, however all of the cells experience a rise in *lit-1* mRNA counts, often doubling the expression observed in equivalent wild type cells. The difference in *lit-1* expression between the *ceh-16*$_{LoxP}$ strain and the *ceh-16(bp323);egl-18(ga97)* strain is of particular interest, considering both strains have a disrupted *ceh-16* gene (Figure 5.8B).

Despite *apr-1* being the gene that resulted in the most significant changes in seam cell number when absent in both mutants, we observed no significant changes in *apr-1* expression in the seam cells of either the *ceh-16(bp323);egl-18(ga97)* double mutants or the *ceh-16*$_{LoxP}$ mutants, except for V5. We also note no anterior-posterior pattern between the cells except in V3 in the wild type and the *ceh-16*$_{LoxP}$ mutant and even then this difference appeared to be minimal (Figure 5.8C).

Given the lack of change in *apr-1* expression in the double mutant, we hypothesized that *apr-1* function was mediated by *egl-18* and *ceh-16*, and not vice versa. As such we employed the *apr-1(bp298)* loss-of-function mutant to compare *egl-18* and *ceh-16* levels to the wild type. *ceh-16* is expressed symmetrically in all cells, except H2, both in the wild type and *apr-1(bp298)*. However, the striking result is the increase in expression in *apr-1(bp298)*. *ceh-16* expression appears to be more than triple what it is in wild type in every seam cell (Figure 5.9 A).

Likewise *egl-18* also exhibits no asymmetry between anterior and posterior cells during the L2 symmetric division, with the exception of H2, in both wild type and *apr-1(bp298)* (Figure

5.9 B). However, we observed a highly significant increase in expression levels of *egl-18* in the absence of *apr-1* (Figure 5.9 B). This increase in expression is less pronounced than that observed with *ceh-16*; however, it is present in all cells except for V4a.

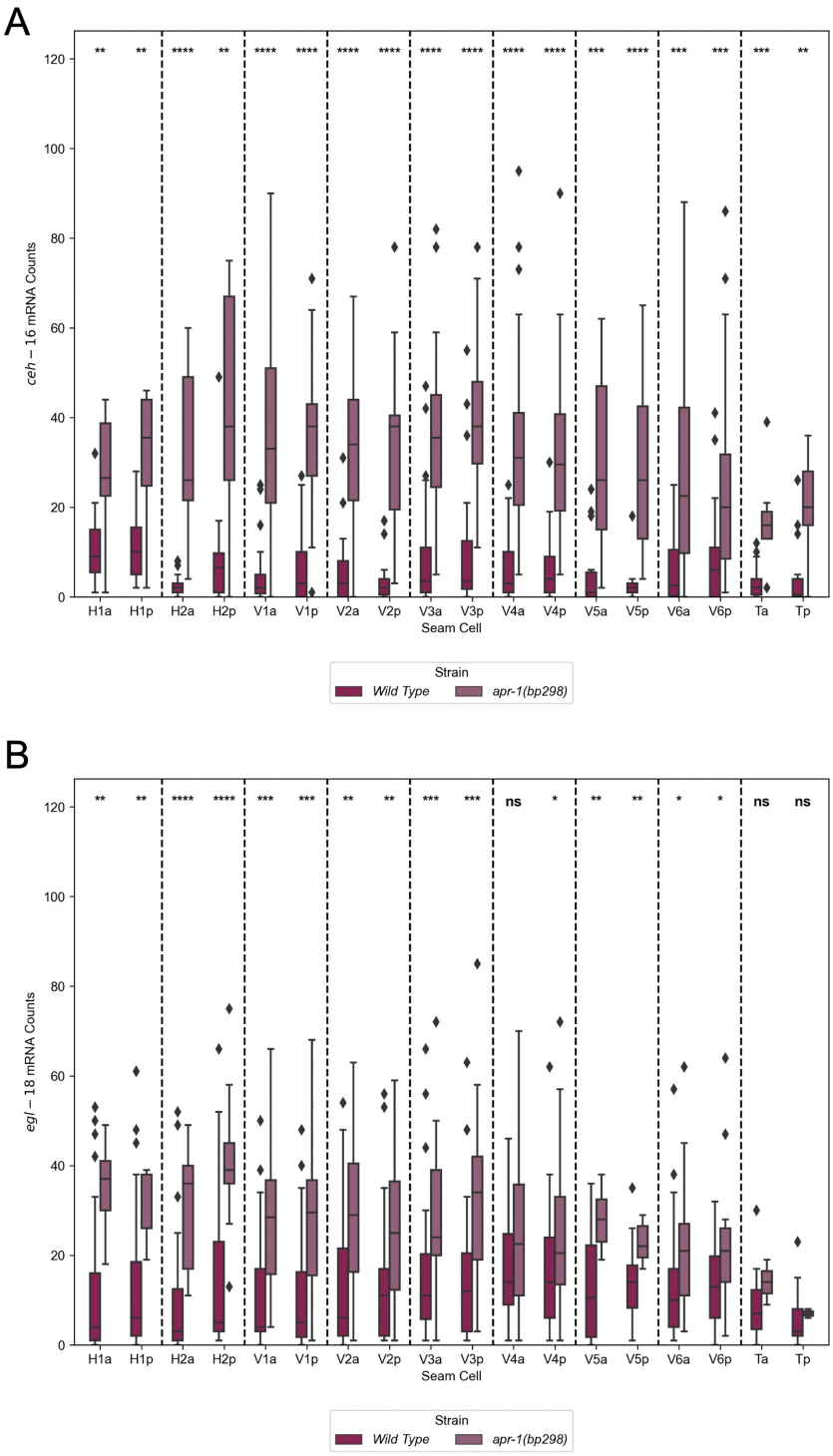


Figure 5.9: **Change in *ceh-16* and *egl-18* levels between wild type and *apr-1(bp298)* at the L2 symmetric division**: **A)** *ceh-16* expression levels and **B)** *egl-18* expression levels in wild type and *apr-1(bp298)*in different cells. The p-values were calculated using Mann-Whitney U, denoted as discussed in methods. The comparison between anterior an posterior cells is shown for each strain in the colour that represents it in the legend and the graph. The comparison between the mutant strain and the wild type is shown above each cell in black.

5.3 Discussion

5.3.1 Robustness of the double mutant

One of the main reasons for our initial interest in the *ceh-16(bp323);egl-18(ga97)* double mutant was our observation of the robustness of its phenotype. For the purpose of this thesis we define a robust phenotype as a phenotype that is consistent and resistant to change when exposed to internal (genetic) and external (environmental) perturbations, whereas a consistent phenotype as one with low variance. This definition is in line with Barkoulas et al. (2013). According to this definition we find that the *ceh-16(bp323);egl-18(ga97)* phenotype is both consistent and robust. The strain showed resistance to both internal perturbation, throught the disruption of core seam cell genes, as well as temperature. Given the opposed patterns of temperature sensitivity in the parental strains, one could attribute the lack of change in seam cell number in the double mutant to a counterbalancing between the sensitivities of the two parental strains. However, upon observation of the cell-specific loss patterns, this appears to be highly unlikely, as the cells lost always follow the same pattern regardless of temperature; a pattern that does not truly reflect either of the parental strains. In fact the cells that the animal hatches with appear to be significantly more resistant to perturbation than either of the parental strains, which are prone to losing cells during larval development. This points to a balancing interaction between the two genes. We theorise that the genes function though a counterbalancing mechanism. This is supported by the interactions we have proposed between them in Chapter 3 and 4 as an activation repression feedback loop would be characteristic of genes that require highly consistent expression levels relative to eachother.

Our hypothesis that the double mutant phenotype is solely attributed to the *ceh-16* mutation being enhanced by the absence of *egl-18* activation is challenged by our observations of the *ceh-16*$_{LoxP}$ animals. While the pattern of seam cell loss in the *ceh-16*$_{LoxP}$ mutant closely resembles that of the *ceh-16(bp323);egl-18(ga97)* mutant, and the temperature sensitivity exhibits similar trends, we observe a significantly higher variability in the tissue-specific *ceh-16* mutant. Although some variability can be attributed to Cre expression noise, and thus

considered as technical error, we observe seam cell losses beyond what can be attributed to the absence of seam cell gain during development. The *ceh-16*$_{LoxP}$ strain frequently loses seam cells that they hatch with, particularly H1, indicating a non-identical mechanism between the two. However, as with the *ceh-16(bp323);egl-18(ga97)* mutant, predominant pattern between the cells in *ceh-16*$_{LoxP}$ is highly indicative of the loss of symmetry at L2.

The difference between the double mutant and the tissue specific *ceh-16* mutant is underscored by the core seam cell gene RNAi experiment. We note differences in variability in all treatments between the double mutant and the LoxP mutant. Notably, the LoxP mutant exhibits a pronounced seam cell phenotype in the absence of *egl-18*, with cell counts significantly below the baseline 10 cells with which the animal hatches. This suggests that the combined reduction in *ceh-16* expression and loss of *egl-18* manifests this robust phenotype only within a specific expression range. Below a certain threshold, it results in the loss of cell fate. Furthermore, LoxP animals treated with *egl-18* RNAi fail to produce viable progeny, further highlighting that the phenotype we observe in the double mutant is threshold specific.

Given that *elt-6* is a paralog of *egl-18*, we initially anticipated similar regulatory mechanisms between these two genes. Consequently, we hypothesized that the observed robustness in seam cell number could be attributed to an upregulation of *elt-6* expression in the absence of *ceh-16*, which we have identified as a repressor of *egl-18*. Nevertheless while the absence of *elt-6* does appear to have an effect on the double mutant, we never see a seam cell count below 9, and even the animals with 9 seam cells have only a missing T cell. This suggests that the robustness of the seam cell loss of symmetry phenotype in the *ceh-16(bp323);egl-18(ga97)* is not contingent upon an increase in *elt-6* expression. Furthermore, we observe no discernible effect from *elt-6* treatment on the LoxP mutant, indicating that the stability of seam cell number in both strains is not mediated by alterations in *elt-6* expression.

The *elt-1* treatment underscores the critical role of *elt-1* in seam cell maintenance, as evidenced by a consistent maximum mean number of 6 seam cells across all strains. It also demonstrates that while the loss of *egl-18* alone leads to a less severe phenotype than the loss of *elt-1*, the simultaneous absence of both genes results in an even lower seam cell

count than either of the individual losses. This highlights the importance of *egl-18* in seam cell maintenance but also indicates that *egl-18* may not be entirely dependent on *elt-1* but rather may be activated in a separate pathway. This aligns with our previous observations in Chapter 3, where we described *egl-18* as conditionally activated by ELT-1 in the presence of CEH-16 repression. Additionally, the *egl-18(ga97)* strain doesn't lay viable progeny when treated with *elt-1* RNAi, once again showing both the importance of the genes in the animals' development, and the importance of the seam cells for *C. elegans* survival.

Despite the fact that the *egl-18(ga97)* mutant suffers an exacerbation of the seam cell response to *elt-1* RNAi compared to the wild type, when this strain is crossed to the *ceh-16(bp323)* the *elt-1* treated seam cell phenotype returns back to wild type levels, again showing that the partial loss of *ceh-16* paired with the complete loss of *egl-18* is more robust than its parentals. When the double mutant is treated with *elt-1* RNAi, the resulting phenotype mirrors that of both the wild type and the *ceh-16(bp323)* mutant, suggesting that this particular reduction in seam cell numbers is primarily driven by *elt-1* alone. The *ceh-16*$_{LoxP}$ exhibits a slightly lower seam cell count compared to the wild type, *ceh-16(bp323)* and *ceh-16(bp323);egl-18(ga97)*, albeit higher than that of the *egl-18(ga97)* mutant. This again disproves the theory that the double mutant behaviour is caused solely by a more severe loss of *ceh-16*, and it shows that a complete loss of *ceh-16* sensitises *C.elegans* to *elt-1* losses. Moreover, the lower phenotype observed in *elt-1* RNAi-treated *ceh-16*$_{LoxP}$ animals compared to wild-type animals treated with *elt-1* suggests that while *ceh-16* is downstream of *elt-1*, its activation may involve additional regulatory mechanisms beyond *elt-1* alone.

Overall, the double mutant displays greater resilience to both environmental and genetic perturbations compared to either of its parental strains. While the phenotype observed in these animals closely resembles that of a complete loss of *ceh-16* in the seam cells, significant differences between the two strains suggest that neither the robustness nor the seam cell loss pattern can be solely attributed to the loss of EGL-18-specific activation of impaired *ceh-16*. To fully comprehend the outcomes of both the double mutant and the tissue-specific *ceh-16* mutant, a comprehensive understanding of seam cell behavior in these strains is essential.

5.3.2 Seam cells lose capacity to divide symmetrically at L2 in *ceh-16(bp323);egl-18(ga97)* and *ceh-16*$_{LoxP}$

In order to understand the robustness of the two strains we had to observe the changes that occurred in their seam cells. After observation of the behaviour of the seam cells over the entire course of development we noted normally occurring seam cell divisions with one exception. The L2 symmetric division. We confirm the asymmetry of this division through the lack of quadruplet formation and we established that the anterior cell initiated the differentiation program into hyp7 through observation of *eff-1*. Specifically, at 26 hours post-hatch, all observed divisions featured *eff-1* expression in the anterior cell, indicative of its commitment to differentiation and fusion. Intriguingly, we identified two distinct divisions during L2, suggesting a conversion from symmetric to asymmetric division. Therefore we conclude that the phenotype we observe both in the *ceh-16(bp323);egl-18(ga97)* and the *ceh-16*$_{LoxP}$ is caused by the conversion of the symmetric division to an asymmetric division. Further observations in the LoxP mutant revealed the consistent absence of H1 and frequent missing seam cells, suggesting impaired seam cell maintenance, a phenomenon that we do not observe in the *ceh-16(bp323);egl-18(ga97)* double mutant.

5.3.3 *ceh-16* and *egl-18* both influence and are influenced by genes known to be involved in division symmetry

To fully understand the behaviour of these two mutants we directed our focus on the L2 symmetric division. In order to do this we knocked down key genes involved in symmetry using RNAi in wild type, *egl-18(ga97), ceh-16(bp323), ceh-16(bp323);egl-18(ga97) and ceh-16*$_{LoxP}$. The genes we selected can be separated into three groups. Those that caused a decrease in seam cell number (*lit-1* and *efl-3*) those that caused an increase in seam cell number (*apr-1* and *pop-1*), and those with little to no effect (*rnt-1* and *eff-1)*. The latter of these are still considered genes of interest as their apparent lack of effect may stem from the RNA interference technique itself being ineffective, given that mutants of both of these genes cause severe seam cell defects.

pop-1 is a pivotal gene involved in the switch between symmetry and asymmetry. Its expression is typically downregulated by the RNT-1/BRO-1 complex, enabling cells to divide symmetrically despite the asymmetrical distribution of many *Wnt* components. We observed an increase in seam cell counts in the *ceh-16(bp323)* mutant upon *pop-1* knockdown, but a decrease in counts in the *egl-18(ga97)* mutant under the same treatment. This is in line with previous findings by Gorrepati et al. (2013) which indicate that *egl-18* may be downstream of *pop-1* and part of the role of *pop-1* in asymmetric divisions, involves the downregulation of *egl-18*. Our seam cell counts indicate that seam cell gains in absence of *pop-1* are dependent on the presence of *ceh-16* and potentially *egl-18*. Given that the absence of *pop-1* causes seam cell gains, and that a loss of *pop-1* is normally associated with an increase in symmetry (van der Horst et al., 2019; Gorrepati et al., 2013) we expected the *pop-1* levels in both the double and the LoxP mutants to be higher than in the wild type. Nevertheless we note highly significant decreases in *pop-1* expression compared to the wild type. This indicates that the loss of symmetry in both these mutants occurs independently of *pop-1* as such we had to interrogate other potential mechanisms. We hypothesise that the decrease of *pop-1* is a response to the loss of symmetry as the system attempts to lower *pop-1* levels to permit the symmetric divisions to occur. We also note slight changes in expression between anterior and posterior cells, however this is unlikely to explain a full loss of symmetric division. This means that the change we observe in *pop-1* expression are the effect of the system trying to regain symmetry. We also show that low *pop-1* is insufficient to induce a symmetric division, unlike what is shown in van der Horst et al. (2019).

In literature, the *ceh-16* loss of symmetry effect is described as being rescued through a complete knockout of *apr-1* (Huang et al., 2009). We see this effect in the *ceh-16*$_{LoxP}$ mutant as well as the *ceh-16(bp323)* mutant through *apr-1* RNAi treatment, although not to the extent described with the *apr-1(bp298)* cross. In our RNAi screen we note that the loss of *apr-1* does indeed allow for an increase in seam cell number in wild type and all three of the *ceh-16* mutants. However, we only see an increase in variability in *egl-18(ga97)* with no gain in seam cell number. This suggests that the function of *apr-1* in promoting symmetrization may be dependent on *egl-18*. However, we also observe an extremely large increase in seam cell

number in the double mutant, which disputes this hypothesis. The extreme effect of *apr-1* RNAi on the double mutant could be partially due to the decrease in *pop-1* levels.

The function of APR-1 is closely linked to *Wnt* signaling and the nuclear export of β-catenins in the anterior cell, its absence leading to the symmetrisation of the final asymmetric division in the seam cells. The fact that *ceh-16(bp323)* has been shown to gain symmetric divisions in absence of *apr-1*, does however show that *apr-1* is able to somewhat compensate for *ceh-16* deficiencies. Both the double mutant and the *ceh-16*$_{LoxP}$ mutant are capable of carrying out symmetric divisions when *apr-1* is removed. This suggests that *ceh-16* and *egl-18* are not essential for all aspects of seam cell symmetry. However, we did not observe symmetrisation of the L2 division even in the absence of *apr-1*. This could be either a result of delayed feeding on the *apr-1* RNAi or it could mean that specifically the L2 division requires *ceh-16* and to a lesser extent *egl-18*.

Despite the strong influence that the absence of *apr-1* has on both the *ceh-16(bp323);egl-18(ga97)* and the *ceh-16*$_{LoxP}$ mutants, we observed no changes in expression of *apr-1* in either of them. Nevertheless we do note strong increases in expression in both *ceh-16* and *egl-18* in absence of *apr-1*. Given the evidence for *ceh-16*'s role in the symmetric division (Huang et al., 2009) and it's rescue when crossed to the *apr-1(bp298)* and the fact that *ceh-16(bp323)* is a partial loss-of-function mutant, it could be that the symmetrisation caused by the loss of *apr-1* in a *ceh-16(bp323)* mutant is caused by an increase in *ceh-16* expression caused by the absence of *apr-1* which compensates for the protein's decreased function. This may explain why the *ceh-16*$_{LoxP}$ mutant does not respond as robustly as *ceh-16(bp323)* and *ceh-16(bp323);egl-18(ga97)* to the loss of *apr-1*. The limited response of *egl-18(ga97)* suggests the importance of this transcription factor (TF) in *apr-1*-dependent symmetrisation. The elevated levels of *egl-18* caused by the loss of *apr-1* could be necessary to induce symmetry the seam cells. However, the knockdown of *ceh-16* together with the removal of *egl-18* overrides *egl-18*'s resistance to *apr-1* downregulation. This could be due to the loss of *pop-1* in the double mutant. As such I believe that the symmetrisation induce by the lack of *apr-1* requires: low *pop-1*, low *egl-18* and low *apr-1*. I believe that knocking out *egl-18* alone doesn't reduce *pop-1* levels to a sufficient level to induce this symmetry.

Another striking aspect of both the *ceh-16(bp323);egl-18(ga97)* and the *ceh-16*$_{LoxP}$ mutants, is their response to the loss of *lit-1*. LIT-1's role in symmetry is often portrayed as a regulator of POP-1. It phosphorylates POP-1, which in turn gets exported, allowing for the posterior cell to continue in its differentiation program (Harandi and Ambros, 2015). In its absence we do observe very severe losses of seam cells which can be explained through the nuclear accumulation of POP-1. Nevertheless we believe this to not be the case at least for the double mutant as *lit-1* levels are unchanged in the *ceh-16(bp323);egl-18(ga97)*. As such we believe that the loss of symmetry in this mutant is independent of *pop-1*. What is striking however is that in the *ceh-16*$_{LoxP}$ mutant we do see increased levels of *lit-1*, which further highlights the differences between the two mutants and hints towards different mechanisms of symmetry loss. The complete loss of *ceh-16* does cause an accumulation of *lit-1* and thus potentially of nuclear POP-1 which could mean that the loss of symmetry in *ceh-16*$_{LoxP}$ could be centred around POP-1. Nevertheless this is unlikely to be the case for *ceh-16(bp323);egl-18(ga97)* despite this mutant also having severe disruptions in *ceh-16*. This ought to be confirmed through protein reporters of POP-1 crossed into *ceh-16(bp323);egl-18(ga97)*. This would be a challenging endeavour due to the physical malformations of the double mutant, however would determine whether the *ceh-16* and *egl-18* provide a *pop-1* independent symmetry pathway. We believe that the mechanism of symmetry loss in the double mutant is linked to its robust phenotype. In absence of the need to carry out a self-renewal division the cells are also less likely than either of its parentals to lose the seam cell fate. This is backed up by the fact that the least consistent cell is the T-cell which maintains a self-renewal division.

Overall, in both the *ceh-16(bp323);egl-18(ga97)* and the *ceh-16*$_{LoxP}$ mutants we observe a loss of symmetry in the L2 division. Despite the seam cell number phenotype being highly similar, there is a difference in variability between the two strains, as well as the additional seam cell losses, particularly of H1, that we observe in the *ceh-16*$_{LoxP}$. Despite the loss of symmetry being similar between the two mutants, the mechanism by which it is induced appears different. The *ceh-16*$_{LoxP}$ loss of symmetry appears to be centred around a gain in *lit-1* expression, whereas in this is not the case in the double mutant. We corroborate

findings by Huang et al. (2009) that APR-1 appears to rescue a *ceh-16*-based loss of symmetry. We believe that the low levels of *pop-1* in the double mutant primes the cells for a symmetric division, which is then triggered when *apr-1* is removed. The symmetrisation of *apr-1(bp298);ceh-16(bp323)* observed in Huang et al. (2009), is likely linked to an overexpression of the partially functional *ceh-16* which explains the lesser response of *ceh-16*$_{LoxP}$.

Chapter 6

General Discussion

6.1 The new quantitative gene regulatory network underlying core seam cell genes at early L1

In order to start the construction of a detailed GRN for the epidermal stem cells in *C. elegans*, we initially focused on three core seam cell genes: *elt-1*, *ceh-16* and *egl-18* at the late L1 stage, which we theorised to represent a steady state. This was later shown to be a difficult assumption to make due to the dynamics of the core seam cell genes, however the 17 hour time point appears to be the closest approximation of a steady-state one can observe in early larval development. *elt-1*, *egl-18* and *ceh-16* are essential for correct seam cell specification (Baugh et al., 2005; Smith et al., 2005; Page et al., 1997a; Gorrepati et al., 2013; Koh and Rothman, 2001; Koh et al., 2002; Huang et al., 2009). However, this does not necessarily indicate that no other genes are involved in the core seam cell network. We initially focus on the interactions between the three core genes in order to further expand the network around them. We rely on smFISH data to infer quantitative relationships between the genes, which we modeled using a new adaptation of the MRA (Kholodenko et al., 2002). This adapted version of the MRA enabled us to work with loss-of-function mutants rather than knockdowns, which is particularly useful for working with essential genes where complete knockdowns are lethal. We applied this approach to infer strengths and polarity of the interactions between these genes. A key consideration in our data analysis is that the perturbations applied are

substantial. Despite the use of the MRA with complete knock outs in Borg et al. (2023), the algorithm was originally developed to deal with smaller perturbations (Kholodenko et al., 2002). This limitation may lead to erroneous estimates regarding the strengths and directionality of the interactions. However we performed additional mathematical testing by utilising a set of ODEs which we simulated using a Bayesian inference approach called SLING (Jørgensen et al., 2023). We initially informed the network using the values obtained from the MRA but then relaxed the constraints on the polarity of the interactions. Both the methods reached the same conclusions as to the polarity and the strengths of the interactions. The ODE approach allowed us to explore different parameter combinations within the network and assess their necessity for a minimal network. We compared our network to the one previously described in the literature (Thompson et al., 2016) and confirmed that our adjustments led to more accurate data predictions even in novel mutants. Our revised network introduced several changes from the literature-based model. Although the ELT-1 activation of *ceh-16* and *egl-18* was consistent with previous findings, the activation of *egl-18* appeared uncertain and occurred only in response to gene repression. We also revised the interaction from CEH-16 to *egl-18* to a repression. Additionally we identified a novel link, an activation from EGL-18 to *ceh-16*.

We use the revised network to make predictions for the genes' response to the removal of *ceh-16* and *egl-18* together. Our predictions accurately matched the expression levels observed in the *ceh-16(bp323);egl-18(ga97)* mutant. Coincidentally, this mutant results in a loss of seam cell phenotype with a highly invariable terminal seam cell number which prompted us to conduct further investigation.

Our protein data at the same time point shows further complexity of the network. While some of our interactions remain consistent at protein level, others do not, like the repression of *egl-18* by CEH-16. While we do test this interaction further at other time points, the fact of the matter remains that we have not explored the post-transcriptional regulation of the system. Given the importance of miRNA in the *C. elegans* regulatory landscape, we believe that these findings are an opportunity to explore miRNA based regulation of the seam cell core genes and a better understanding of the behaviour of the proteins and not just the transcripts of the

the three core genes.

6.2 The seam cell GRN changes over time and position

Having generated the model of the seam cell GRN at the late L1 stage, we compared this network to the GRN at other time points. We sought to identify significant changes in the network's state due to the differing roles of seam cells across larval development. We note that the most consistent interaction across all time points is the repression of *egl-18* by CEH-16, which aligns with the established function of its homolog *engrailed* (Markel et al., 2002; Alexandre and Vincent, 2003; Genestine et al., 2015). However, we also note significant time-specific differences. Notably, in earlier stages, there is limited activity directed towards *elt-1*, while activation of *elt-1* occurs during both symmetric and asymmetric divisions. The dynamics of these interactions are complicated by high noise levels in the data.

In later stages, we observed a shift in the role of *elt-1*. Despite its identification as an activator, our studies at 17 hours indicate strong repression of *ceh-16* and *egl-18* by ELT-1 in late L2. This suggests a potential change in behavior or the activation of an intermediate regulator, possibly NHR-25, which has shown peaks in expression in ELT-1 ChIP-seq data and has been reported to negatively regulate *egl-18* (Luo et al., 2020; ENCODE Project Consortium, 2012; Katsanos and Barkoulas, 2022). We do not however have any data indicating it could downregulate *ceh-16*. Another candidate we would strongly recommend for further study is *unc-37*, whose homolog *groucho* represses the *ceh-16* homolog *engrailed*.

Additionally despite our knowledge on the binding sites of both *ceh-16* and *elt-1*, we have no such information on *egl-18* thus in order to confirm the directness of our interactions a proposed approach would be using techniques like TaDa (Marshall et al., 2016) to identify EGL-18 binding sites and confirm in particular whether it could bind the *ceh-16* promoter region.

We observe that the peak gene expression occurs during the self-renewal division. We also note that the gene expression levels fall at the end of L2. We predict that in order for the seam cell genes to exit their seam cell fate and differentiate terminally at the end of L4,

the core gene levels need to be extremely low. As such they are downregulated over time through repressive loops starting from the symmetric division onwards. To further investigate this an a follow up study on the core genes behaviour in the two later larval stages would be required.

Furthermore, we observed that differences in gene expression and loss frequencies among individual seam cells are more closely linked to position than to lineage or downstream roles. Using graded RNA interference (RNAi) to induce progressive losses of seam cells, we focused on *egl-18* and *elt-1*. Notable differences in sensitivity were observed between anterior (V1 and V2) and posterior (V3 and V4) V cells. Prior research has often treated V1-V4 as a homogenous group (Katsanos and Barkoulas, 2022; Ren and Zhang, 2010; Koneru et al., 2021b; Mallick and Yi, 2014). However, our findings suggest that this grouping can obscure critical nuances necessary for understanding seam cell behaviors, including cell fate determination. Our data show that the amount of active *elt-1* correlates directly with seam cell retention by the end of L4. Anterior cells (V1 and V2) express lower *elt-1* levels and are more prone to loss than posterior cells (V3 and V4), particularly following asymmetric division. Similar expression patterns between H1 and H2 and the anterior V cells further support this trend. Thus, it is beneficial to group V1 and V4 together while treating head cells separately for the L1 stage, while later larval development may require single-cell resolution.

We also show that the anterior-posterior patterning of the *elt-1* loss of seam cells is driven by Wnt. While *Wnt* has been previously shown to interact with the core gene network through the *egl-18/pop-1* (Gorrepati et al., 2013) interaction a *Wnt* effect has not previously been shown on *elt-1*. We note that a loss of any of the three *Wnt* effectors we have tested results in increasing losses in the posterior cells in absence of *elt-1*. A further study we believe would greatly inform the mechanism behind seam cell patterning in absence of *elt-1* is testing whether overexpression of *Wnt* effectors, decreases the sensitivity of anterior seam cells like H1p and the V1s.

Another important study that ought to be carried out to fully characterise the seam cell network is in-vivo imaging. This could be done using MCP/GFP and MS2 constructs (Tutucci et al., 2018; Kinney et al., 2023; Lee et al., 2019; Tocchini and Mango, 2024; Keil et al.,

2017). Our observations indicate highly variable dynamics of the three genes, and the repression/activation links between *ceh-16* and *egl-18* particularly at 17 hours could be indicative of oscillatory gene expression. As such live imaging would shed further light on the gene regulatory patterns of the three genes as well as illustrating their dynamics.

In order to fully also understand the spatial patterning of the seam cells one would have to generate gene knockouts specific to individual cells. We believe that while currently this approach is not feasible we believe that the LoxP based approach is the most promising for driving this forward. One possible technique that could be evaluated is *Wnt* or *Hox* gene driven CRE. This could allow regional targetting and *Wnt* specific targetting on the seam cells. One approach we would recommend is the generation of a MAB-5:CRE which would allow the targeting of V3 and V4 (Smith and Kratsios, 2024; Cowing and Kenyon, 1992). This crossed with a *ceh-16*, or *elt-1* LoxP strain might be used to disrupt the robust behaviour of these two cells. In summary, while we have made substantial progress in understanding the core seam cell GRN, the network's complexity suggests the need for further research.

Finally if we compare our temporal analysis of the seam cell core gene network to the Yamanaka network, we note that the late L1 timepoints more closely resembles the pluripotent cells. This is due to the large amount of positive interactions which get aren't present at other time points. As such it would be an interesting consideration to evaluate whether the L2 and early L1 cells could be more prone to differentiation. An experiment that could address this could be a temperature based perturbation and seam cell quantification at various stages.

6.3 The complete loss of *egl-18* coupled with a partial loss of *ceh-16* causes a robust loss of symmetry

Most mutants with an aberrant seam cell count exhibit considerable variability in cell counts across different animals (Huang et al., 2009; Hintze et al., 2020, 2021; Katsanos et al., 2017). We have identified a mutant that not only shows low variability in seam cell number in the wild type and under normal conditions, but that also is resistant to environmental and genetic

perturbation. The robust phenotype of the *ceh-16(bp323);egl-18(ga97)* double mutant was initially thought to arise from an exacerbation of symmetry loss due to the absence of *ceh-16* (Huang et al., 2009) through the loss of the 17-hour activation of *ceh-16* by EGL-18 (as discussed in Chapter 3). To test this we generated a *ceh-16*$_{LoxP}$ seam cell specific knockout. This mutant did have a similar phenotype to what we observed with *ceh-16(bp323);egl-18(ga97)* however it lacked the robustness we observed in the double mutant. Both mutants were shown to lose the L2 symmetric division, however *ceh-16*$_{LoxP}$ also occasionally failed to maintain seam cells. The loss of *egl-18* combined with a partial loss of *ceh-16* generated seam cells that while failing to self-renew at L2, were a lot more robust.

The question that arises here is whether a stem cell that loses is capacity for self-renewal remains a stem cell and whether the cells fail to self-renew under the current conditions or are inherently incapable of doing so. We know from our seam cell specific quantifications of losses and gains in the *ceh-16(bp323);egl-18(ga97)* that the T cell maintains the capacity to self renew, even if the other seam cells don't. We also show that in absence of either *pop-1* or *apr-1* self-renewal divisions do occur, however this involves genetic perturbation beyond what would be considered a normal self-renewal, nevertheless we do show that these cells are capable of dividing. The ultimate proof of concept here would be identifying conditions, such as a change in temperature, or the ablation of one of the seam cells that would trigger a self-renewal division. However, given that the T cell does divide and that with RNAi based perturbation of *apr-1* and *pop-1* we do observe some self-renewal divisions, we conclude that these cells remain stem-like cells and in fact become a more robust version of their wild type equivalent.

To fully understand this phenotype, it would be valuable to identify a mutation that can restore the L2 symmetric division to its normal behavior in the *ceh-16(bp323);egl-18(ga97)* mutant. So far the best candidate for this is targeting the *apr-1* gene. While the RNAi converts L3 and L4 divisions to symmetric in the mutant, we believe that a cross to the *apr-1(bp298)* mutant could rescue the L2 division as it does in (Huang et al., 2009) and we believe this would be an essential experiment despite the difficulty of crossing *ceh-16(bp323);egl-18(ga97)* due to its physical malformations. The mechanism by which the mutation remains so robust, we

believe is centred around its loss of symmetry, as such a key observation that would need to be made is whether the robustness is maintained even when the symmetry loss is rescued. If that is the case that would be evidence that robustness in seam cell number is centred around *ceh-16* and *egl-18*, nevertheless if in regaining symmetry we see a loss of robustness then the robustness itself is due to the cells avoiding the self-renewal division.

6.4 The core seam cell gene network regulates essential genes involved with asymmetry

There are many key players involved in seam cell divisions. We have chosen a few key genes to examine their behavior in our two loss-of-symmetry mutants. We showed that these genes in a state that would be expected to induce hyperplasia: decreased *pop-1* levels (Calvo et al., 2001; van der Horst et al., 2019), elevated *rnt-1* levels (van der Horst et al., 2019), elevated *lit-1* for *pop-1* nuclear export (Lo et al., 2004). These changes would normally be indicative of animals predisposed to hyperplasia. However, only downregulation of *apr-1*, whose expression was unaltered in our mutants, resulted in increased seam cell counts.

Downregulating *apr-1* had a particularly strong seam cell gain effect in the *ceh-16(bp323);egl-18(ga97)* double mutant, which we believe to be linked to the fact that the other genes involved with symmetry are primed for a symmetric division. However, in order for the double mutant to carry out the symmetric division *apr-1* has to be downregulated. We observe elevated expression of both *egl-18* and *ceh-16* in the *apr-1(bp298)* mutant, which explains why the partial loss of function *ceh-16(bp323)* mutant is rescued in the absence of *apr-1* (Huang et al., 2009) as the increase in expression, could compensate for the partial loss of function.

While we know that *apr-1* is essential for the maintenance of asymmetry in the L4 (Ren and Zhang, 2010), further studies of *apr-1* need to be carried out to fully crystallise its role in symmetry in the L2. The observed increase in *ceh-16* expression in the absence of *apr-1* suggests that *apr-1* acts as a repressor of *ceh-16*. As such the duplications in the seam cells

at L4 paired with the decrease in *ceh-16* expression over larval development, could mean that the duplications in the *apr-1(bp298)* are caused by an increase in *ceh-16* levels at L4. This could be tested through *ceh-16* smFISH at L4 stage in the *apr-1(bp298)*.

Additionally given that *apr-1* appears to repress both *egl-18* and *ceh-16* that makes it a candidate for an L2 regulator that results in the repression of the core seam cell GRN. The only caveat for this is the lack of ELT-1 chip-seq peaks in the *apr-1* promoter region.

We would suggest identifying the time point at which the seam cell duplications occur in a *ceh-16(bp323);apr-1(bp298)* mutant as they have only previously been counted at L3 (Huang et al., 2009). We would also expect this mutant to no longer have the L4 duplication seen in the *apr-1(bp298)* mutant as we'd expect the increase in *ceh-16* levels to not compensate for the decrease in protein functionality at that stage. Moreover, RNAi targeting of *Wnt* factors in the *apr-1(bp298)* mutant could provide insights into its role in the choice between symmetric and asymmetric divisions.

Nevertheless we have identified crucial links between the core seam cell genes and symmetry and we have furthered the understanding of the L2 self-renewal division. We have broadened our understanding of *ceh-16* as both a regulator of seam cell self-renewal and a critical factor in seam cell maintenance. Moreover we have identified *apr-1* as a key player in the switch between symmetric and asymmetric divisions and have identified it as a repressor of *ceh-16* and *egl-18*, further linking the core seam cell network to the *Wnt* pathway and expanding our understanding of the seam cell GRN.

References

A. Aboobaker and M. Blaxter. Hox gene evolution in nematodes: novelty conserved. *Current opinion in genetics & development*, 13(6):593–598, 2003.

T. Akiyama. Wnt/β-catenin signaling. *Cytokine & growth factor reviews*, 11(4):273–282, 2000.

T. Akutsu, S. Miyano, and S. Kuhara. *Identification of genetic networks from a small number of gene expression patterns under the Boolean network model*, pages 17–28. Biocomputing'99. World Scientific, 1999.

C. Alexandre and J.-P. Vincent. Requirements for transcriptional repression and activation by engrailed in drosophila embryos. *Development*, 130(4):729–739, 2003.

U. Alon. *An introduction to systems biology: design principles of biological circuits*. CRC press, 2019.

L. Alsina, E. Israelsson, M. C. Altman, K. K. Dang, P. Ghandil, L. Israel, H. Von Bernuth, N. Baldwin, H. Qin, Z. Jin, et al. A narrow repertoire of transcriptional modules responsive to pyogenic bacteria is impaired in patients carrying loss-of-function mutations in myd88 or irak4. *Nature immunology*, 15(12):1134–1142, 2014.

S. Ando, E. Sakamoto, and H. Iba. Evolutionary modeling and inference of gene network. *Information Sciences*, 145(3-4):237–259, 2002.

A. A. Avilion, S. K. Nicolis, L. H. Pevny, L. Perez, N. Vivian, and R. Lovell-Badge. Multipotent cell lineages in early mouse development depend on sox2 function. *Genes I& development*, 17(1):126–140, 2003.

A. Bafico, G. Liu, A. Yaniv, A. Gazit, and S. A. Aaronson. Novel mechanism of wnt signalling inhibition mediated by dickkopf-1 interaction with lrp6/arrow. *Nature cell biology*, 3(7): 683–686, 2001.

A. T. Baldwin and B. T. Phillips. The tumor suppressor apc differentially regulates multiple β-catenins through the function of axin and ckiα during c. elegans asymmetric stem cell divisions. *Journal of cell science*, 127(12):2771–2781, 2014.

G. Balázsi, A. V. Oudenaarden, and J. J. Collins. Cellular decision making and biological noise: from microbes to mammals. *Cell*, 144(6):910–925, 2011.

M. Barkoulas, J. S. van Zon, J. Milloz, A. van Oudenaarden, and M.-A. Félix. Robustness and epistasis in the *C. elegans* vulval signaling network revealed by pathway dosage modulation. *Developmental Cell*, 24(1):64–75, 2013.

W. Bateson. *Materials for the study of variation: treated with especial regard to discontinuity in the origin of species*. Macmillan and Company, 1894.

L. R. Baugh, J. C. Wen, A. A. Hill, D. K. Slonim, E. L. Brown, and C. P. Hunter. Synthetic lethal analysis of *Caenorhabditis elegans* posterior embryonic patterning genes identifies conserved genetic interactions. *Genome biology*, 6(5):1–8, 2005.

T. Bayes. Lii. an essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, frs communicated by mr. price, in a letter to john canton, amfr s. *Philosophical transactions of the Royal Society of London*, (53):370–418, 1763.

K. N. Bekas and B. T. Phillips. unc-37/groucho and lsy-22/aes repress wnt target genes in c. elegans asymmetric cell divisions. *bioRxiv*, pages 2022–01, 2022.

J.-P. Borg, J. Colinge, and P. Ravel. Modular response analysis reformulated as a multilinear regression problem. *Bioinformatics*, 39(4):btad166, 04 2023. ISSN 1367-4811. doi: 10.1093/bioinformatics/btad166. URL https://doi.org/10.1093/bioinformatics/btad166.

T. Boulin, J. F. Etchberger, and O. Hobert. Reporter gene fusions. *WormBook: The Online Review of C.elegans Biology [Internet]*, 2006.

T. Boveri. *Zellen-Studien: Die Befruchtung und Teilung des Eies von Ascaris megalocephala*. G. Fischer, 1888.

L. A. Boyer, T. I. Lee, M. F. Cole, S. E. Johnstone, S. S. Levine, J. P. Zucker, M. G. Guenther, R. M. Kumar, H. L. Murray, R. G. Jenner, D. K. Gifford, D. A. Melton, R. Jaenisch, and R. A. Young. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, 122 (6):947–956, 2005.

C. Brabin, P. J. Appleford, and A. Woollard. The *Caenorhabditis elegans* GATA factor ELT-1 works through the cell proliferation regulator BRO-1 and the Fusogen EFF-1 to maintain the seam stem-like fate. *PLoS genetics*, 7(8), 2011.

S. Brenner. The genetics of *Caenorhabditis elegans*. *Genetics*, 77(1):71–94, 1974.

S. Brenner. *My life in science*. Citeseer, 2001.

E. H. Bresnick, H.-Y. Lee, T. Fujiwara, K. D. Johnson, and S. Keles. Gata switches as developmental drivers. *Journal of Biological Chemistry*, 285(41):31087–31093, 2010.

A. Brożek, A. Ceccarelli, A. C. S. Jørgensen, M. Hintze, M. Barkoulas, and V. Shahrezaei. Inference of the core gene regulatory network underlying seam cell development in caenorhabditis elegans. *bioRxiv*, 2023.

K. Brunschwig, C. Wittmann, R. Schnabel, T. R. Bürglin, H. Tobler, and F. Müller. Anterior organization of the caenorhabditis elegans embryo by the labial-like hox gene ceh-13. *Development*, 126(7):1537–1546, 1999.

K. M. Cadigan and R. Nusse. Wnt signaling: a common theme in animal development. *Genes & development*, 11(24):3286–3305, 1997.

D. Calvo, M. Victor, F. Gay, G. Sui, M. P.-S. Luke, P. Dufourcq, G. Wen, M. Maduro, J. Rothman, and Y. Shi. A pop-1 repressor complex restricts inappropriate cell type-specific gene transcription during caenorhabditis elegans embryogenesis. *The EMBO journal*, 2001.

G. Casella and E. I. George. Explaining the gibbs sampler. *The American Statistician*, 46(3): 167–174, 1992.

G. Cassata, G. Shemer, P. Morandi, R. Donhauser, B. Podbilewicz, and R. Baumeister. *ceh-16*/engrailed patterns the embryonic epidermis of *Caenorhabditis elegans*. *Development*, 132(4):739–749, 2005.

J. F. d. Celis and M. Ruiz-Gómez. groucho and hedgehog regulate engrailed expression in the anterior compartment of the drosophila wing. *Development*, 121(10):3467–3476, 1995.

M. Chalfie, Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher. Green fluorescent protein as a marker for gene expression. *Science*, 263(5148):802–805, 1994.

I. Chambers and A. Smith. Self-renewal of teratocarcinoma and embryonic stem cells. *Oncogene*, 23(43):7150–7160, 2004.

M. Chambers, W. Turki-Judeh, M. W. Kim, K. Chen, S. D. Gallaher, and A. J. Courey. Mechanisms of groucho-mediated repression revealed by genome-wide analysis of groucho binding and activity. *BMC genomics*, 18:1–17, 2017.

A. J. Chang, N. Chronis, D. S. Karow, M. A. Marletta, and C. I. Bargmann. A distributed chemosensory circuit for oxygen preference in c. elegans. *PLOS Biology*, 4(9):1–15, 08 2006.

X. Chen, H. Xu, P. Yuan, F. Fang, M. Huss, V. B. Vega, E. Wong, Y. L. Orlov, W. Zhang, J. Jiang, Y.-H. Loh, H. C. Yeo, Z. X. Yeo, V. Narang, K. R. Govindarajan, B. Leong, A. Shahab, Y. Ruan, G. Bourque, W.-K. Sung, N. D. Clarke, C.-L. Wei, and H.-H. Ng. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*, 133(6):1106–1117, Jun 13 2008.

V. Chickarmane, C. Troein, U. A. Nuber, H. M. Sauro, and C. Peterson. Transcriptional dynamics of the embryonic stem cell switch. *PLoS computational biology*, 2(9):e123, 2006.

A. Chisholm. Control of cell fate in the tail region of c. elegans by the gene egl-5. *Development*, 111(4):921–932, 1991.

A. D. Chisholm and J. Hardin. Epidermal morphogenesis. *WormBook: The Online Review of C. elegans Biology [Internet]*, 2005.

A. D. Chisholm and T. I. Hsiao. The c. elegans epidermis as a model skin. *I: development*, 2012.

E. Cinnamon and Z. Paroush. Context-dependent regulation of groucho/tle-mediated repression. *Current opinion in genetics & development*, 18(5):435–440, 2008.

J. C. Claussen, J. Skiecevičienė, J. Wang, P. Rausch, T. H. Karlsen, W. Lieb, J. F. Baines, A. Franke, and M.-T. Hütt. Boolean analysis reveals systematic interactions among low-abundance species in the human gut microbiome. *PLoS computational biology*, 13(6): e1005361, 2017.

H. Clevers and R. Nusse. Wnt/β-catenin signaling and disease. *Cell*, 149(6):1192–1205, 2012.

M. L. Cohen, S. Kim, K. Morita, S. H. Kim, and M. Han. The gata factor *elt-1* regulates *C. elegans* developmental timing by promoting expression of the *let-7* family microrna. *PLoS Genet*, 11(3):e1005099, 2015.

D. W. Cowing and C. Kenyon. Expression of the homeotic gene mab-5 during caenorhabditis elegans embryogenesis. *Development*, 116(2):481–490, 1992.

E. Culetto and D. B. Sattelle. A role for caenorhabditis elegans in understanding the function and interactions of human disease genes. *Human molecular genetics*, 9(6):869–877, 2000.

M. I. Davidich and S. Bornholdt. Boolean network model predicts cell cycle sequence of fission yeast. *PloS one*, 3(2), 2008.

E. Davidson and M. Levin. Gene regulatory networks. *Proceedings of the National Academy of Sciences*, 102(14):4935, 2005.

E. H. Davidson. *The regulatory genome: gene regulatory networks in development and evolution*. Elsevier, 2010.

F. M. Delgado and F. Gómez-Vela. Computational methods for gene regulatory networks reconstruction and analysis: A review. *Artificial Intelligence in Medicine*, 95:133–145, 2019.

Y. Deng, H. Zenil, J. Tegnér, and N. A. Kiani. Hidi: an efficient reverse engineering schema for large-scale dynamic regulatory network reconstruction using adaptive differentiation. *Bioinformatics*, 33(24):3964–3972, 2017.

M. Di Bonito, J. C. Glover, and M. Studer. Hox genes and region-specific sensorimotor circuit formation in the hindbrain and spinal cord. *Developmental Dynamics*, 242(12):1348–1368, 2013.

M. J. Dunlop, R. S. C. III, J. H. Levine, R. M. Murray, and M. B. Elowitz. Regulatory activity revealed by dynamic correlations in gene expression noise. *Nature genetics*, 40(12): 1493–1498, 2008.

G. S. Eichler, S. Huang, and D. E. Ingber. Gene expression dynamics inspector (gedi): for integrative analysis of expression profiles. *Bioinformatics (Oxford, England)*, 19(17): 2321–2322, Nov 22 2003.

D. M. Eisenmann and S. K. Kim. Protruding vulva mutants identify novel loci and wnt signaling factors that function during *Caenorhabditis elegans* vulva development. *Genetics*, 156(3): 1097–1116, 2000.

R. Eker. The recessive mutant engrailed in drosophila melanogaster. *Hereditas*, 12(1-2): 217–222, 1929.

M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, 2002.

ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414):57, 2012.

J. J. Faith, B. Hayete, J. T. Thaden, I. Mogno, J. Wierzbowski, G. Cottarel, S. Kasif, J. J. Collins, and T. S. Gardner. Large-scale mapping and validation of escherichia coli transcriptional regulation from a compendium of expression profiles. *PLOS Biology*, 5(1):1–13, 01 2007.

E. H. Feinberg, M. K. VanHoven, A. Bendesky, G. Wang, R. D. Fetter, K. Shen, and C. I.

Bargmann. Gfp reconstitution across synaptic partners (grasp) defines cell contacts and synapses in living nervous systems. *Neuron*, 57(3):353–363, 2008.

A. E. Friedland, Y. B. Tzur, K. M. Esvelt, M. P. Colaiácovo, G. M. Church, and J. A. Calarco. Heritable genome editing in c. elegans via a crispr-cas9 system. *Nature methods*, 10(8): 741–743, 2013.

N. Friedman. Inferring cellular networks using probabilistic graphical models. *Science*, 303 (5659):799–805, 2004.

E. Fuchs. Scratching the surface of skin development. *Nature*, 445(7130):834–842, 2007.

T. Fujiwara. Gata transcription factors: basic principles and related human disorders. *The Tohoku journal of experimental medicine*, 242(2):83–91, 2017.

M.-A. Félix and A. Wagner. Robustness and evolution: concepts, insights and challenges from a developmental model system. *Heredity*, 100(2):132–140, 2008.

G. Gardner, di Bernardo di Bernardo, L. Lorenz, and C. Collins. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science*, 301(5629): 102–105, 2003.

T. S. Gardner and J. J. Faith. Reverse-engineering transcription control networks. *Physics of life reviews*, 2(1):65–88, 2005.

S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6): 721–741, 1984.

M. Genestine, L. Lin, M. Durens, Y. Yan, Y. Jiang, S. Prem, K. Bailoor, B. Kelly, P. K. Sonsalla, and P. G. Matteson. Engrailed-2 (en2) deletion produces multiple neurodevelopmental defects in monoamine systems, forebrain structures and neurogenesis and behavior. *Human molecular genetics*, 24(20):5805–5827, 2015.

J. S. Gilleard and J. D. McGhee. Activation of hypodermal differentiation in the caenorhabditis

elegans embryo by gata transcription factors elt-1 and elt-3. *Molecular and cellular biology*, 21(7):2533–2544, 2001.

D. A. Glauser. Temperature sensing and context-dependent thermal behavior in nematodes. *Current opinion in neurobiology*, 73:102525, 2022.

J. E. Gleason and D. M. Eisenmann. Wnt signaling controls the stem cell-like asymmetric division of the epithelial seam cells during c. elegans larval development. *Developmental biology*, 348(1):58–66, 2010.

F. González, S. Boué, and J. C. I. Belmonte. Methods for making induced pluripotent stem cells: reprogramming a la carte. *Nature Reviews Genetics*, 12(4):231–242, 2011.

B. C. Goodwin. Temporal organization in cells. a dynamic theory of cellular control processes. *Temporal organization in cells.A dynamic theory of cellular control processes.*, 1963.

L. Gorrepati, K. W. Thompson, and D. M. Eisenmann. *C. elegans* gata factors *egl-18* and *elt-6* function downstream of wnt signaling to maintain the progenitor fate during larval asymmetric divisions of the seam cells. *Development*, 140(10):2093–2102, 2013.

E. Haeckel. Schöpfungsgeschichte. *Leipzig. Vide also Translation. The History of Creation. King and Co., London*, 1876.

J. K. Hale. *Ordinary differential equations*. Courier Corporation, 2009.

D. H. Hall and R. L. Russell. The posterior nervous system of the nematode caenorhabditis elegans: serial reconstruction of identified neurons and complete pattern of synaptic interactions. *Journal of Neuroscience*, 11(1):1–22, 1991.

O. F. Harandi and V. R. Ambros. Control of stem cell self-renewal and differentiation by the heterochronic genes and the cellular asymmetry machinery in caenorhabditis elegans. *Proceedings of the National Academy of Sciences*, 112(3):E287–E296, 2015.

J. Harris, L. Honigberg, N. Robinson, and C. Kenyon. Neuronal cell migration in c. elegans: regulation of hox gene expression and cell position. *Development*, 122(10):3117–3131, 1996.

M. Harterink, D. H. Kim, T. C. Middelkoop, T. D. Doan, A. van Oudenaarden, and H. C. Korswagen. Neuroblast migration along the anteroposterior axis of c. elegans is controlled by opposing gradients of wnts and a secreted frizzled-related protein. *Development*, 138 (14):2915–2924, 2011.

D. I. Hastie and P. J. Green. Model choice using reversible jump markov chain monte carlo. *Statistica Neerlandica*, 66(3):309–338, 2012.

W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. 1970.

R. Heinrich and T. A. Rapoport. A linear steady-state treatment of enzymatic chains: general properties, control and effector strength. *European journal of biochemistry*, 42(1):89–95, 1974.

M. A. Herman. Wnt signaling in c. elegans. *Wnt signaling in Development*, pages 187–212, 2003.

A. V. Hill. The possible effects of the aggregation of the molecules of hemoglobin on its dissociation curves. *j. physiol.*, 40:iv–vii, 1910.

M. Hintze, S. L. Koneru, S. P. Gilbert, D. Katsanos, J. Lambert, and M. Barkoulas. A cell fate switch in the *Caenorhabditis elegans* seam cell lineage occurs through modulation of the wnt asymmetry pathway in response to temperature increase. *Genetics*, 214(4):927–939, 2020.

M. Hintze, D. Katsanos, V. Shahrezaei, and M. Barkoulas. Phenotypic robustness of epidermal stem cell number in c. elegans is modulated by the activity of the conserved n-acetyltransferase *nath-10*/nat10. *Frontiers in Cell and Developmental Biology*, 9:640856, 2021.

D. Hirsh and R. Vanderslice. Temperature-sensitive developmental mutants of caenorhabditis elegans. *Developmental biology*, 49(1):220–235, 1976.

E. F. Hoier, W. A. Mohler, S. K. Kim, and A. Hajnal. The caenorhabditis elegans apc-related gene *apr-1* is required for epithelial cell migration and hox gene expression. *Genes & development*, 14(7):874–886, 2000.

M. Huang and M. Chalfie. Gene interactions affecting mechanosensory transduction in caenorhabditis elegans. *Nature*, 367(6462):467–470, 1994.

S. Huang, P. Shetty, S. M. Robertson, and R. Lin. Binary cell fate specification during c. elegans embryogenesis driven by reiterated reciprocal asymmetry of tcf pop-1 and its coactivatorβ-catenin sys-1. *Development*, 134(14):2685–2695, 2007.

X. Huang, E. Tian, Y. Xu, and H. Zhang. The *C. elegans* engrailed homolog *ceh-16*regulates the self-renewal expansion division of stem cell-like seam cells. *Developmental biology*, 333(2):337–347, 2009.

J. Huelsken and J. Behrens. The wnt signalling pathway. *Journal of cell science*, 115(21): 3977–3978, 2002.

S. Hughes, C. Brabin, P. J. Appleford, and A. Woollard. CEH-20/Pbx and UNC-62/Meis function upstream of *rnt-1*/Runx to regulate asymmetric divisions of the *C. elegans* stem-like seam cells. *Biology open*, 2(7):718–727, 2013.

T. R. Hughes, M. J. Marton, A. R. Jones, C. J. Roberts, R. Stoughton, C. D. Armour, H. A. Bennett, E. Coffey, H. Dai, Y. D. He, M. J. Kidd, A. M. King, M. R. Meyer, D. Slade, P. Y. Lum, S. B. Stepaniants, D. D. Shoemaker, D. Gachotte, K. Chakraburtty, J. Simon, M. Bard, and S. H. Friend. Functional discovery via a compendium of expression profiles. *Cell*, 102 (1):109–126, 2000.

M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier. A programmable dual-rna–guided dna endonuclease in adaptive bacterial immunity. *science*, 337(6096):816–821, 2012.

A. John, S. T. Smith, and J. B. Jaynes. Inserting the ftz homeodomain into engrailed creates a dominant transcriptional repressor that specifically turns off ftz target genes in vivo.

Development, 121(6):1801–1813, 1995.

A. C. S. Jørgensen, M. Sturrock, A. Ghosh, and V. Shahrezaei. Bayesian model discovery for reverse-engineering biochemical networks from data. *bioRxiv*, 2023.

P. M. Joshi, M. R. Riddle, N. J. Djabrayan, and J. H. Rothman. *Caenorhabditis elegans* as a model for stem cell biology. *Developmental dynamics: an official publication of the American Association of Anatomists*, 239(5):1539–1554, 2010.

D. Jupiter, H. Chen, and V. VanBuren. Star net 2: a web-based tool for accelerating discovery of gene regulatory networks using microarray co-expression data. *BMC bioinformatics*, 10 (1):1–7, 2009.

H. Kacser. The control of flux. In *Symp Soc Exp Biol*, volume 27, page 65, 1973.

H. Kagoshima, R. Nimmo, N. Saad, J. Tanaka, Y. Miwa, S. Mitani, Y. Kohara, and A. Woollard. The *C. elegans* CBFÎ² homologue BRO-1 interacts with the Runx factor, RNT-1, to promote stem cell proliferation and self-renewal. *Development*, 134(21):3905–3915, 2007.

T. Kaletta and M. O. Hengartner. Finding function in novel targets: C. elegans as a model organism. *Nature reviews Drug discovery*, 5(5):387–399, 2006.

R. S. Kamath and J. Ahringer. Genome-wide rnai screening in caenorhabditis elegans. *Methods*, 30(4):313–321, 2003.

R. S. Kamath, A. G. Fraser, Y. Dong, G. Poulin, R. Durbin, M. Gotta, A. Kanapin, N. Le Bot, S. Moreno, M. Sohrmann, et al. Systematic functional analysis of the caenorhabditis elegans genome using rnai. *Nature*, 421(6920):231–237, 2003.

D. Katsanos. *Identification of transcription factor targets, gene expression profiles and accessible chromatin regions in the Caenorhabditis elegans epidermis using targeted DamID*. PhD thesis, Imperial College London, 2021.

D. Katsanos and M. Barkoulas. Targeted DamID in *C. elegans* reveals a direct role for LIN-22 and NHR-25 in antagonizing the epidermal stem cell fate. *Science advances*, 8(5): eabk3141, 2022.

D. Katsanos, S. L. Koneru, L. M. Boukhibar, N. Gritti, R. Ghose, P. J. Appleford, M. Doitsidou, A. Woollard, J. S. van Zon, R. J. Poole, and M. Barkoulas. Stochastic loss and gain of symmetric divisions in the *C. elegans* epidermis perturbs robustness of stem cell number. *Plos Biology*, 15(11):e2002429, 2017.

S. Kauffman. Homeostasis and differentiation in random genetic control networks. *Nature*, 224(5215):177–178, 1969.

S. A. Kauffman. Behaviour of randomly constructed genetic nets: binary element nets. *Towards a Theoretical Biology, ed.CH Waddington (Aldine, Chicago, 1970) pp*, pages 18–37, 1970.

W. Keil, L. M. Kutscher, S. Shaham, and E. D. Siggia. Long-term high-resolution imaging of developing c. elegans larvae with microfluidics. *Developmental cell*, 40(2):202–214, 2017.

C. Kenyon. A gene involved in the development of the posterior body region of c. elegans. *Cell*, 46(3):477–487, 1986.

B. N. Kholodenko, O. V. Demin, G. Moehren, and J. B. Hoek. Quantification of short term signaling by the epidermal growth factor receptor. *Journal of Biological Chemistry*, 274(42): 30169–30181, 1999.

B. N. Kholodenko, A. Kiyatkin, F. J. Bruggeman, E. Sontag, H. V. Westerhoff, and J. B. Hoek. Untangling the wires: a strategy to trace functional interactions in signaling and gene networks. *Proceedings of the National Academy of Sciences*, 99(20):12841–12846, 2002.

P. Khosravi, V. H. Gazestani, L. Pirhaji, B. Law, M. Sadeghi, B. Goliaei, and G. D. Bader. Inferring interaction type in gene regulatory networks using co-expression data. *Algorithms for molecular biology*, 10(1):1–11, 2015.

A. R. Kidd, J. A. Miskowski, K. R. Siegfried, H. Sawa, and J. Kimble. A β-catenin identified by functional rather than sequence criteria and its role in wnt/mapk signaling. *Cell*, 121(5): 761–772, 2005.

J. Kim, J. Chu, X. Shen, J. Wang, and S. H. Orkin. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell*, 132(6):1049–1061, 2008.

B. Kinney, S. Sahu, N. Stec, K. Hills-Muckey, D. W. Adams, J. Wang, M. Jaremko, L. Joshua-Tor, W. Keil, and C. M. Hammell. A circadian-like gene network programs the timing and dosage of heterochronic mirna transcription during c. elegans development. *Developmental Cell*, 58(22):2563–2579, 2023.

B. Klinger and N. Blüthgen. Reverse engineering gene regulatory networks by modular response analysis–a benchmark. *Essays in Biochemistry*, 62(4):535–547, 2018.

K. Koh and J. H. Rothman. ELT-5 and ELT-6 are required continuously to regulate epidermal seam cell differentiation and cell fusion in *C. elegans*. *Development*, 128(15):2867–2880, 2001.

K. Koh, S. M. Peyrot, C. G. Wood, J. A. Wagmaister, M. F. Maduro, D. M. Eisenmann, and J. H. Rothman. Cell fates and fusion in the *C. elegans* vulval primordium are regulated by the *egl-18* and ELT-6 GATA factors are apparent direct targets of the LIN-39 Hox protein. *Development*, 129(22):5171–5180, 2002.

S. L. Koneru, M. Hintze, D. Katsanos, and M. Barkoulas. Cryptic genetic variation in a heat shock protein modifies the outcome of a mutation affecting epidermal stem cell development in c. elegans. *Nature Communications*, 12(1):3263, 2021a.

S. L. Koneru, F. X. Quah, R. Ghose, M. Hintze, N. Gritti, J. S. van Zon, and M. Barkoulas. A role for the fusogen *eff-1* in epidermal stem cell number robustness in *Caenorhabditis elegans*. *Scientific reports*, 11(1):1–13, 2021b.

H. C. Korswagen. Canonical and non-canonical wnt signaling pathways in caenorhabditis elegans: Variations on a common signaling theme. *Bioessays*, 24(9):801–810, 2002.

D. E. Koshland Jr, G. Némethy, and D. Filmer. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry*, 5(1):365–385, 1966.

S. Kramer, E. Meyer-Natus, C. Stigloher, H. Thoma, A. Schnaufer, and M. Engstler. Parallel monitoring of RNA abundance, localization and compactness with correlative single molecule FISH on LR White embedded samples. *Nucleic acids research*, 49(3):e14, 2021.

R. Krumlauf. Hox genes in vertebrate development. *Cell*, 78(2):191–201, 1994. ISSN 0092-8674.

M. M. Kudron, A. Victorsen, L. Gevirtzman, L. W. Hillier, W. W. Fisher, D. Vafeados, M. Kirkey, A. S. Hammonds, J. Gersch, and H. Ammouri. The ModERN resource: genome-wide binding profiles for hundreds of Drosophila and *Caenorhabditis elegans* transcription factors. *Genetics*, 208(3):937–949, 2018.

C. Lanctôt, C. Kaspar, and T. Cremer. Positioning of the mouse hox gene clusters in the nuclei of developing embryos and differentiating embryoid bodies. *Experimental cell research*, 313(7):1449–1459, 2007.

F. Landmann, S. Quintin, and M. Labouesse. Multiple regulatory elements with spatially and temporally distinct activities control the expression of the epithelial differentiation gene *lin-26* in c. elegans. *Developmental biology*, 265(2):478–490, 2004.

C. Lee, H. Shin, and J. Kimble. Dynamics of notch-dependent transcriptional bursting in its native context. *Developmental Cell*, 50(4):426–435, 2019.

S. F. Levy and M. L. Siegal. Network hubs buffer environmental variation in saccharomyces cerevisiae. *PLoS biology*, 6(11):e264, 2008.

E. B. Lewis. A gene complex controlling segmentation in drosophila. *Nature*, 276(5688): 565–570, 1978.

S. Liang, S. Fuhrman, and R. Somogyi. Reveal, a general reverse engineering algorithm for inference of genetic network architectures. 1998.

J. Liu and J. I. Murray. Mechanisms of lineage specification in caenorhabditis elegans. *Genetics*, 225(4):iyad174, 2023.

M.-C. Lo, F. Gay, R. Odom, Y. Shi, and R. Lin. Phosphorylation by the β-catenin/mapk complex promotes 14-3-3-mediated nuclear export of tcf/pop-1 in signal-responsive cells in c. elegans. *Cell*, 117(1):95–106, 2004.

Y.-H. Loh, Q. Wu, J.-L. Chew, V. B. Vega, W. Zhang, X. Chen, G. Bourque, J. George, B. Leong, J. Liu, K.-Y. Wong, K. W. Sung, C. W. H. Lee, X.-D. Zhao, K.-P. Chiu, L. Lipovich, V. A. Kuznetsov, P. Robson, L. W. Stanton, C.-L. Wei, Y. Ruan, B. Lim, and H.-H. Ng. The oct4 and nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature genetics*, 38(4):431–440, 2006.

C. A. Loomis, E. Harris, J. Michaud, W. Wurst, M. Hanks, and A. L. Joyner. The mouse engrailed-1 gene and ventral limb patterning. *Nature*, 382(6589):360–363, 1996.

Y. Luo, B. C. Hitz, I. Gabdank, J. A. Hilton, M. S. Kagda, B. Lam, Z. Myers, P. Sud, J. Jou, and K. Lin. New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic acids research*, 48(D1):D882–D889, 2020.

Y. Maki, D. Tominaga, M. Okamoto, S. Watanabe, and Y. Eguchi. *Development of a system for the inference of large scale genetic networks*, pages 446–458. Biocomputing 2001. World Scientific, 2000.

S. A. Malekpour, A. Alizad-Rahvar, and M. Sadeghi. Logicnet: probabilistic continuous logics in reconstructing gene regulatory networks. *BMC Bioinformatics*, 21(1):318, 2020.

H. Mallick and N. Yi. A new Bayesian lasso. *Statistics and its Interface*, 7(4):571–582, 2014. ISSN 19387997.

J. N. Maloof and C. Kenyon. The hox gene lin-39 is required during c. elegans vulval induction to select the outcome of ras signaling. *Development*, 125(2):181–190, 1998.

J. N. Maloof, J. Whangbo, J. M. Harris, G. D. Jongeward, and C. Kenyon. A wnt signaling pathway controls hox gene expression and neuroblast migration in c. elegans. *Development*, 126(1):37–49, 1999.

A. A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R. D. Favera, and A. Califano. Aracne: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, 7(1):S7, 2006.

H. Markel, J. Chandler, and W. Werr. Translational fusions with the engrailed repressor domain efficiently convert plant transcription factors into dominant-negative functions. *Nucleic acids research*, 30(21):4709–4719, 2002.

O. J. Marshall, T. D. Southall, S. W. Cheetham, and A. H. Brand. Cell-type-specific profiling of protein–DNA interactions without cell isolation using targeted DamID with next-generation sequencing. *Nature protocols*, 11(9):1586–1598, 2016.

P. Martinez and C. T. Amemiya. Genomics of the hox gene cluster. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 133(4):571–580, 2002. ISSN 1096-4959.

I. Masse, L. Barbollat-Boutrand, M. E. Kharbili, O. Berthier-Vergnes, D. Aubert, and J. Lamartine. Gata3 inhibits proliferation and induces expression of both early and late differentiation markers in keratinocytes of the human epidermis. *Archives of Dermatological Research*, 306:201–208, 2014.

H. Matsumoto, H. Kiryu, C. Furusawa, M. S. Ko, S. B. Ko, N. Gouda, T. Hayashi, and I. Nikaido. Scode: an efficient regulatory network inference algorithm from single-cell rna-seq during differentiation. *Bioinformatics*, 33(15):2314–2321, 2017.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6): 1087–1092, 1953.

K. Mitsui, Y. Tokuzawa, H. Itoh, K. Segawa, M. Murakami, K. Takahashi, M. Maruyama, M. Maeda, and S. Yamanaka. The homeoprotein nanog is required for maintenance of pluripotency in mouse epiblast and es cells. *Cell*, 113(5):631–642, 2003.

K. Mizumoto and H. Sawa. Two βs or not two βs: regulation of asymmetric division by

β-catenin. *Trends in Cell Biology*, 17(10):465–473, 2007a. ISSN 0962-8924.

K. Mizumoto and H. Sawa. Two βs or not two βs: regulation of asymmetric division by β-catenin. *Trends in cell biology*, 17(10):465–473, 2007b.

M. Mojtahedi, A. Skupin, J. Zhou, I. G. Castaño, R. Leong-Quong, H. Chang, K. Trachana, A. Giuliani, and S. Huang. Cell fate decision as high-dimensional critical state transition. *PLOS Biology*, 14(12):e2000640, 2016.

J. Monod, J. Wyman, and J.-P. Changeux. On the nature of allosteric transitions: a plausible model. *Journal of molecular biology*, 12(1):88–118, 1965.

J. D. Montenegro. *Gene co-expression network analysis*, pages 387–404. Plant bioinformatics: methods and protocols. Springer, 2022.

M. K. Montgomery, S. Xu, and A. Fire. Rna as a target of double-stranded rna-mediated genetic interference in caenorhabditis elegans. *Proceedings of the national academy of sciences*, 95(26):15502–15507, 1998.

J. M. Mooij, J. Peters, D. Janzing, J. Zscheischler, and B. Schölkopf. Distinguishing cause from effect using observational data: methods and benchmarks. *Journal of Machine Learning Research*, 17(32):1–102, 2016.

R. Moreno-Sánchez, E. Saavedra, S. Rodríguez-Enríquez, V. Olín-Sandoval, et al. Metabolic control analysis: a tool for designing strategies to manipulate metabolic pathways. *BioMed Research International*, 2008, 2008.

N. Moris, C. Pina, and A. M. Arias. Transition states and cell fate decisions in epigenetic landscapes. *Nature Reviews Genetics*, 17(11):693–703, 2016.

B. Mugat, M.-L. Parmentier, N. Bonneaud, H. Y. E. Chan, and F. Maschat. Protective role of engrailed in a drosophila model of huntington's disease. *Human molecular genetics*, 17 (22):3601–3616, 2008.

B. Munsky, B. Trinh, and M. Khammash. Listening to the noise: random fluctuations reveal gene network parameters. *Molecular systems biology*, 5(1):318, 2009.

T. Nakamura, F. Hamada, T. Ishidate, K.-i. Anai, K. Kawahara, K. Toyoshima, and T. Akiyama. Axin, an inhibitor of the wnt signalling pathway, interacts with β-catenin, gsk-3β and apc and reduces the β-catenin level. *Genes to Cells*, 3(6):395–403, 1998.

J. Narula, C. J. Williams, A. Tiwari, J. Marks-Bluth, J. E. Pimanda, and O. A. Igoshin. Mathematical model of a gene regulatory network reconciles effects of genetic perturbations on hematopoietic stem cell emergence. *Developmental biology*, 379(2):258–269, 2013.

P. Navarro, N. Festuccia, D. Colby, A. Gagliardi, N. P. Mullin, W. Zhang, V. Karwacki-Neisius, R. Osorno, D. Kelly, M. Robertson, et al. Oct4/sox2-independent nanog autorepression modulates heterogeneous nanog gene expression in mouse es cells. *The EMBO journal*, 31(24):4547–4562, 2012.

J. Nichols, B. Zevnik, K. Anastassiadis, H. Niwa, D. Klewe-Nebenius, I. Chambers, H. Schöler, and A. Smith. Formation of pluripotent stem cells in the mammalian embryo depends on the pou transcription factor oct4. *Cell*, 95(3):379–391, 1998.

R. Nimmo, A. Antebi, and A. Woollard. mab-2 encodes rnt-1, a c. elegans runx homologue essential for controlling cell proliferation in a stem cell-like developmental lineage. 2005.

S. Okawa, S. Nicklas, S. Zickenrott, J. C. Schwamborn, and A. Del Sol. A generalized gene-regulatory network model of stem cell differentiation for predicting lineage specifiers. *Stem cell reports*, 7(3):307–315, 2016.

K. Okita, T. Ichisaka, and S. Yamanaka. Generation of germline-competent induced pluripotent stem cells. *Nature*, 448(7151):313–317, 2007.

C. Olsen, P. E. Meyer, and G. Bontempi. On the impact of entropy estimation on transcriptional regulatory network inference based on mutual information. *EURASIP Journal on Bioinformatics and Systems Biology*, 2009:1–9, 2008.

B. D. Page, W. Zhang, K. Steward, T. Blumenthal, and J. R. Priess. Elt-1, a gata-like transcription factor, is required for epidermal cell fates in caenorhabditis elegans embryos. *Genes and Development*, 11(13):1651–1661, 1997a.

B. D. Page, W. Zhang, K. Steward, T. Blumenthal, and J. R. Priess. ELT-1, a GATA-like transcription factor, is required for epidermal cell fates in *Caenorhabditis elegans* embryos. *Genes I& development*, 11(13):1651–1661, 1997b.

A. M. Pani and B. Goldstein. Direct visualization of a native wnt in vivo reveals that a long-range wnt gradient forms by extracellular dispersal. *Elife*, 7:e38325, 2018.

E. S. Pearson. Bayes' theorem, examined in the light of experimental sampling. *Biometrika*, pages 388–442, 1925.

I. S. Peter and E. H. Davidson. Evolution of gene regulatory networks controlling body plan development. *Cell*, 144(6):970–985, 2011.

K. Petratou, T. Subkhankulova, J. A. Lister, A. Rocco, H. Schwetlick, and R. N. Kelsh. A systems biology approach uncovers the core gene regulatory network governing iridophore fate choice from the neural crest. *PLoS genetics*, 14(10):e1007402, 2018.

B. Podbilewicz, E. Leikina, A. Sapir, C. Valansi, M. Suissa, G. Shemer, and L. V. Chernomordik. The *C. elegans* developmental fusogen EFF-1 mediates homotypic fusion in heterologous cells and in vivo. *Developmental cell*, 11(4):471–481, 2006.

H. D. Preisler and S. Kauffman. A proposal regarding the mechanism which underlies lineage choice during hematopoietic differentiation. *Leukemia research*, 23(8):685–694, 1999.

X. Qiu, A. Rahimzamani, L. Wang, B. Ren, Q. Mao, T. Durham, J. L. McFaline-Figueroa, L. Saunders, C. Trapnell, and S. Kannan. Inferring causal gene regulatory networks from coupled single-cell expression dynamics using scribe. *Cell systems*, 10(3):265–274.e11, Mar 25 2020.

A. Raj, P. V. D. Bogaard, S. A. Rifkin, A. V. Oudenaarden, and S. Tyagi. Imaging individual mRNA molecules using multiple singly labeled probes. *Nature methods*, 5(10):877–879, 2008.

A. Raj, S. A. Rifkin, E. Andersen, and A. V. Oudenaarden. Variability in gene expression underlies incomplete penetrance. *Nature*, 463(7283):913–918, 2010.

R. Ram and M. Chetty. Mcmc based bayesian inference for modeling gene networks. In *Pattern Recognition in Bioinformatics: 4th IAPR International Conference, PRIB 2009, Sheffield, UK, September 7-9, 2009. Proceedings 4*, pages 293–306. Springer, 2009.

A. Rao, A. O. H. III, D. J. States, and J. D. Engel. *Using directed information to build biologically relevant influence networks*, pages 145–156. Computational Systems Bioinformatics: (Volume 6). World Scientific, 2007.

H. Rekaik, F.-X. B. de Thé, A. Prochiantz, J. Fuchs, and R. L. Joshi. Dissecting the role of engrailed in adult dopaminergic neurons–insights into parkinson disease pathogenesis. *FEBS letters*, 589(24):3786–3794, 2015.

H. Ren and H. Zhang. Wnt signaling controls temporal identities of seam cells in caenorhabditis elegans. *Developmental Biology*, 345(2):144–155, 2010. ISSN 0012-1606.

Q. Ren, J. Chen, and Y. Liu. Lrp5 and lrp6 in wnt signaling: similarity and divergence. *Frontiers in cell and developmental biology*, 9:670960, 2021.

I. Rissling, K. Strauch, C. Höft, W. H. Oertel, and J. C. Möller. Haplotype analysis of the engrailed-2 gene in young-onset parkinson's disease. *Neurodegenerative Diseases*, 6(3): 102–105, 2009.

C. E. Rocheleau, J. Yasuda, T. H. Shin, R. Lin, H. Sawa, H. Okano, J. R. Priess, R. J. Davis, and C. C. Mello. Wrm-1 activates the lit-1 protein kinase to transduce anterior/posterior polarity signals in c. elegans. *Cell*, 97(6):717–726, 1999.

D. J. Rodda, J.-L. Chew, L.-H. Lim, Y.-H. Loh, B. Wang, H.-H. Ng, and P. Robson. Transcriptional regulation of nanog by oct4 and sox2 *. *Journal of Biological Chemistry*, 280(26): 24731–24737, 2005.

K. Sachs, O. Perez, D. Pe'er, D. A. Lauffenburger, and G. P. Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721):523–529, 2005.

A. M. Saffer, D. H. Kim, A. van Oudenaarden, and H. R. Horvitz. The caenorhabditis elegans

synthetic multivulva genes prevent ras pathway activation by tightly repressing global ectopic expression of lin-3 egf. *PLoS Genetics*, 7(12):e1002418, 2011.

T. Santra, W. Kolch, and B. N. Kholodenko. Integrating bayesian variable selection with modular response analysis to infer biochemical network topology. *BMC systems biology*, 7: 1–19, 2013.

T. Santra, O. Rukhlenko, V. Zhernovkov, and B. N. Kholodenko. Reconstructing static and dynamic models of signaling pathways using modular response analysis. *Current Opinion in Systems Biology*, 9:11–21, 2018.

D. Sarmah, G. R. Smith, M. Bouhaddou, A. D. Stern, J. Erskine, and M. R. Birtwistle. Network inference from perturbation time course data. *NPJ Systems Biology and Applications*, 8(1): 42, 2022.

B. Sauer. Functional expression of the cre-lox site-specific recombination system in the yeast saccharomyces cerevisiae. *Molecular and cellular biology*, 1987.

B. Sauer and N. Henderson. Site-specific dna recombination in mammalian cells by the cre recombinase of bacteriophage p1. *Proceedings of the National Academy of Sciences*, 85 (14):5166–5170, 1988.

W. R. Schafer. Mechanosensory molecules and circuits in c. elegans. *Pflügers Archiv-European Journal of Physiology*, 467:39–48, 2015.

J. Schäfer and K. Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. 4(1), 2005.

N. Serrano and F. Maschat. Molecular mechanism of polyhomeotic activation by engrailed. *The EMBO journal*, 17(13):3704–3713; 3713, 1998.

V. Shahrezaei and P. S. Swain. The stochastic nature of biochemical networks. *Current opinion in biotechnology*, 19(4):369–374, 2008.

A. K. Shalek, R. Satija, X. Adiconis, R. S. Gertner, J. T. Gaublomme, R. Raychowdhury,

S. Schwartz, N. Yosef, C. Malboeuf, D. Lu, J. J. Trombetta, D. Gennert, A. Gnirke, A. Goren, N. Hacohen, J. Z. Levin, H. Park, and A. Regev. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*, 498(7453):236–240, 2013.

J. Shao, K. He, H. Wang, W. S. Ho, X. Ren, X. An, M. K. Wong, B. Yan, D. Xie, J. Stamatoyannopoulos, and Z. Zhao. Collaborative regulation of development but independent control of metabolism by two epidermis-specific transcription factors in caenorhabditis elegans*. *Journal of Biological Chemistry*, 288(46):33411–33426, 2013.

D. D. Shaye and I. Greenwald. Ortholist: a compendium of c. elegans genes with human orthologs. *PloS one*, 6(5):e20085, 2011.

P. Shetty, M.-C. Lo, S. M. Robertson, and R. Lin. C. elegans tcf protein, pop-1, converts from repressor to activator as a result of wnt-induced lowering of nuclear levels. *Developmental biology*, 285(2):584–592, 2005.

I. Shmulevich, E. R. Dougherty, and W. Zhang. From boolean to probabilistic boolean networks as models of genetic regulatory networks. *Proceedings of the IEEE*, 90(11): 1778–1792, 2002.

H. H. Simon, H. Saueressig, W. Wurst, M. D. Goulding, and D. D. M. O'Leary. Fate of midbrain dopaminergic neurons controlled by the engrailed genes. *The Journal of Neuroscience*, 21 (9):3126–3134, 2001.

M. C. Simon, L. Pevny, M. V. Wiles, G. Keller, F. Costantini, and S. H. Orkin. Rescue of erythroid development in gene targeted gata–1 mouse embryonic stem cells. *Nature genetics*, 1(2):92–98, 1992.

J. A. Smith, P. McGarr, and J. S. Gilleard. The *Caenorhabditis elegans* GATA factor *elt-1* is essential for differentiation and maintenance of hypodermal seam cells and for normal locomotion. *Journal of cell science*, 118(24):5709–5719, 2005.

J. J. Smith and P. Kratsios. Hox gene functions in the c. elegans nervous system: From early patterning to maintenance of neuronal identity. In *Seminars in Cell & Developmental*

Biology, volume 152, pages 58–69. Elsevier, 2024.

S. Y. Sokol. Analysis of dishevelled signalling pathways during xenopus development. *Current Biology*, 6(11):1456–1467, 1996.

J. Spieth, Y. H. Shim, K. Lea, R. Conrad, and T. Blumenthal. *elt-1*, an embryonically expressed *Caenorhabditis elegans* gene homologous to the GATA transcription factor family. *Molecular and cellular biology*, 11(9):4651–4659, 1991.

M. P. Stumpf. Inferring better gene regulation networks from single-cell data. *Current Opinion in Systems Biology*, 27:100342, 2021.

J. Sulston and J. White. Regulation and cell autonomy during postembryonic development of caenorhabditis elegans. *Developmental biology*, 78(2):577–597, 1980.

J. E. Sulston and H. R. Horvitz. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Developmental biology*, 56(1):110–156, 1977.

J. E. Sulston, E. Schierenberg, J. G. White, and J. N. Thomson. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Developmental biology*, 100(1):64–119, 1983.

N. Suzuki, A. Kanai, Y. Suzuki, H. Ogino, and H. Ochi. Adrenergic receptor signaling induced by klf15, a regulator of regeneration enhancer, promotes kidney reconstruction. *Proceedings of the National Academy of Sciences*, 119(33):e2204338119, 2022.

M. Sáez, R. Blassberg, E. Camacho-Aguilar, E. D. Siggia, D. A. Rand, and J. Briscoe. Statistically derived geometrical landscapes capture principles of decision-making dynamics during cell fate transitions. *Cell Systems*, 13(1):12–28. e3, 2022.

I. Tabus, J. Rissanen, and J. Astola. A classifier based on normalized maximum likelihood model for classes of boolean regression models. In *2002 11th European Signal Processing Conference*, pages 1–4. IEEE, 2002.

K. Takahashi and S. Yamanaka. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126(4):663–676, 2006.

F. Tang, C. Barbacioru, Y. Wang, E. Nordman, C. Lee, N. Xu, X. Wang, J. Bodeau, B. B. Tuch, A. Siddiqui, et al. mrna-seq whole-transcriptome analysis of a single cell. *Nature methods*, 6(5):377–382, 2009.

S. Tao, D. Tang, Y. Morita, T. Sperka, O. Omrani, A. Lechel, V. Sakk, J. Kraus, H. A. Kestler, M. Kühl, et al. Wnt activity and basal niche position sensitize intestinal stem and progenitor cells to dna damage. *The EMBO journal*, 34(5):624–640, 2015.

S. Theresa. Maintenance of c. elegans. *WormBook, February*, 11, 2006.

K. W. Thompson, P. Joshi, J. S. Dymond, L. Gorrepati, H. E. Smith, M. W. Krause, and D. M. Eisenmann. The paired-box protein pax-3 regulates the choice between lateral and ventral epidermal cell fates in *C. elegans*. *Developmental biology*, 412(2):191–207, 2016.

C. J. Thorpe, A. Schlesinger, and B. Bowerman. Wnt signalling in caenorhabditis elegans: regulating repressors and polarizing the cytoskeleton. *Trends in cell biology*, 10(1):10–17, 2000.

B. Tihanyi, T. Vellai, Á. Regős, E. Ari, F. Müller, and K. Takács-Vellai. The c. elegans hox gene ceh-13 regulates cell migration and fusion in a non-colinear way. implications for the early evolution of hox clusters. *BMC developmental biology*, 10:1–14, 2010.

L. Timmons and A. Fire. Specific interference by ingested dsrna. *Nature*, 395(6705):854, 1998.

C. Tocchini and S. E. Mango. An adapted ms2-mcp system to visualize endogenous cytoplasmic mrna with live imaging in caenorhabditis elegans. *PLoS biology*, 22(3):e3002526, 2024.

E. Tutucci, M. Vera, J. Biswas, J. Garcia, R. Parker, and R. H. Singer. An improved ms2 system for accurate reporting of the mrna life cycle. *Nature methods*, 15(1):81–89, 2018.

K. Van Auken, D. C. Weaver, L. G. Edgar, and W. B. Wood. Caenorhabditis elegans embryonic axial patterning requires two recently discovered posterior-group hox genes. *Proceedings of the National Academy of Sciences*, 97(9):4499–4503, 2000.

S. E. van der Horst, J. Cravo, A. Woollard, J. Teapal, and S. van den Heuvel. *C. elegans* runx/cbfβ suppresses pop-1 tcf to convert asymmetric to proliferative division of stem cell-like seam cells. *Development*, 146(22):dev180034, 2019.

L. Verny, N. Sella, S. Affeldt, P. P. Singh, and H. Isambert. Learning causal networks with latent variables from multivariate information in genomic data. *PLoS computational biology*, 13(10):e1005662, 2017.

R. S. Viger, S. M. Guittot, M. Anttonen, D. B. Wilson, and M. Heikinheimo. Role of the gata family of transcription factors in endocrine development, function, and disease. *Molecular Endocrinology*, 22(4):781–798, 2008.

D. Vipin, L. Wang, G. Devailly, T. Michoel, and A. Joshi. Causal transcription regulatory network inference using enhancer activity as a causal anchor. *International Journal of Molecular Sciences*, 19(11):3609, 2018.

C. H. Waddington. Canalization of development and the inheritance of acquired characters. *Nature*, 150(3811):563–565, 1942.

C. H. Waddington. *The strategy of the genes*. Routledge, 1957.

C. H. Waddington. *Genetic assimilation*, volume 10 of *Advances in genetics*, pages 257–293. Elsevier, 1961.

J. Wei, X. Hu, X. Zou, and T. Tian. Inference of genetic regulatory network for stem cell using single cells expression data. In *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 217–222, 2016. ISBN null.

J. Whangbo and C. Kenyon. A wnt signaling system that specifies two patterns of cell migration in c. elegans. *Molecular cell*, 4(5):851–858, 1999.

J. G. White, E. Southgate, J. N. Thomson, S. Brenner, et al. The structure of the nervous system of the nematode caenorhabditis elegans. *Philos Trans R Soc Lond B Biol Sci*, 314 (1165):1–340, 1986.

M. Wildwater, N. Sander, G. de Vreede, and S. van den Heuvel. Cell shape and wnt signaling redundantly control the division axis of c. elegans epithelial stem cells. *Development*, 138 (20):4375–4385, 2011.

A. R. Winnier, J. Y.-J. Meir, J. M. Ross, N. Tavernarakis, M. Driscoll, T. Ishihara, I. Katsura, and D. M. Miller. Unc-4/unc-37-dependent repression of motor neuron-specific genes controls synaptic choice in caenorhabditis elegans. *Genes & development*, 13(21):2774–2786, 1999.

Y. Xiao. A tutorial on analysis and simulation of boolean gene regulatory network models. *Current Genomics*, 10(7):511–525, 2009.

Y. Xie, Y. Liu, and W. Valdar. Joint estimation of multiple dependent gaussian graphical models with applications to mouse genomics. *Biometrika*, 103(3):493–511, 2016.

M. Zecca, K. Basler, and G. Struhl. Direct and long-range action of a wingless morphogen gradient. *Cell*, 87(5):833–844, 1996.

J. Zhang, R. E. Campbell, A. Y. Ting, and R. Y. Tsien. Creating new fluorescent probes for cell biology. *Nature Reviews Molecular Cell Biology*, 3(12):906–918, 2002. doi: $10.1038/\mathrm{nrm976}$. URL `https://doi.org/10.1038/nrm976`. ID: Zhang2002.

Chapter 7

Appendix

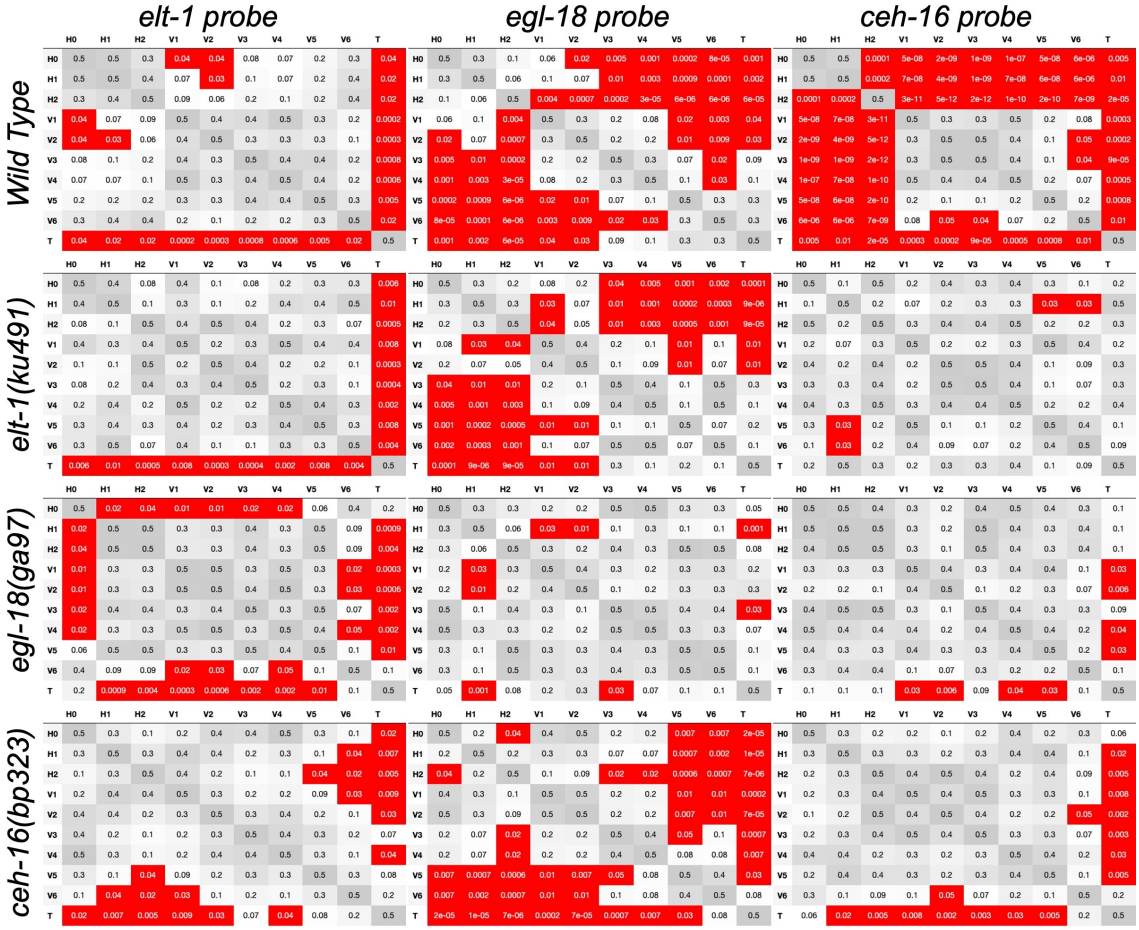


Figure 7.1: **P-values of the comparisons between the late L1 seam cells' gene expression patterns:** A matrix of Mann-Whitney U p-values for each of the three genes in wild type and mutant backgrounds comparing each cell to each other cell to determine which cells' expression patterns can be considered identical. The red colour reflects highly significant values and the gray colour reflects non-significant values.

Table 7.1: **Additional description of the various ODE models used: A)** Parameter names and definition. **B)** Description of the AND/OR logic of different models tested based on the initial MRA.

A

Parameter name	Parameter meaning
b_1	Basal production of *elt-1*, i.e. the amount of *elt-1* produced independently of the presence of the other two genes
b_2	Basal production of *egl-18*, i.e. the amount of *egl-18* produced independently of the presence of the other two genes
b_3	Basal production of ceh-*16*, i.e. the amount of *ceh-16* produced independently of the presence of the other two genes
k_1	The quantity of ELT-1 to produce half occupation on the activation sites, rescaled to be proportional to *elt-1* levels
k_2	The quantity of EGL-18 to produce half occupation on the activation sites, rescaled to be proportional to *egl-18* levels
k_3	The quantity of CEH-16 to produce half occupation on the repression sites, rescaled to be proportional to *ceh-16* levels
k_{3s}	The quantity of CEH-16 to produce half occupation on the activation sites, rescaled to be proportional to *ceh-16* levels
v_{1a2}	Production of *elt-1* activated by EGL-18 (OR link)
v_{1r3}	Production of *elt-1* repressed by CEH-16 (OR link)
v_1	Production of *elt-1* activated by EGL-18 AND repressed by CEH-16 (AND link)
v_{2a1}	Production of *egl-18* activated by ELT-1 (OR link)
v_{2r3}	Production of *egl-18* repressed by CEH-16 (OR link)
v_2	Production of *egl-18* activated by ELT-1 AND repressed by CEH-16 (AND link)
v_{3a1}	Production of *ceh-16* activated by ELT-1 (OR link)
v_{3a2}	Production of *ceh-16* activated by EGL-18 (OR link)
v_{3a3}	Production of *ceh-16* activated by CEH-16 (OR link)
v_3	Production of *ceh-16* activated by the AND action of 2 or 3 of the PROTEINS

B

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
EGL-18 activation, CEH-16 repression of *elt-1*	OR	OR	OR	AND	AND	AND	OR	OR	OR	OR	OR	AND	AND	AND	AND	AND	OR	OR	AND	AND
ELT-1 activation, CEH-16 repression of *egl-18*	OR	OR	OR	OR	OR	OR	AND	AND	AND	OR	OR	AND	AND	AND	OR	OR	AND	AND	AND	AND
ELT-1 activation, EGL-18 activation of *ceh-16*	OR	OR	OR	OR	OR	OR	OR	OR	OR	AND	AND	OR	OR	OR	AND	AND	AND	AND	AND	AND
ELT-1 activation, CEH-16 activation of *ceh-16*	OR	OR	AND	OR	OR	AND	OR	OR	AND	OR	AND	OR	OR	AND	OR	AND	OR	AND	OR	AND
EGL-18 activation, CEH-16 activation of *ceh-16*	OR	AND	OR	OR	AND	OR	OR	AND	OR	OR	AND	OR	AND	OR	OR	AND	OR	AND	OR	AND

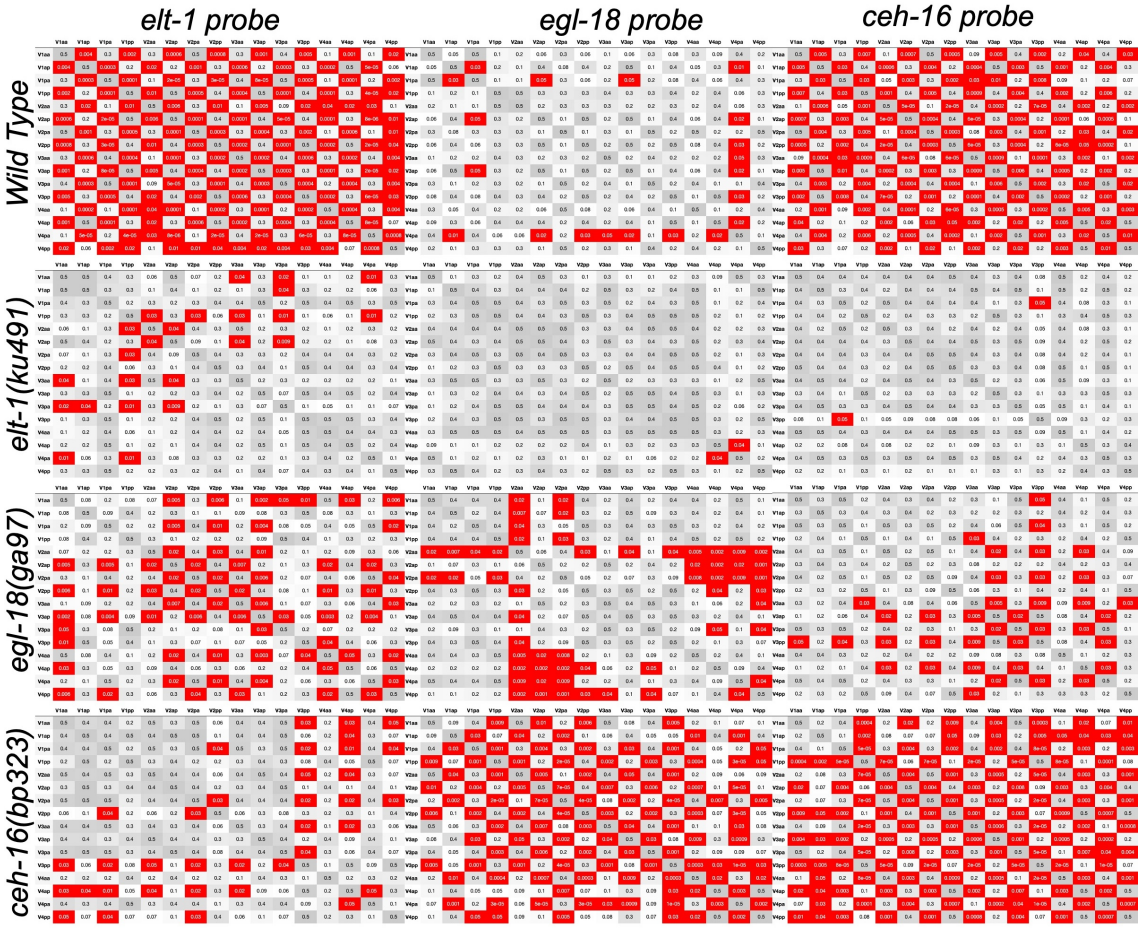


Figure 7.2: **P-values of the comparisons between the L2 asymmetric daughter seam cells' gene expression patterns including the differentiating cells:** A matrix of Mann-Whitney U p-values for each of the three genes in wild type and mutant backgrounds comparing each cell to each other cell to determine which cells' expression patterns can be considered identical. The red colour reflects highly significant values and the gray colour reflects non-significant values.

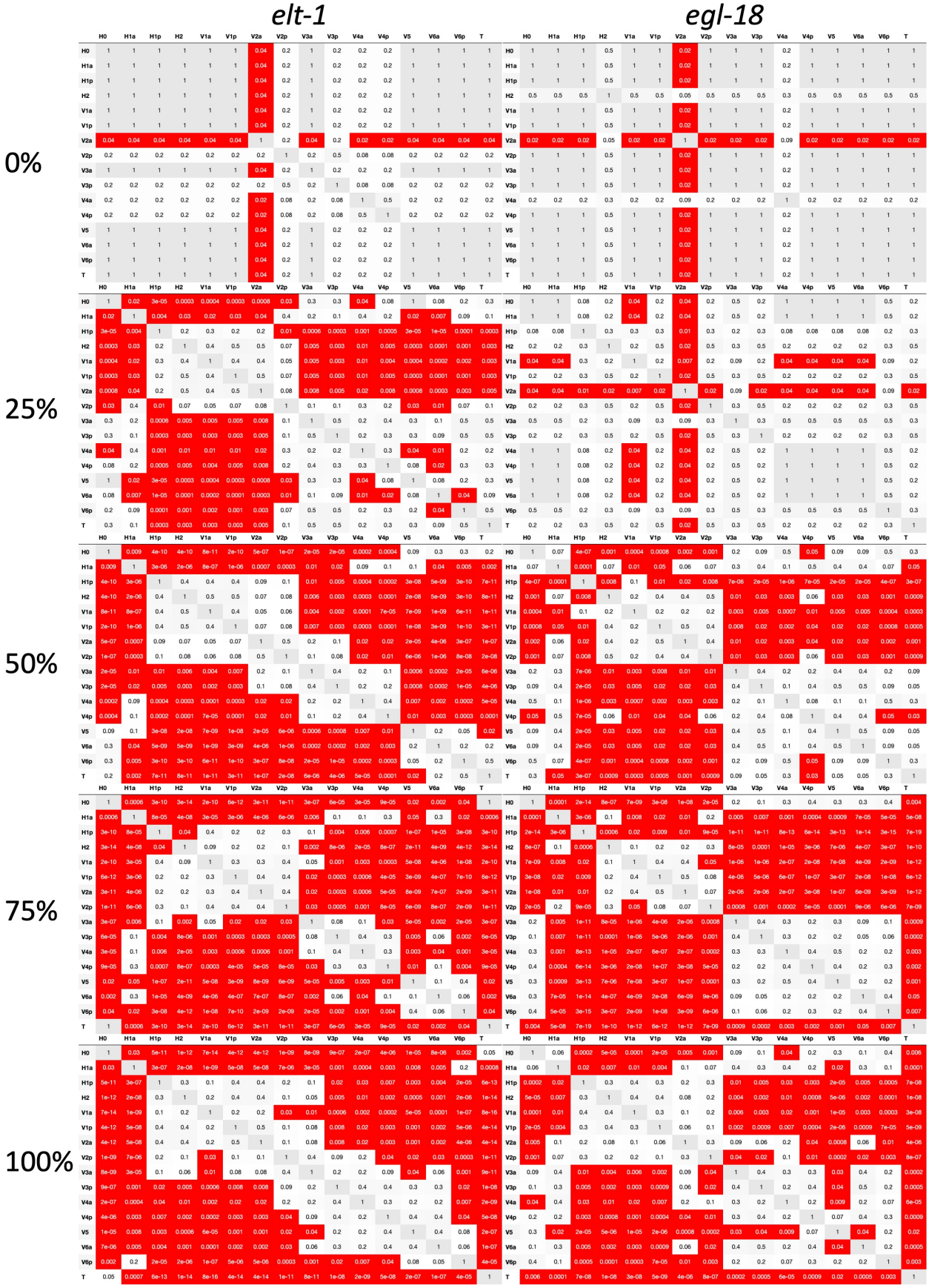


Figure 7.3: **P-values of the comparison of each cell's frequency of abnormal counts with each other cell's frequency of abnormal counts at increasing concentration of *elt-1* and *egl-18* RNAi:** We showcase the p-values of the cell-to-cell comparison of abnormal counts as calculated using Mann-Whitney U. Green indicates non-significant values and the redder the value the more significant the difference between two cells.

egl-5(n945)

	H0	H1a	H1p	H2	V1a	V1p	V2a	V2p	V3a	V3p	V4a	V4p	V5	V6a	V6p	T
Control Vector	1	1	1	0.8	1	1	0.6	0.8	0.8	1	0.8	1	1	0.6	1	1
50% elt-1 RNAi	0.9	1	0.4	0.04	0.3	0.5	0.3	0.2	0.2	0.3	0.2	0.2	0.7	0.8	0.9	0.8

lin-39(n1760)

	H0	H1a	H1p	H2	V1a	V1p	V2a	V2p	V3a	V3p	V4a	V4p	V5	V6a	V6p	T
Control Vector	1	1	0.8	1	0.8	0.8	0.6	0.6	0.8	0.9	1	1	0.8	1	1	1
50% elt-1 RNAi	0.9	0.7	0.08	0.4	0.9	0.5	0.1	0.09	0.04	0.02	0.04	0.4	0.7	0.8	0.3	0.8

lin-39(n1760); mab-5(e1239); egl-5(n945)

	H0	H1a	H1p	H2	V1a	V1p	V2a	V2p	V3a	V3p	V4a	V4p	V5	V6a	V6p	T
Control Vector	1	1	1	1	1	0.8	0.9	1	0.5	1	0.8	1	1	1	1	1
50% elt-1 RNAi	0.9	0.7	0.2	0.009	0.1	0.9	0.06	0.1	0.01	0.03	0.2	1	1	0.9	0.3	1

bar-1(ga80)

	H0	H1a	H1p	H2	V1a	V1p	V2a	V2p	V3a	V3p	V4a	V4p	V5	V6a	V6p	T
Control Vector	1	1	0.8	1	0.8	1	0.6	0.8	0.5	0.8	1	1	0.8	1	1	1
50% elt-1 RNAi	0.9	0.2	0.9	0.2	0.4	0.5	0.3	1	0.6	0.9	0.6	0.9	0.09	0.02	0.003	1

egl-20(hu105)

	H0	H1a	H1p	H2	V1a	V1p	V2a	V2p	V3a	V3p	V4a	V4p	V5	V6a	V6p	T
Control Vector	1	1	1	1	1	0.8	0.7	1	1	1	0.8	1	0.8	1	1	1
50% elt-1 RNAi	0.5	0.008	0.06	0.2	0.7	0.4	0.6	0.08	0.04	0.2	0.06	0.03	0.0006	1e-05	0.001	0.6

lin-17(n671)

	H0	H1a	H1p	H2	V1a	V1p	V2a	V2p	V3a	V3p	V4a	V4p	V5	V6a	V6p	T
Control Vector	1	1	1	1	1	1	0.7	1	1	0.9	1	1	1	1	1	0.4
50% elt-1 RNAi	0.9	0.2	0.04	0.3	0.09	0.3	0.5	0.3	0.03	0.5	0.005	0.005	0.0002	0.001	0.0006	3e-05

Figure 7.4: **P-values comparing the mean number of cells in wild type and Hox and Wnt mutant strains, with and without 50% *elt-1* treatment:** We showcase the p-values of the each cells comparison of abnormal counts relative to the same treatment in the wild type as calculated using Mann-Whitney U. Green indicates non-significant values and the redder the value the more significant the difference between two cells.

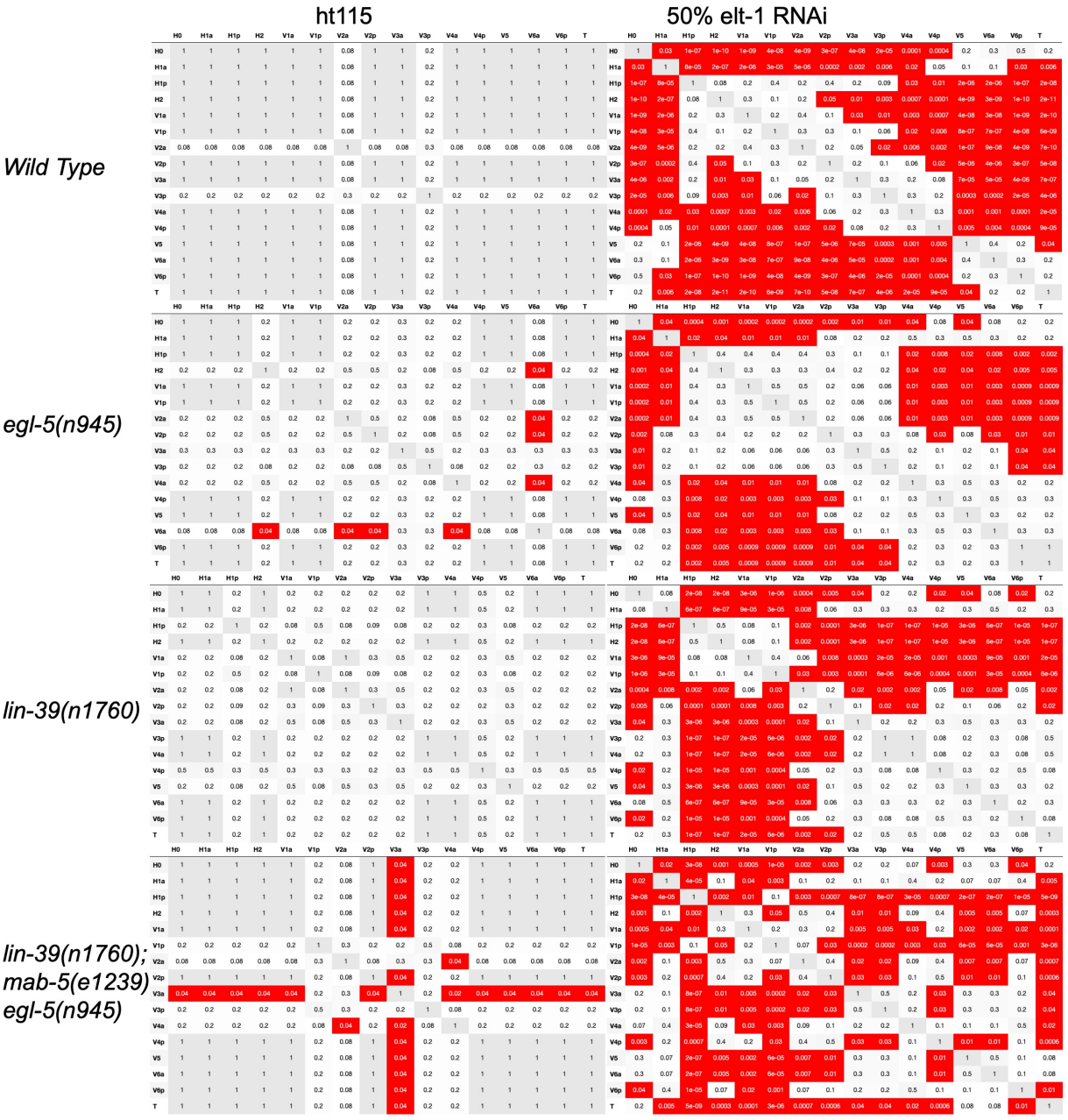


Figure 7.5: **P-values of the comparison of each cell's frequency of abnormal counts with each other cell's frequency of abnormal counts in wild type and Hox mutants *(egl-5(n945), lin-39(n1880) and lin-39(n1760);mab-5(e1239);egl-5(n945))* treated with the empty RNAi vector (ht115) and 50% *elt-1* RNAi:** We showcase the p-values of the cell-to-cell comparison of abnormal counts as calculated using Mann-Whitney U. Green indicates non-significant values and the redder the value the more significant the difference between two cells.

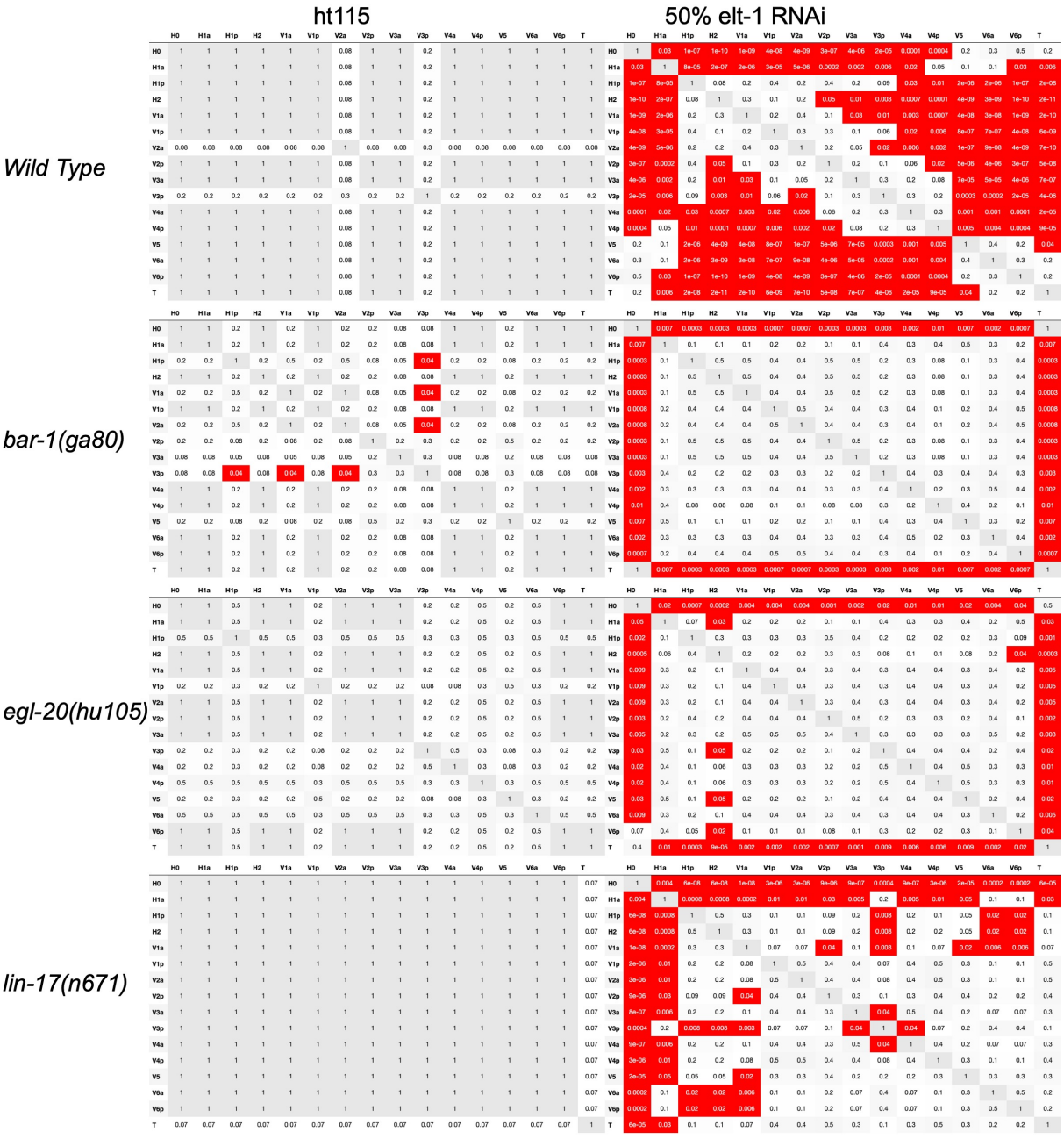


Figure 7.6: **P-values of the comparison of each cell's frequency of abnormal counts with each other cell's frequency of abnormal counts in wild type and Wnt mutants *(bar-1(ga80), egl-20(hu105) and lin-17(n671))* treated with the empty RNAi vector (ht115) and 50% *elt-1* RNAi:** We showcase the p-values of the cell-to-cell comparison of abnormal counts as calculated using Mann-Whitney U. Green indicates non-significant values and the redder the value the more significant the difference between two cells.

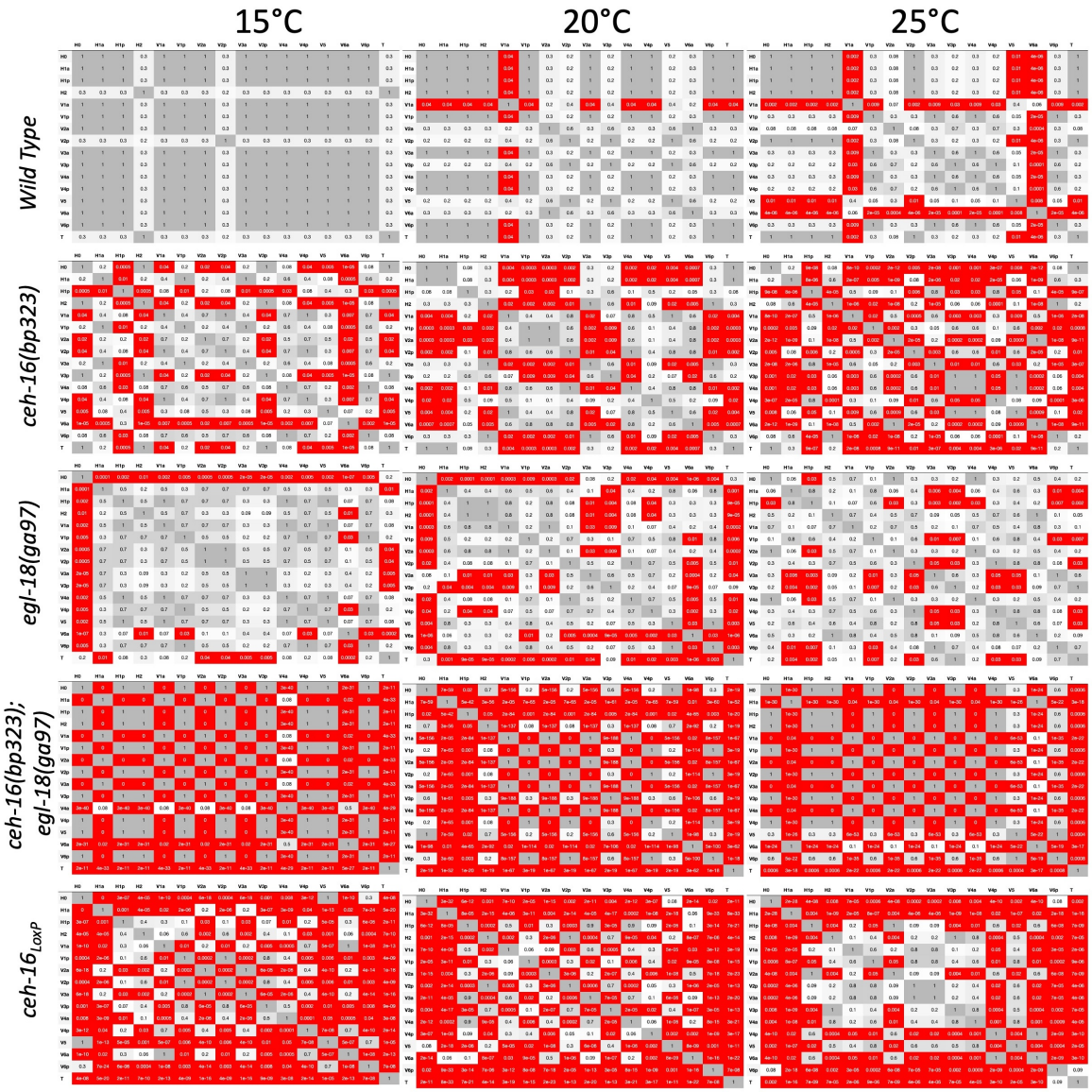


Figure 7.7: **P-values between cells of *wild type*, *ceh-16(bp323)*, *egl-18(ga97)*, *ceh-16(bp323);egl-18(ga97)* and *ceh-16*$_{LoxP}$ at 15℃, 20℃ and 25℃:** These were calculated using Mann-Whitney U to establish whether the frequency of the different numbers of seam cells was the same between cells