

**Imperial College
London**

Investigating the role of the fusogen *eff-1* and
natural genetic variation in
Caenorhabditis elegans seam cell development

Sneha Latha Koneru

Department of Life Sciences

Submitted in part fulfilment of the requirements
for the degree of Doctor of Philosophy at
Imperial College London, September 2019

Declaration of Originality

I confirm that the content of my thesis is a product of the research conducted by myself in the lab of Dr. Michalis Barkoulas at Imperial College London except when stated otherwise. I acknowledge the contribution of students and collaborators in the context of the appropriate experiments.



Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution-Non Commercial-No Derivatives licence (CC BY-NC-ND). Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the license terms of this work.

Abstract

Robustness is the ability of biological systems to produce invariant phenotypes despite perturbations. Development is especially robust to internal perturbations, like stochastic gene expression or mutations, and external perturbations, such as changes in environmental factors including temperature and nutrition. The highly invariant developmental patterning in *Caenorhabditis elegans* offers an ideal system to study the genetic and molecular mechanisms underlying developmental robustness. This work describes an experimental paradigm to discover the mechanistic basis and consequences of developmental robustness using the *C. elegans* seam cells as a model. Seam cells are lateral epidermal cells that are stem cell-like in their ability to produce differentiated cells and maintain proliferative potential. Through a forward genetic screen, I describe a novel role for the fusogen gene *eff-1*, which was previously known to drive cell fusion events, in the robustness of seam cell patterning. Furthermore, I show that *eff-1* is not required for differentiation of seam cells, therefore I demonstrate that fusion is uncoupled from the differentiation programme. In another set of experiments, I show for the first time that the terminal number of seam cells in *C. elegans* is robust to standing genetic variation. A consequence of developmental robustness is the acquisition of cryptic genetic variation that does not modify the phenotype under normal conditions but manifests phenotypically upon perturbation. I demonstrate that the genetic background affects seam cell number at a higher developmental temperature of 25 °C or upon mutations in the GATA transcription factor and target of the Wnt pathway, *egl-18*. CB4856 (Hawaii) suppressed the effect of temperature on the seam cell number compared to the lab reference N2 (United Kingdom), as well as lowered the expressivity of *egl-18* mutations. Multiple regions of the genome were found to interact epistatically to modify *egl-18* mutation expressivity, suggesting that a complex genetic archi-

ture underlies seam cell development. Taken together, this work increases our knowledge on the robustness of seam cell patterning to various sources of variation.

Acknowledgements

PhD in life sciences has been a humbling yet rewarding experience. First and foremost, I would like to thank my supervisor, Dr. Michalis Barkoulas, for providing me with the opportunity to work on exciting research problems and pursue a doctoral degree in his lab. This thesis would not have been as rigorous without the help, guidance and encouragement of my supervisor. I appreciate the trust he placed in me to allow me the independence and flexibility to work. Most importantly, Michalis has been a great mentor who led by the example of being an attentive and dedicated researcher.

My heartfelt gratitude to the present and past members of the research group – Michael, Iqrah, Sophie, Mark, Manish, Florence, Ritobrata, Lamia, Evangelina and Christina – who made my time in the lab fun and engaging. Especially, my fellow PhD student Dimitris for stimulating academic and non-academic discussions. I am indebted to my peers whose feedback helped me improve my research, writing, presentation and scientific communication skills. I also thank undergraduate and master's students Fu, Mingke, Jia and Judy for their assistance with my research. I express thanks to my colleagues and friends Marine, Natalie and Ari who have provided me with advice and helped me navigate graduate school.

I thank the European Research Council (ERC) for funding my research and supporting me financially. I acknowledge Stephen Rothery from FILM facility for his assistance with image analysis. I am grateful to the worm community for being helpful, collegial and sharing resources readily. I particularly want to thank Lise Frezal for sharing code and Marie Anne Felix for her encouragement and insightful discussions about my research. Thanks to Dr. Mark Sterken from Wageningen University and Chris Grove from Wormbase for helping me find relevant research material.

Special thanks to my friends in London, Aruna, Ranjani, Ashish, Deepak, Vivek, Kruthika and Arun who have provided me with a respite from research on so many weekends to recharge and given me a perspective outside of academia. I am particularly grateful to Archana and Manas for their friendship and support. I appreciate the encouragement from my friends from around the world, Andrea, Keertee, Shravya, Vikram and Swetha.

I would like to thank my grandma for her unconditional love and passing on her sense of confidence to me. I am grateful to my parents for their understanding and for keeping me grounded. I am thankful to my sister, Shravs, for being entertaining and loving. Her presence in my life sparked my interest in biology. I would like to thank extended family members, Amrutha and Bhargava for their kindness. Finally, and most importantly, I would like to thank my partner Vamsi for his endless patience, support and encouragement during the PhD. His companionship and unwavering belief in me have been instrumental in achieving my academic pursuits.

Epigraph

"The genome is certainly not a collection of 100000 commandments with everything carried out by dead reckoning." - Excerpt from Loose Ends by Sydney Brenner

Contents

Abstract	4
Acknowledgements	6
Epigraph	8
List of Figures	16
List of Tables	18
Abbreviations	20
List of gene names used in this work	22
1 General Introduction	25
1.1 General introduction	26
1.1.1 Robustness in development	26
1.1.2 Genetic modifiers hinder the predictability of phenotype from genotype .	28
1.1.3 Cryptic genetic variation is a consequence of developmental robustness .	31
1.1.4 Quantitative genetics as tool to map cryptic genetic variation	33
1.2 Introduction to <i>C. elegans</i>	34
1.2.1 The life cycle of <i>C. elegans</i>	34
1.2.2 <i>C. elegans</i> genome and genetics	35
1.2.3 Natural genetic variation in <i>C. elegans</i>	36
1.3 Postembryonic development of hypodermis	38
1.3.1 Overview of Seam cell patterning	38

1.4	Genetic control of postembryonic development	46
1.4.1	Heterochronic pathway regulates the temporal seam cell division patterns	46
1.4.2	Transcription factors involved in seam cell patterning	47
1.4.3	Wnt/ β -catenin asymmetry regulates the polarity of asymmetric seam cell divisions	48
1.4.4	Role of cell fusion in seam cell development	52
1.5	<i>C. elegans</i> seam cell development as a model system for studying developmental robustness	55
1.6	Aims of this research	56
2	Materials and Methods	58
2.1	General methods used in <i>C. elegans</i>	59
2.1.1	Maintenance	59
2.1.2	Cryopreservation	59
2.1.3	Transgenesis by microinjection	60
2.1.4	RNA interference (RNAi) by feeding	60
2.1.5	Genetics	61
2.1.6	Temperature treatment on seam cell development	61
2.1.7	Chemical mutagenesis, mutant screening and mapping	62
2.2	Microscopy and image analysis	62
2.2.1	Scanning electron microscopy (SEM) microscopy	62
2.2.2	Counting seam cell number	63
2.2.3	Lineaging seam cells	63
2.2.4	Long-term time-lapse microscopy of <i>eff-1</i>	63
2.2.5	Single molecule mRNA fluorescent in situ hybridisation (smFISH)	64
2.2.6	Seam-cell shape analysis	64
2.2.7	Measurement of angles between the seam cells during divisions	65
2.2.8	Confocal microscopy	65
2.2.9	<i>POPHHOP</i> marker intensity analysis	65
2.3	Molecular methods	66

2.3.1	Genomic DNA extraction from a single worm and polymerase chain reaction (PCR)	66
2.3.2	Sanger sequencing of targeted DNA sequences	66
2.3.3	Cloning <i>egl-18</i>	67
2.3.4	Cloning RNAi constructs	68
2.3.5	Design and cloning of single-guide RNA for Clustered Regularly Interspaced short Palindromic Repeats (CRISPR)-Cas9 mediated genome editing	68
2.3.6	co-CRISPR and mutant screening	69
2.3.7	Genomic DNA extraction from large numbers of worms	69
2.4	Whole genome sequencing	70
2.5	Quantitative genetics methods	70
2.5.1	Introgression of the seam cell marker <i>SCMp::GFP</i> into wild-isolates	70
2.5.2	Introgression of mutations in a GATA transcription factor (<i>egl-18</i>), and a fusogen (<i>eff-1</i>) into wild-isolates	71
2.5.3	Generation of recombinant inbred lines (RILs) containing <i>egl-18(ga97)</i> mutation	71
2.5.4	Phenotyping the RILs for seam cell number (<i>scn</i>) and pooling strategy for bulk segregant analysis	72
2.5.5	Quantitative trait locus QTL mapping	72
2.5.6	Bioinformatic analysis of whole genome sequencing data	73
2.5.7	Bulk segregant analysis	73
2.6	Molecular genetics	74
2.6.1	Designing genetic markers based on single nucleotide polymorphisms (SNPs) in the CB4856 genome	74
2.6.2	Designing genetic markers based on deletions/insertions in the CB4856 genome	75
2.6.3	Generation of near isogenic lines (NILs) containing quantitative trait loci	75
2.6.4	Generating smaller QTLs using genetic crosses and screening	77

2.7	Statistical analysis	77
3	Identification of a novel role for fusogen <i>eff-1</i> in the robustness of seam cell number	78
3.1	Introduction	79
3.2	Results	81
3.2.1	Forward genetic screen for mutants with variable seam cell number identifies MBA21	81
3.2.2	Mutation in MBA21 strain maps to the fusogen <i>eff-1</i>	81
3.2.3	<i>icb4</i> fails to complement a mutant allele of <i>eff-1</i>	83
3.2.4	Phenotypic characterisation of <i>eff-1(icb4)</i> mutant	85
3.2.5	Quantitative analysis of cell shape and cell division axis	87
3.2.6	Developmental basis for the seam-cell-number variability in <i>eff-1(icb4)</i>	91
3.2.7	Quantification of <i>eff-1</i> mRNA transcripts in seam cells with single molecule fluorescence in situ hybridisation imaging	93
3.2.8	<i>eff-1</i> is not required for the differentiation of anterior seam daughters	97
3.2.9	Differentiation of seam cells to hypodermal cells in <i>eff-1(icb4)</i> may be dependent on <i>nhr-25</i>	97
3.2.10	Overexpression of <i>egl-18</i> in anterior seam daughters phenocopies <i>eff-1(icb4)</i>	99
3.3	Discussion	101
3.3.1	EFF-1 is required for robust seam cell patterning	101
3.3.2	<i>eff-1</i> is not required for differentiation of anterior seam cell daughters	102
3.3.3	Loss of <i>eff-1</i> may affect seam cell patterning indirectly by disrupting seam cell contacts	104
4	Role of Natural Genetic Variation in Seam Cell Development	106
4.1	Introduction	107
4.2	Results	108
4.2.1	Seam cell number is robust to standing genetic variation	108

4.2.2	There is a $G \times E$ interaction between natural genetic variation and temperature that affects seam cell number	110
4.2.3	The expressivity of mutations in <i>lin-22</i> and <i>bro-1</i> did not differ between wild isolates	115
4.2.4	The expressivity of <i>eff-1(icb4)</i> did not differ between wild isolates	118
4.2.5	The expressivity of mutations in GATA transcription factor <i>egl-18</i> was lower in CB4856	120
4.2.6	The differential expressivity of <i>egl-18(ga97)</i> between wild isolates requires <i>elt-6</i> and <i>pop-1</i>	122
4.2.7	The differential expressivity of <i>egl-18(ga97)</i> is postembryonic	122
4.2.8	The differential expressivity of <i>egl-18(ga97)</i> phenotype is specific to the seam tissue	125
4.2.9	The differential expressivity of <i>egl-18(ga97)</i> seam phenotype is not dependent on <i>haw6805</i> polymorphism in <i>nath-10</i>	125
4.2.10	The differential expressivity of <i>egl-18(ga97)</i> phenotype is not dependent on increase in expression of its paralogue <i>elt-6</i>	126
4.3	Discussion	127
4.3.1	Effect of natural genetic variation on seam cell number	127
4.3.2	No differences in the expressivity of mutations in <i>lin-22</i> , <i>bro-1</i> and <i>eff-1</i>	128
4.3.3	There are differences in the expressivity of mutations in <i>egl-18</i> between wild isolates	130
4.3.4	Genetic basis of cryptic genetic variation affecting SCN in wild isolates carrying <i>egl-18(ga97)</i>	131

5 Mapping Genetic Variation Underlying Differences in Expressivity of the *egl-18(ga97)* mutation Between N2 And CB4856 **133**

5.1	Introduction	134
5.2	Results	136
5.2.1	Construction and phenotyping of the recombinant inbred lines between Bristol and Hawaii strains carrying <i>egl-18(ga97)</i>	136

5.2.2	Bulk segregant analysis of RILs between N2 and CB4856 carrying <i>egl-18(ga97)</i> identifies four QTLs	136
5.2.3	Genetic marker design and validation for QTL fine mapping	138
5.2.4	Depooling the high and low-bulk RILs using genetic markers	140
5.2.5	The QTLs on chr II, III and X represent cryptic genetic variation	142
5.2.6	Breaking the large genomic intervals of major QTLs residing on chromosomes II and III into smaller intervals	144
5.2.7	Identifying candidate genes in the QTLs on chromosomes II and III through an RNAi screen	147
5.2.8	Prioritisation of candidate genes in QTLs	154
5.2.9	QTLs may modify the <i>egl-18(ga97)</i> phenotype in N2 and CB4856 through the Wnt pathway	160
5.3	Discussion	163
5.3.1	Towards understanding genetic basis of differential expressivity of <i>egl-18(ga97)</i> by a quantitative genetics approach	163
5.3.2	<i>dsh-2</i> and <i>egl-27</i> are potential candidate genes on chromosome II affecting <i>egl-18(ga97)</i> expressivity	167
5.3.3	Novel role for <i>sor-1</i> and <i>hsp-110</i> for modifying seam cell number in an <i>egl-18</i> mutant background	169
6	General Discussion	171
6.1	The fusogen EFF-1 contributes to robustness of seam cell patterning	172
6.2	Cell differentiation is uncoupled from cell fusion in the seam	174
6.3	Cryptic genetic variation influences seam cell development	175
6.4	The genetic basis for the differential expressivity of <i>egl-18</i> mutations between N2 and CB4856	177
6.5	A developmental model for buffering seam cell number in <i>egl-18</i> loss-of-function mutants	178
	Bibliography	181

Appendix A Resources	209
A.1 List of strains	209
A.2 Lists of primers	212
Appendix B Additional tables	220
B.1 ANOVA table	220
B.2 PCA table	221
B.3 List of RILs	221
B.4 List of genes with natural variation on chromosome II	223
B.5 List of genes with natural variation on chromosome III	238
Appendix C Graphs	244
Appendix D Permissions	248

List of Figures

1.1	Illustration of concepts of penetrance and expressivity (reproduced from Fournier and Schacherer (2017))	29
1.2	Life cycle of <i>C. elegans</i> from an embryo to adult at 22°C (reproduced from WormAtlas)	35
1.3	Genome-wide phylogeny of <i>C. elegans</i> wild isolates (adapted from CeNDR, https://www.elegansvariation.org/data/release/latest)	39
1.3	Genome-wide phylogeny of <i>C. elegans</i> wild isolates (adapted from CeNDR, https://www.elegansvariation.org/data/release/latest)	40
1.4	The epidermis of <i>C. elegans</i> L1 larvae (reproduced from WormAtlas)	42
1.5	First L1 seam cell division in <i>C. elegans</i> visualised by <i>ajm-1::GFP</i> which marks the apical junctions (reproduced from WormAtlas)	43
1.6	<i>C. elegans</i> seam cell lineage. Adapted from Katsanos et al. (2017)	44
1.7	A model of asymmetric cell division controlled by $W\beta$ pathway (reproduced from Lam and Phillips (2017))	51
3.1	Seam cell number mutants recovered from our mutagenesis screen focusing on developmental variability	82
3.2	Causative mutation in MBA21 maps to chromosome II (6 Mb – 10 Mb)	84
3.3	The <i>icb4</i> mutation represents a strong loss of function of <i>eff-1</i>	86
3.4	<i>eff-1(icb4)</i> animals display tail, alae and seam cell defects	88
3.5	Seam cell shape is affected in <i>eff-1(icb4)</i> mutants	90
3.6	Seam cells become progressively misaligned between L2 symmetric and L2 asymmetric cell division in <i>eff-1(icb4)</i>	92

3.7	Developmental basis of SCN variability in <i>eff-1(icb4)</i>	94
3.8	Seam cell number increase in <i>eff-1</i> depends on <i>egl-18</i>	95
3.9	<i>eff-1</i> is expressed in anterior seam cell daughters	96
3.10	Anterior seam daughters differentiate in the absence of EFF-1	98
3.11	Synthetic interaction between loss-of-function of <i>eff-1</i> and <i>nhr-25</i>	99
3.12	Overexpression of <i>egl-18</i> in anterior seam daughters increase SCN variability . .	100
4.1	Seam cell number is robust to standing genetic variation in <i>C. elegans</i>	109
4.2	Presence of a $G \times E$ interaction between natural genetic variation and temper- ature in <i>C. elegans</i> that affects seam cell number	111
4.3	Percentage of animals with extra seam cells in wild isolates of <i>C. elegans</i> grown at 20°C and 25°C	113
4.4	Percentage of animals with extra seam cells in a subset of wild isolates of <i>C.</i> <i>elegans</i> grown at 20°C and 25°C	114
4.5	No difference in the expressivity of <i>bro-1</i> or <i>lin-22</i> mutations between wild isolates	117
4.6	Expressivity of <i>eff-1(icb4)</i> and <i>egl-18(ga97)</i> mutation in SCN between wild isolates	119
4.7	Differential expressivity of <i>egl-18(ga97)</i> is not recapitulated by <i>egl-18</i> RNAi . . .	121
4.8	The differential expressivity of <i>egl-18(ga97)</i> mutation in N2 and CB4856 is postembryonic and specific to seam cell number	123
4.9	The differential expressivity of <i>egl-18(ga97)</i> seam phenotype is not due to <i>nath(haw6805)</i> or <i>elt-6</i> expression differences	126
5.1	Generation and phenotyping of recombinant inbred lines	137
5.2	Bulk segregant analysis of recombinant inbred lines	139
5.3	Validation of genetic markers based on deletions in Hawaiian (Cb4856) genome .	141
5.4	Genotyping of recombinant inbred lines with genetic markers based on deletions in Hawaiian (CB4856) genome	143
5.5	Phenotypic analysis of near isogenic lines indicates two major quantitative trait loci	145

5.6	Phenotypic and genotypic analysis of NILs carrying genomic fragments of chromosome II from CB4856	148
5.7	Phenotypic and genotypic analysis of NILs carrying genomic fragments of chromosome III from CB4856	149
5.8	Subtle differences in sensitivity of somatic RNAi between N2 and CB4856	151
5.9	RNAi screen to identify candidate genes in QTL on chromosome II	153
5.10	RNAi screen to identify candidate genes in QTL on chromosome III (first set) .	155
5.11	RNAi screen to identify candidate genes in QTL on chromosome III (second set)	157
5.12	Genotyping of NILs carrying genomic fragments of chromosomes II and III from CB4856 using SNP-based genetic markers	159
5.13	No differences in overall Wnt pathway activity between N2 and CB4856	161
5.14	N2 and CB4856 show differences in SCN upon RNAi knockdown of Wnt components	164
6.1	A schematic of the developmental model by which seam cell number may be buffered in <i>egl-18(ls)</i> mutants	180
C.1	Additional seam cell lineages of <i>eff-1(icb4)</i> animals	245
C.2	Additional Seam cell lineages of <i>eff-1(icb4)</i> animals	246
C.3	Proportion of CB4856 SNPs along the six chromosomes in stringent condition .	247

List of Tables

3.1	Mutation in MBA21 maps to third exon of <i>eff-1</i>)	83
A.1	Strains used in this thesis	210
A.2	Cloning and sequencing primers	212
A.3	Fluorescent labelled oligos used in smFISH	213
A.4	List of genetic markers based on SNPs in CB4856 genome	215
A.5	List of genetic markers based on snip-SNPs in CB4856 genome	217
A.6	List of genetic markers based on indels in CB4856 genome	218
B.1	ANOVA table pertaining to cell division angles in <i>eff-1(icb4)</i> compared to wild type	220
B.2	PCA table for seam cell shape	221
B.3	List of RILs	221
B.4	List of genes with natural variation in the genomic interval chr. II (4794415 – 8173823)	224
B.5	List of genes with natural variation in the genomic interval chr. III (7631003 – 8917452)	239

Abbreviations

ANOVA	<u>A</u> nalysis <u>O</u> f <u>V</u> ariance
chr.	<u>C</u> hromosome
CGC	<u>C</u> aenorhabditis <u>G</u> enetics <u>C</u> enter
CGV	<u>C</u> ryptic <u>G</u> enetic <u>V</u> ariation
CTCF	<u>C</u> orrected <u>T</u> otal <u>C</u> ell <u>F</u> luorescence
dsRNA	<u>D</u> ouble <u>S</u> tranded <u>R</u> ibonucleic <u>A</u> cid
EMS	<u>E</u> thyl <u>M</u> ethane <u>S</u> ulfonate
GCR	<u>G</u> enetic <u>C</u> ompensation <u>R</u> esponse
GFP	<u>G</u> reen <u>F</u> luorescent <u>P</u> rotein
GWAS	<u>G</u> enome <u>W</u> ide <u>A</u> ssociation <u>S</u> tudy
hyp7	<u>H</u> ypodermal syncytium 7
IPTG	<u>I</u> so <u>P</u> ropyl β -D-1- <u>T</u> hiogalactopyranoside
Indel	<u>I</u> nsertion- <u>D</u> eletion
LB	<u>L</u> ysogeny <u>B</u> roth
miRNA	<u>M</u> icro <u>R</u> NA
min	<u>M</u> inutes
NGM	<u>N</u> ematode <u>G</u> rowth <u>M</u> edium
NIL	<u>N</u> ear <u>I</u> sogenic <u>L</u> ines
NMD	<u>N</u> onsense- <u>M</u> ediated <u>D</u> ecay
PCA	<u>P</u> rincipal <u>C</u> omponent <u>A</u> nalysis
PCR	<u>P</u> olymerase <u>C</u> hain <u>R</u> eaction
PDE	<u>P</u> ost <u>d</u> eid

PTC Premature-Termination Codon
QTL Quantitative Trait Locus
RIL Recombinant Inbred Lines
RFLP Restriction Fragment Length Polymorphism
RPM Rotations Per Minute
RNAi Ribonucleic Acid Interference
SCN Seam Cell Number
sgRNA Single Guide RNA
snip-SNP RFLP-based SNP
SNP Single Nucleotide Polymorphism

List of Genes

<i>aff-1</i>	<u>A</u> nchor cell <u>F</u> usion <u>F</u> ailure
<i>arf-3</i>	<u>A</u> DP- <u>R</u> ibosylation <u>F</u> actor related
<i>bar-1</i>	<u>B</u> eta-catenin/ <u>A</u> rmadillo <u>R</u> elated
BRCA	<u>B</u> Reast <u>C</u> Ancer gene
<i>bro-1</i>	<u>B</u> R <u>O</u> ther (Drosophila tx factor partner) homolog
<i>cdc-14</i>	<u>C</u> ell <u>D</u> ivision <u>C</u> ycle related
CDC14	<u>C</u> ell <u>D</u> ivision <u>C</u> ycle 14A
<i>cdh-3</i>	<u>C</u> a <u>D</u> Herin family
<i>ceh-16</i>	<u>C.</u> <u>e</u> legans <u>H</u> omeobox
CFTR	<u>C</u> ystic <u>F</u> ibrosis <u>T</u> fibrosis conductance <u>R</u> egulator
<i>clp-1</i>	<u>C</u> a <u>L</u> Pain family
<i>del-10</i>	<u>D</u> Egenerin <u>L</u> ike
<i>dgk-5</i>	<u>D</u> iacyl <u>G</u> lycerol <u>K</u> inase
<i>dsh-2</i>	<u>D</u> i <u>S</u> Hvelled related
DVL	<u>D</u> ishe <u>V</u> e <u>L</u> led segment polarity protein
ECM	<u>E</u> xtra <u>C</u> ellular <u>M</u> atrix
<i>eef-1A.1</i>	<u>E</u> karyotic translation <u>E</u> longation <u>F</u> actor
<i>eff-1</i>	<u>E</u> pithelial <u>F</u> usion <u>F</u> ailure
EGF	<u>E</u> pidermal <u>G</u> rowth <u>F</u> actor
EGFR	<u>E</u> pidermal <u>G</u> rowth <u>F</u> actor <u>R</u> eceptor
<i>egl-5</i>	<u>E</u> Gg <u>L</u> aying defective
<i>egl-18</i>	<u>E</u> Gg <u>L</u> aying defective

<i>egl-27</i>	<u>EG</u> g <u>L</u> aying defective
<i>elt-6</i>	<u>E</u> rythroid- <u>L</u> ike <u>T</u> ranscription factor family
HOX	<u>H</u> Omeobo <u>X</u>
<i>hsp-110</i>	<u>H</u> eat <u>S</u> hock <u>P</u> rotein
HTT	<u>H</u> un <u>T</u> ing <u>T</u> in
<i>kle-2</i>	<u>KLE</u> isin (abnormal closure) family
LEF	<u>L</u> ymphoid <u>E</u> nhancer binding <u>F</u> actor
<i>let-23</i>	<u>LE</u> Thal
<i>let-60</i>	<u>LE</u> Thal
<i>lin-9</i>	abnormal cell <u>LI</u> Neage
<i>lin-36</i>	abnormal cell <u>LI</u> Neage
<i>lit-1</i>	<u>L</u> oss of <u>I</u> n <u>T</u> estine
<i>lnkn-1</i>	conserved transmembrane adhesion protein involved in <u>LI</u> <u>N</u> <u>K</u> <u>I</u> <u>N</u> g cells together
<i>mab-5</i>	<u>M</u> ale <u>AB</u> normal
<i>mig-10</i>	abnormal cell <u>MIG</u> ration
<i>mrt</i>	<u>Mo</u> <u>RT</u> al germline
<i>nath-10</i>	<u>NAT</u> nath (vertebrate N-AcetylTransferase) <u>H</u> omolog
<i>ncl-1</i>	abnormal <u>Nu</u> <u>CL</u> eoli
<i>nhr-25</i>	<u>N</u> uclear <u>H</u> ormone <u>R</u> eceptor family
NLS	<u>N</u> uclear <u>L</u> ocalisation <u>S</u> ignal
NuRD	<u>NU</u> cleosome <u>R</u> emodeling <u>D</u> eacetylase
PcG	<u>P</u> oly <u>C</u> omb <u>G</u> roup
<i>pop-1</i>	<u>PO</u> sterior <u>P</u> harynx defect
PRC	<u>P</u> olycomb <u>R</u> epressive <u>C</u> omplex
<i>pry-1</i>	<u>P</u> oly <u>Ra</u> <u>Y</u>
<i>pvl</i>	<u>P</u> rotruding <u>Vu</u> <u>L</u> va
RERE	arginine – glutamic acid <u>RE</u> dipeptide <u>RE</u> peats
<i>rol</i>	<u>ROL</u> ler

<i>rnt-1</i>	<u>RuNT</u> related
<i>set-24</i>	<u>SET</u> (trithorax/polycomb) domain containing
<i>sma-2</i>	<u>SMAl</u>
<i>sor-1</i>	<u>SO</u> p-2 <u>R</u> elated (ectopic expression of Hox genes)
<i>tbb-2</i>	<u>TuB</u> ulin, <u>B</u> eta
TCF	<u>T-C</u> ell <u>F</u> actor
<i>unc-54</i>	<u>UNC</u> oordinated
<i>vab-10</i>	<u>V</u> ariable <u>AB</u> normal morphology
<i>wve-1</i>	<u>WAVE</u> (actin cytoskeleton modulator) homolog
<i>zfp-1</i>	<u>Z</u> inc <u>F</u> inger <u>P</u> rotein
<i>ztf-30</i>	<u>Z</u> inc finger putative <u>T</u> ranscription <u>F</u> actor family

Chapter 1

General Introduction

1.1 General introduction

Genetic variation is the ultimate substrate on which evolution acts. However, this genetic variation resides within individual organisms which serve as vessels to propagate their genetic information. Therefore, it is of utmost importance that individual animals and plants develop optimally to successfully reproduce and transmit their genetic information. Developmental biology is the study of the molecular processes underlying the transition from a single-celled embryo to a multicellular organism. Two intriguing aspects of organismal development are highly relevant for this thesis: first, how complex the relationship between genotype and phenotype is, which dictates the building of multicellular organisms based on the genetic code. Understanding how the genetic code encodes phenotypes involves the discovery of genes that govern cell growth, differentiation and morphogenesis. Second, normal development often proceeds reproducibly despite stochastic gene expression, standing genetic variation in genes and changing environmental conditions. C. H. Waddington proposed the concept of canalisation as a buffering mechanism to explain the lack of phenotypic diversity despite standing genetic variation in development (Waddington, 1942). Uncovering buffering mechanisms in developmental biology is an open and interesting question in this field and is one of the aims of this thesis.

1.1.1 Robustness in development

Robustness is a fundamental property of biological systems and can be defined as their ability to produce and maintain consistent phenotypes despite perturbations (Whitacre, 2012; Kitano, 2004; Visser et al., 2003). For example, focusing at the molecular level, tight synchronisation of gene expression across spatiotemporal scales in a changing environment is essential for robust development of multicellular organisms from embryos (Maduro, 2015). Large deviations in quantity and timing of gene expression can cause developmental errors that may lead to inviable or less fit offspring. The ability to precisely control gene expression is therefore a challenge for cells as gene expression has been shown to be stochastic and noisy, even in isogenic cells grown in a constant environment (Symmons and Raj, 2016; Raj and Oudenaarden, 2008; Raj, Rifkin, et al., 2010; Elowitz et al., 2002).

Concerning possible perturbations, mutations or natural genetic variation segregating in the

population might pose a problem to a particular developmental system. Continuing with the same example, mutations or natural variation in a promoter region may increase the expression of a critical developmental gene multifold. Moreover, the environment in which an organism develops is also variable and may lead to a similar effect. Phenotypic variation in development can therefore remain low despite noisy gene expression, standing genetic variation in the population and changing environment. This highlights that development can be robust to both internal perturbations (mutations or molecular stochasticity) and external perturbations in the environment, such as changes in temperature and food (Félix and Wagner, 2008; Masel and Siegal, 2009).

It has been debated whether robustness is a distributed property of developmental gene networks or whether single genes can confer robustness (Wagner, 2005; Masel and Siegal, 2009). Siegal and Bergman (2002) and Bergman and Siegal (2003) have shown through theoretical work that developmental robustness to genetic variation can emerge in highly connected networks. Systematic screens to identify the basis of buffering of morphological traits to perturbations have been previously undertaken in *Saccharomyces cerevisiae*, but not in multicellular animals (Levy and Siegal, 2008; Giaever et al., 2002). Another study in *S. cerevisiae* found that there was greater functional compensation for duplicate genes than for singletons, suggesting that a paralogue can confer robustness to mutation in its gene duplicate (Gu et al., 2003). Knocking out 5% of *S. cerevisiae* genes increased phenotypic variance to environmental variation, and these genes were highly connected in the protein-protein interaction network (Levy and Siegal, 2008). An alternative mechanism to compensation by gene duplicates is through distributed robustness that involves backup pathways in metabolic or gene regulatory networks (Wagner, 2005).

The most studied robustness-conferring single gene is *Hsp90* (heat shock protein 90), which encodes a molecular chaperone that helps proteins fold correctly, thus maintaining protein homeostasis. Several studies in *Drosophila melanogaster*, *Arabidopsis thaliana* and *Caenorhabditis elegans* have shown that loss of *Hsp90* reveals phenotypic variation due to loss of buffering (Rutherford and Lindquist, 1998; Queitsch, Sangster, and Susan Lindquist, 2002; Katsanos et al., 2017). These loci that suppress phenotypic variation are called phenotypic capacitors. The

mechanisms by which these “robustness genes” buffer different types of variations to produce robust phenotypic outcomes remain mostly unknown. Developmental robustness may thus be a result of a single gene or epistatic interactions of multiple genes.

1.1.2 Genetic modifiers hinder the predictability of phenotype from genotype

Since the complete sequence of the human genome was completed in 2003 by Human Genome Project (HGP), there has been a gold rush to use genomic data to predict phenotypic traits and predisposition to diseases. Personalised genomics companies like 23andMe[®] and Veritas Genetics[™] offer direct-to-consumer genetic testing for anyone who wants to find out their predisposition to certain well-characterised diseases and morphological phenotypes, like height and eye colour. The goal of personalised medicine is to use an individual’s genotype to predict disease predisposition and even inform preventative or therapeutic interventions. However, there is a fundamental challenge of predicting phenotypes from the genomic sequence as genotype to phenotype relationship is complex. Genomics companies have taken different approaches to predict phenotypes from genotypes. Veritas Genetics[™] scans the entire genome (3.2 Gb) compared to 23andMe[®], which scans only predetermined SNPs (690 000 bp) across the genome. Other companies have chosen the middle ground by sequencing only the exome or protein coding genes (50 Mb). Most genomics companies predict phenotypes using scientific knowledge from public databases such as ClinVar or based on statistical models based on user-reported phenotypes (Landrum and Kattman, 2018). However, the published literature on the genotype-phenotype relationship for diseases is often based on small sample sizes of genetically related individuals and might not be representative of the whole human population. In the era of personalised medicine, the genetic background dependence of phenotypes poses the biggest challenge to precisely predict the effect of a genetic variant on an individual’s phenotype (Gasch, Payseur, and Pool, 2016). This is because phenotypic outcomes of developmental diseases, and diseases such as cancer are variable and dependent on the genetic background of the individual.

Genetic background comprising of natural genetic variation can produce vastly different effects on a developmental or disease phenotype in different individuals (Chandler, Chari, and Dworkin, 2013). Even in monogenic disorders, which are caused by mutations in a single gene,

there can be phenotypic differences either in penetrance or expressivity as illustrated in Fig. 1.1. For example, mutations in breast cancer gene 1 (BRCA1), which increase the risk for breast and ovarian cancer in humans can display incomplete penetrance – only a subset of the population carrying the mutation will go on to develop cancer – this is due to epistatic interactions and suppressors in the genome (Chandler, Chari, and Dworkin, 2013). In the case of cystic fibrosis, a disease characterised by damage to lungs, digestive system and other organs, the penetrance of mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene is complete or 100%. However, individuals carrying mutations in the CFTR gene show differential expressivity or different levels of severity of the disease (Fournier and Schacherer, 2017). This differential expressivity of CFTR mutations is due to the presence of modifiers (genes that affect the phenotype) in the genetic background of individuals (Cutting, 2010).

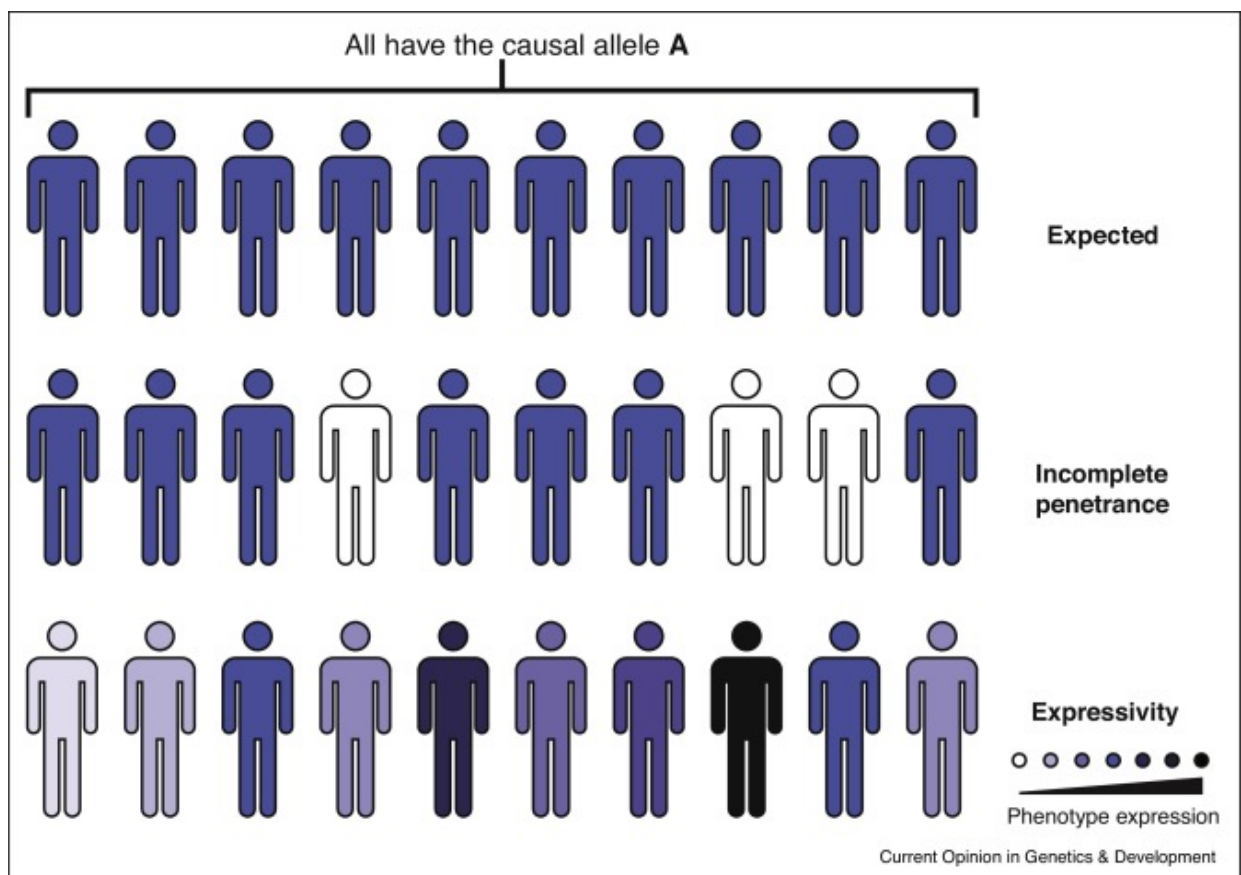


Figure 1.1: **Illustration of concepts of penetrance and expressivity (reproduced from Fournier and Schacherer (2017))**. In a population, all individuals carrying the causal allele A may show the expected phenotype which constitutes 100% penetrance. In the case of incomplete penetrance, only a proportion of the population carrying the causal allele A develop the disease. Expressivity is a special case of 100% penetrance, all individuals carrying the causal allele A will go on to develop the disease but to a variable extent.

There have been a growing number of studies in model organisms, which highlight differences in phenotypic expressivity of mutant alleles in different backgrounds and environments. In *C. elegans* development, RNAi knockdown of essential embryonic genes leads to variability of embryonic lethality across different wild isolates (Paaby, A. G. White, et al., 2015). Similarly, in *S. cerevisiae*, genes considered essential in one isolate were dispensable for survival in the other (Dowell et al., 2010). Phenotypic variability in the number of bristles per haltere was discovered by introgression of the homeotic mutation in a hox gene *Ultrabithorax Ubx* in 29 wild-type *D. melanogaster* lines (Gibson and Helden, 1997). Such background effects on phenotype have also been observed in *Mus musculus* (Montagutelli, 2000). For example, the severity of symptoms associated with a ferrochelatase deficiency mutation in *M. musculus* varied between three different genetic backgrounds (Abitbol et al., 2005). A problem therefore with using single reference laboratory strains is that they might not be sufficient to yield functional information about all genes or only provide a partial description of gene function. For instance, even though *C. elegans* was the first multicellular organism to have a completely sequenced genome, more than 40% of its genes have no functional annotation. Similarly, 32% and 22% of genes in *D. melanogaster* and *M. musculus*, respectively do not have curated functional annotation whereas *S. cerevisiae*, which has been the subject of many quantitative genetics studies, has only 8% annotated (Petersen, Dirksen, and Schulenburg, 2015). The number of functionally annotated genes might be increased by studying mutant alleles in a broader range of backgrounds. For example, loss of function of genes that may not have a phenotype in one *S. cerevisiae* strain but may have a phenotype in another (Dowell et al., 2010). It is also possible that the genes lacking functional annotation are essential in responses to environmental variation in a natural habitat that may never be encountered in standard laboratory conditions (Petersen, Dirksen, and Schulenburg, 2015).

Currently, we lack an understanding of how natural genetic variation affects phenotypic outcomes limiting our ability to predict phenotypic outcomes based on genotype. This is understandable in the context of humans because it is not possible to study phenotypes in scenarios where the environment and other non-heritable factors are well controlled. However, the phenotypic consequences of genetic variation have been largely ignored even in model or-

ganisms, where it is possible to control environment and other contributing factors in favour of reproducibility gained from the use of genetically isogenic laboratory lines. Mutant alleles are studied in one laboratory strain to understand the effect of a single genetic variable while holding everything else constant. Genetically isogenic lines are an excellent resource for elucidating gene function as they reduce the confounding effects of other genes in the background. However, laboratory-aided evolution of model organisms can fix alleles that have pleiotropic effects on unrelated traits such as life-history traits like fecundity (Duveau and Félix, 2012; Sterken et al., 2015). Therefore, the biology learned from the reference laboratory strains, even though interesting, might not be representative of the wild isolates of the species (Gasch, Payseur, and Pool, 2016) or incorrect conclusions may be drawn. For example, a study in *C. elegans* found that the gene *pha-1*, which was considered to be involved in pharyngeal development was found to be an antidote expressed in the zygote to a maternal-effect toxin *sup-35* (Ben-David, Burga, and Kruglyak, 2017). This is supported by the fact that all phenotypes associated with *pha-1* loss-of-function are suppressed by mutations in *sup-35* and overexpression of *sup-35* phenocopies *pha-1* loss-of-function (Kuzmanov, Yochem, and Fay, 2014). In a *M. musculus* model, alleles of genes implicated in human psychiatric illnesses were studied in different genetic backgrounds, wherein they found strong interactions of alleles with the genetic background that supported opposing conclusions (Sittig et al., 2016). Therefore, findings from studies examining gene function and allelic effects on disease should be validated in different genetic backgrounds.

1.1.3 Cryptic genetic variation is a consequence of developmental robustness

Developmental buffering results in a nonlinear genotype to phenotype map (Félix and Barkoulas, 2015). This has been studied previously in the context of robust phenotypes in *C. elegans* like vulval cell fate patterning. In this case, it was shown that the system tolerates a 4-fold variation in the genetic dose of a key signalling molecule, LIN-3/epidermal growth factor (EGF), without a change in the output of the cell fate pattern (Barkoulas et al., 2013). Developmental buffering of phenotypes can allow the underlying gene network to accumulate conditionally neutral genetic variation, which does not manifest phenotypically in wild-type

conditions (Félix and Wagner, 2008). This type of natural genetic variation that does not affect phenotypes in the wild-type condition but can be revealed upon genetic perturbation or environmental perturbation is called cryptic genetic variation (CGV) (Félix and Wagner, 2008; Gibson and Dworkin, 2004). CGV is a hidden source of variation for natural selection to act upon and may facilitate adaptation (Paaby and Rockman, 2014). Moreover, CGV is responsible for differences in penetrance/expressivity of mutant alleles (Gibson and Dworkin, 2004). Consistent with this idea, the severity of RNAi phenotypes observed in *C. elegans* is dependent on the genetic background (Vu et al., 2015). CGV is relevant for human disease as the severity of human diseases varies between individuals because genetic modifiers interact with disease causing loci. The emergence of a complex disease like type II diabetes is thought to be a result of the CGV that has accumulated over the evolution and is now revealed by modern life-style (Queitsch, Carlson, and Girirajan, 2012; Gibson and Dworkin, 2004).

There are two methods to uncover CGV in a developmental system. The first is to subject different genetic backgrounds to environmental perturbations such as changes in temperature, nutrition and growth conditions. These experiments typically involve the study of genotype-by-environment ($G \times E$) interactions affecting the phenotype of interest. The second is to introduce genetic perturbations, such as mutations to different genetic backgrounds. These experiments involve the investigation of genotype-by-genotype ($G \times G$) interactions affecting the phenotype of interest (Gibson and Dworkin, 2004). Introgression of mutations into different genetic backgrounds requires generations of backcrossing (Gibson and Dworkin, 2004). This method has been used to discover CGV affecting photoreceptor determination by introgression of a mutation in epidermal growth factor receptor (EGFR) gene *Egfr* into two wild-type genetic backgrounds of *D. melanogaster* (Dworkin, Palsson, et al., 2003). A study in *C. elegans* also discovered CGV affecting the vulval cell-fate induction of mutations in EGFR gene *let-23* by introgressing a mutation from the lab reference strain into a wild-type background (Duveau and Félix, 2012). Multiple other studies have discovered CGV in vulval patterning gene network in *C. elegans* (Milloz et al., 2008; Braendle and Félix, 2008; Félix and Barkoulas, 2012; Grimbirt and Braendle, 2014), although in most cases the molecular determinants of CGV remain elusive.

The environmental conditions can be manipulated easily to investigate $G \times E$ interactions.

Genetic introgressions of mutations to reveal $G \times G$ interactions is a labour-intensive task in *C. elegans* because it relies on backcrossing a mutation at least ten times, which even with the short generation time of *C. elegans* takes on an average about five weeks per strain. The advent of gene editing technology such as CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas9 allows rapid generation of the precise mutations in different wild-type backgrounds (Friedland et al., 2013; Dickinson et al., 2013). CRISPR-Cas9 is conceivably more precise than introgression of mutations because genomic regions flanking the gene of interest may be carried over even after 10 generations of backcrossing, which precludes the study of their effect on the phenotype.

1.1.4 Quantitative genetics as tool to map cryptic genetic variation

Most quantitative traits are complex or polygenic, that is they do not follow simple Mendelian inheritance laws based on a single allele (U. Kim et al., 2003). These traits do not fall into discrete classes and show a continuous range of phenotypic variation that can be influenced by both the environment and genetics (Mackay, Stone, and Ayroles, 2009). Mendelian (or monogenic) traits can be more easily predicted based on the genotype; however, prediction of complex traits is not straightforward. In humans, height is an example of a complex trait, which is influenced by ≈ 700 common variants located in hundreds of loci (Wood et al., 2014). Furthermore, height is also influenced by environmental factors such as nutrition. Quantitative genetics methods allow the detection and quantification of the contribution of genotype and environment to the phenotypic variance.

The genetic basis of phenotypic variation can be investigated with either quantitative trait loci (QTL) mapping or genome wide association studies (GWAS), reviewed in Bazakos et al. (2017). QTL mapping involves the generation of recombinant inbred lines (RILs), which are typically derived by several generations of inbreeding of F1 progeny from a cross between phenotypically and genetically divergent parents. RILs contain randomly shuffled parental genomes allowing the derivation of a genotype-to-phenotype correlation. QTLs are then detected by statistical analysis of this correlation. QTLs often contain large genomic fragments and need to be fine mapped by breaking down the genomic fragments through additional genetic recombination. This is done by producing near isogenic lines (NILs) which contain only

the QTL of interest in an otherwise homozygosed genome of one parent. NILs also allow the quantification of the contribution of potentially multiple individual QTLs to the phenotype. In addition, epistasis between the QTLs can be revealed through NILs by analysing different QTL combinations in NILs. QTL mapping has the potential to identify novel genes involved in phenotypic variation. However, one major drawback of QTL mapping is the low resolution of QTLs, which may contain thousands of genes. Resolving the QTL to a single gene level requires laborious classical genetics to breakdown the genomic interval identified.

GWAS is a complementary approach to QTL mapping, which takes advantage of natural genetic variation segregating in wild populations. The most significant advantage of GWAS over QTL mapping is the higher resolution of mapping of the causative variants of the phenotype, especially for organisms with substantial genetic diversity, without the need to perform genetic crosses, which of course is not possible for human studies. Depending on the number of natural isolates included in the GWAS design, this method has the capacity to detect single candidate genes. One major drawback of GWAS is its inability to detect rare alleles because the statistical power of GWAS depends both on the effect of variation on the phenotype and the frequency of the variation in the population (Bazakos et al., 2017). Both QTL and GWAS approaches have massively benefited from the reduction in the cost of whole genome sequencing, such that the effect of variation can be assessed statistically for its effect on the phenotype throughout the genome, instead of relying on selected genetic markers.

1.2 Introduction to *C. elegans*

1.2.1 The life cycle of *C. elegans*

C. elegans is a free-living nematode that thrives in microbe-rich rotting vegetation in the wild (Schulenburg and Félix, 2017). In the lab, *C. elegans* is maintained on nematode growth medium (NGM) in petri dishes (Brenner, 1974). Its diet consists of a strain of *Escherichia coli* OP50 exclusively, a uracil auxotroph that does not grow well and therefore allows for clear observation of *C. elegans* (Stiernagle, 2006). It develops from a single-celled embryo to a young adult in ≈ 3 d at 20°C in the presence of food (Hall and Altun, 2008). Developmental speed decreases when animals are grown at 15°C and increases at 25°C. The life cycle is shown

in Fig. 1.2. *C. elegans* go through four larval developmental (L1 – L4) stages, in which a new cuticle is secreted and the old cuticle is shed. In the absence of food, L1 larvae undergo developmental arrest and can either develop into dauers, which are starvation and desiccation resistant alternative life forms or enter to L2 stage upon the reintroduction of food. *C. elegans* have a boom and bust life cycle in the wild due to the ephemeral nature of food and dauer larvae are likely to be the dispersal stage that aid migration to a new ecological niche.

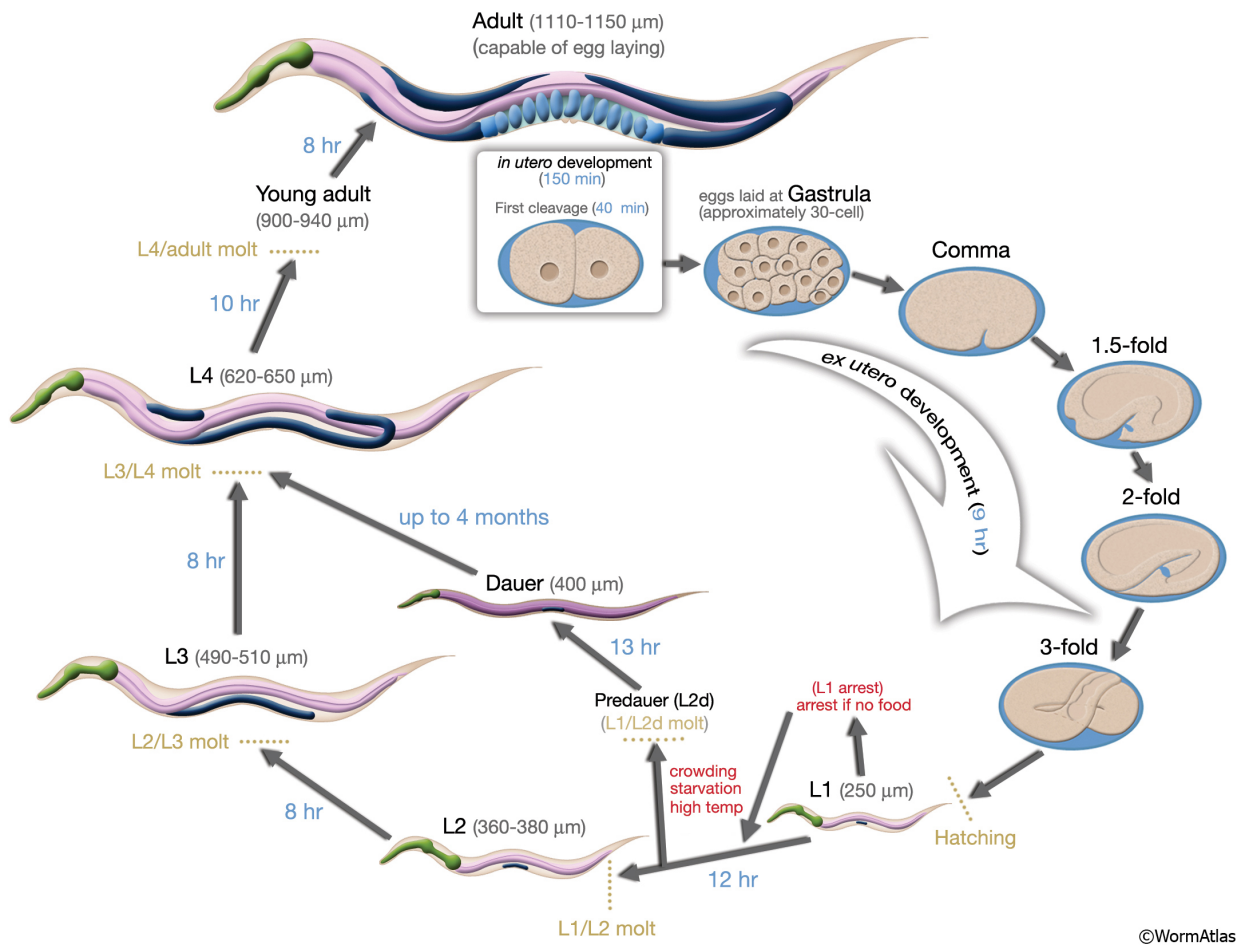


Figure 1.2: **Life cycle of *C. elegans* from an embryo to adult at 22°C (reproduced from WormAtlas).** Numbers in blue colour indicate the number of hours the animal spends in a particular larval stage. The developmental time for *C. elegans* from hatching to a young adult is 38 h in the presence of food. The number in the parenthesis next to the larval stage indicates the length of the animal.

1.2.2 *C. elegans* genome and genetics

The *C. elegans* genome is ≈ 100 Mb in size and is organised into five autosomes (I, II, III, IV, V) and one sex chromosome (X). Chromosomes are holocentric and have only a single crossover event per meiosis. *C. elegans* has $\approx 20\,127$ protein-coding genes (Wormbase release WS273,

Sep. 2019) and at least 38% (7663) have predicted orthologs in the human genome (Shaye and Greenwald, 2011). Many of these genes are involved in conserved signalling pathways like receptor tyrosine kinase (RTK), Notch, TGF- β , Wnt and insulin signalling pathways. Therefore, discoveries made from *C. elegans* research can have broader implications for human health and diseases. Common post-transcriptional modifications in *C. elegans* include trans-splicing and alternative polyadenylation. Approximately 70% of *C. elegans* mRNAs are trans-spliced, and about 15% of genes are organised in operons. A 22-nucleotide leader sequence (SL1) is added at the 5' end of mRNA. Genes residing in operons are transcribed as polycistronic mRNA transcripts wherein the first mRNA is mostly spliced to SL1 and downstream mRNAs to SL2 (Blumenthal, 2005). Ubiquitously transcribed genes harbour miRNA targets in their 3' UTRs and use alternative polyadenylation to fine-tune tissue-specific protein expression (Blazie et al., 2017).

C. elegans is an androdieocious nematode species consisting of two sexes, self-fertilising sequential hermaphrodites and males. *C. elegans* has XX/XO sex-determination system, wherein hermaphrodites have two sex chromosomes (XX), and males have one (XO). Both sexes are diploid for the five autosomes. Hermaphroditism in *C. elegans* is sequential. This is because hermaphrodites are anatomically females that evolved to produce sperm during the L4 stage and store them in their spermatheca. The hermaphrodites switch to oogenesis and produce ≈ 300 self-fertilised offspring. The number of progeny is limited by the number of hermaphrodite-derived sperm and can be increased to ≈ 1000 by mating with males. *C. elegans* is propagated as selfing hermaphrodites, which allows for maintenance of isogenic populations. The frequency of males in the lab is low at $\approx 0.1\%$ and these are produced spontaneously due to the non-disjunction of the X chromosome. Males are essential for genetic crosses to introduce new mutations or genetic variation into new genetic backgrounds.

1.2.3 Natural genetic variation in *C. elegans*

Overall *C. elegans* as a species has low genetic diversity owing to its selfing mode of reproduction. Its genetic diversity is 20-fold lower than in a gonochoristic species like *D. melanogaster* (Barrière and Félix, 2005b). Despite its low genetic diversity, *C. elegans* has enough natural genetic variation allowing its use for genetic mapping of mutations. CB4856 (a polymorphic

strain) and N2 differ by 327 050 (one polymorphism per 307 bp or 3.26 polymorphisms per kb on an average) single nucleotide variants (SNVs) and 79 529 insertions-deletions (indels). SNVs and indels are not uniformly distributed along the chromosomes, with the arms showing greater average variation compared to the centre due to higher recombination rate exhibited by chromosomal arms (Rockman and Kruglyak, 2009; O. A. Thompson et al., 2015). Interestingly, SNVs and indels are also not uniformly distributed across the six chromosomes with chromosome V harbouring highest total number of variants (O. A. Thompson et al., 2015). N2 is a laboratory reference strain that has undergone laboratory-aided evolution to acquire many beneficial nucleotide changes not found in wild isolates (Sterken et al., 2015). For example, wild isolates aggregate on the *E. coli* lawn due to their preference for low oxygen conditions found on the edge of the lawn. N2, in contrast, prefers higher oxygen concentration and to disperse across the bacterial lawn because of a gain-of-function allele (215V) in the neuropeptide receptor gene *npr-1* (Andersen, Bloom, et al., 2014). This variation (F215V) in the *npr-1* causes many pleiotropic effects on behaviour (aggregation, pathogen and heat avoidance), phenotypic trait (body size) and life-history traits (lifetime fecundity) (reviewed in Sterken et al. (2015)). *C. elegans* homolog of vertebrate N-acetyltransferase (Nat10) *nath-10* is another gene that involves laboratory-derived alleles in N2 that has pleiotropic effects on life-history traits such as fecundity, brood size and age at maturity (Duveau and Félix, 2012).

C. elegans has been, and continues to be sampled, extensively around the world by researchers in the community (Hodgkin and Doniach, 1997; Haber et al., 2005; Barrière and Félix, 2005a; Andersen, Gerke, et al., 2012; Crombie et al., 2019). *C. elegans* is a cosmopolitan species whose allelic frequencies are only weakly correlated to geographical origin (Sivasundar and Hey, 2003; Haber et al., 2005; Andersen, Gerke, et al., 2012). Moreover, there is no association between genetic and geographical distances (Sivasundar and Hey, 2003). The *Caenorhabditis elegans* Natural Diversity Resource (CeNDR, <http://www.elegansvariation.org/>) collects maintains and distributes wild isolates of *C. elegans* (Cook et al., 2017). 766 *C. elegans* strains are available from CeNDR, which fall into 330 isotypes (Fig. 1.3 and Fig. 1.3) that have been sequenced by whole genome sequencing technologies (WGS) according to CeNDR's latest release (20180527). There are 3 396 485 SNVs in the 330 strains sequenced that repre-

sents a resource for genotype-phenotype correlations. Interestingly, there is a hotspot of genetic diversity around geographically isolated Hawaiian Islands as strains isolated here do not share haplotypes with other isolates (Andersen, Gerke, et al., 2012; Cook et al., 2017). It is thought that the Hawaiian *C. elegans* population represents ancestral genetic diversity because they contain approximately three times more diversity than the non-Hawaiian population (Crombie et al., 2019).

1.3 Postembryonic development of hypodermis

1.3.1 Overview of Seam cell patterning

Seam cells are epidermal precursor cells that are born from the AB lineage in the *C. elegans* embryo (Sulston, Schierenberg, et al., 1983) and give rise to the epidermis of the worm. At the end of embryonic development, *C. elegans* hatches as an L1 larva without ventral hypodermis (note the absence of hypodermis (hyp7) on the ventral side in Fig. 1.4A, C). Instead, there is a single layer of P cells (1.4 4B), a subset of which give eventually give rise to the vulva. The dorsal side is covered with hyp7 cells (23 cells), born during embryogenesis and fused to form a syncytium that covers the excretory pore and the anus (Fig. 1.4A, B). Seam cells and P cells in L1 are in contact with each other as they have apical junctions between them (Fig. 1.4D). Apical junctions are a single structure that combines adhesive and barrier functions of adherens and tight junctions (Armenti and Nance, 2012). Apical junctions in *C. elegans* contain cadherin-catenin and DLG-AJM complexes which are thought to mediate cell adhesion and cell polarity redundantly (Michaux, Legouis, and Labouesse, 2001).

The L1 larva hatches with ten seam cells present bilaterally. These carry out divisions throughout postembryonic development except for H0 which does not divide. The first asymmetric division of seam cell in L1 stage occurs after 5 h of hatching (Austin and C. Kenyon, 1994) (Fig. 1.5). Upon asymmetric cell division, the anterior seam cell daughters differentiate, except in the case of H1 wherein the posterior seam cell daughter differentiates, to hyp7. The non-differentiating daughter cell remains a seam cell. V5 is typically the first seam cell to divide, and V2 – V4 divide before V1 (Gritti et al., 2016; Austin and C. Kenyon, 1994). V2 – V6 seam cells contribute hyp7 cells exclusively to the ventral side (Fig. 1.5B,C). They achieve

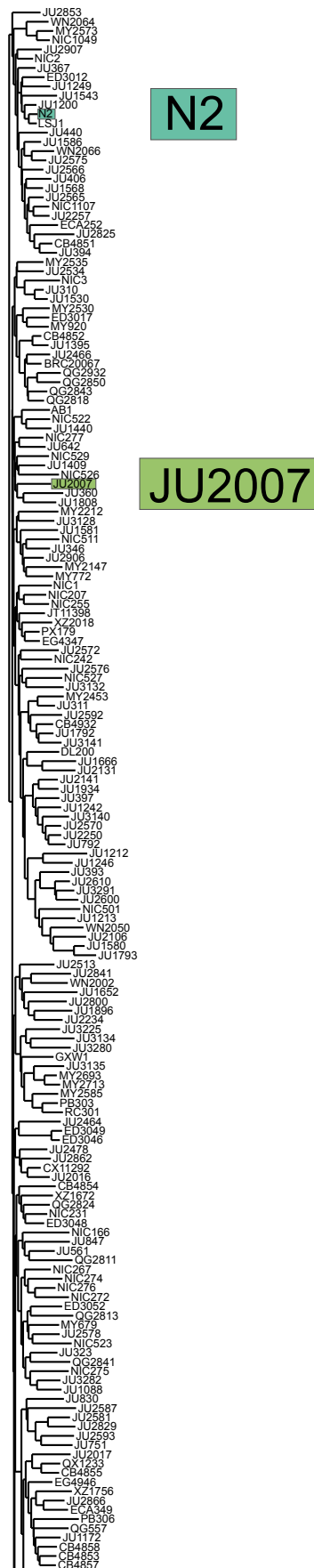


Figure 1.3: Genome-wide phylogeny of *C. elegans* wild isolates (adapted from CeNDR, <https://www.elegansvariation.org/data/release/latest>). Continued on next page.

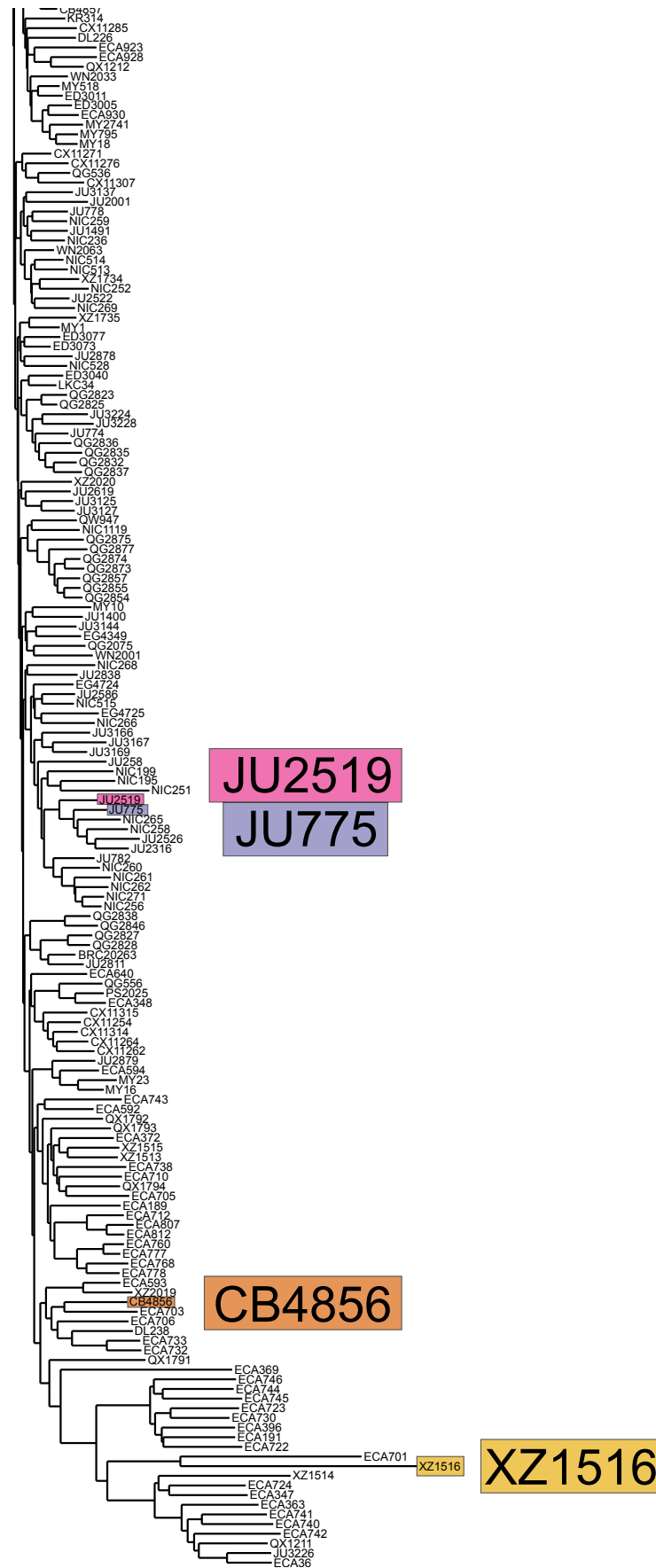


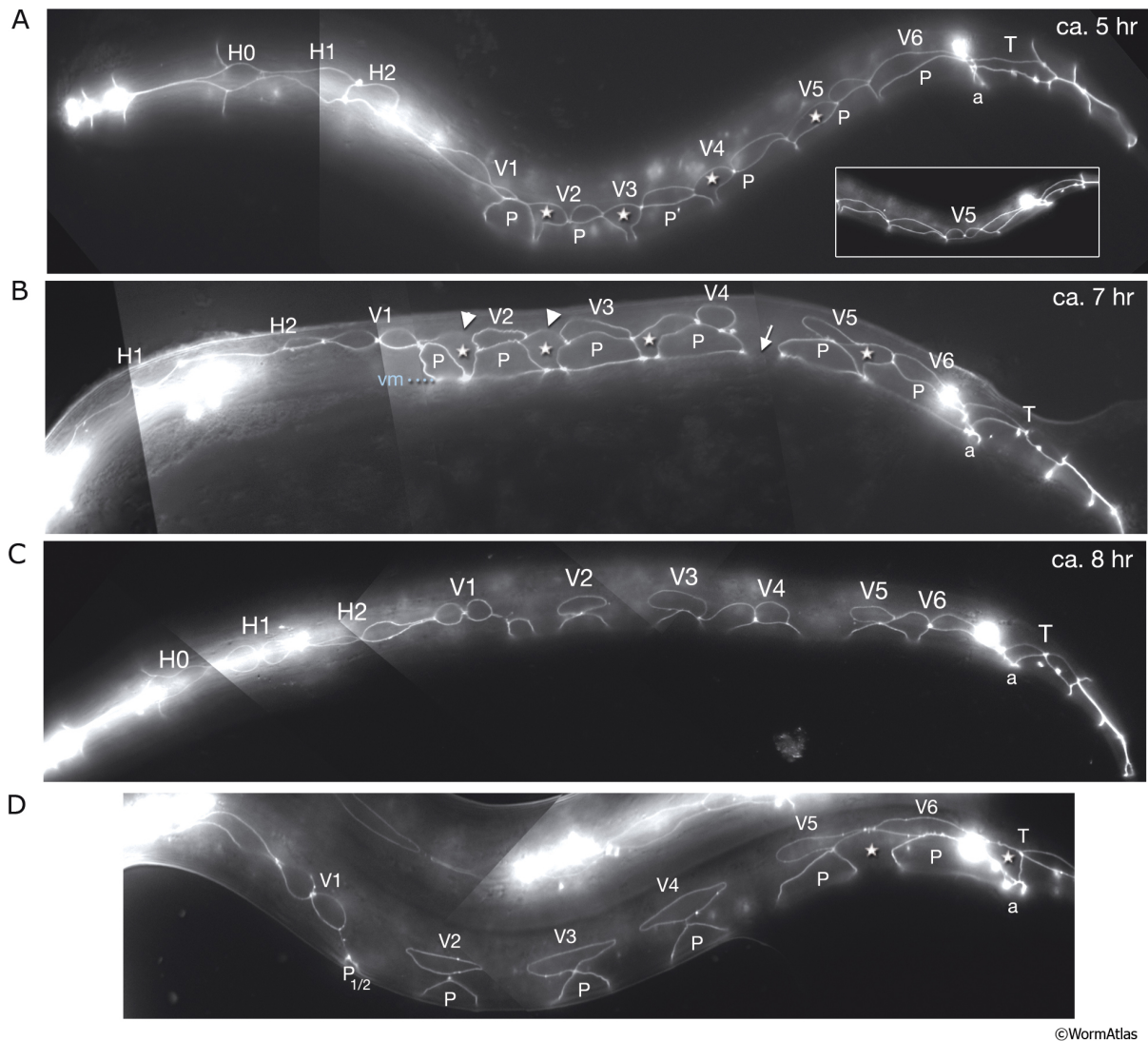
Figure 1.3: Genome-wide phylogeny of *C. elegans* wild isolates (adapted from CeNDR, <https://www.elegansvariation.org/data/release/latest>). Continued on next page.

Figure 1.3: **Genome-wide phylogeny of *C. elegans* wild isolates (Adapted from CeNDR, <https://www.elegansvariation.org/data/release/latest>).** Six strains (N2-Bristol, UK; JU2007-Isle of Wight, UK; JU2519-Lisbon, Portugal; JU775-Lisbon, Portugal; CB4856-Honolulu, USA; XZ1516-Kekaha, USA) that have been used in this thesis are highlighted. These isolates were selected because they are genetically divergent. N2 was originally isolated from Bristol, UK and is used as the laboratory reference strain. CB4856 is a highly polymorphic isolate that is used to map mutations in *C. elegans*. XZ1516 is the most divergent wild isolate.

this by sending cytoplasmic processes from the anterior part of the cells that open the apical junctions between the neighbouring P cells isolating P cell pairs and reach the ventral midline (Podbilewicz and J. G. White, 1994). The anterior seam daughters that will differentiate to hyp7 endoreduplicate their DNA and become tetraploid and start expressing hypodermal markers such as *dpy-7* and *elt-3* (Gilleard, Barry, and Johnstone, 1997; Hedgecock and J. G. White, 1985; Rijnberk et al., 2017; Yamamoto, Takeshita, and Sawa, 2011). Embryonically derived hyp7 cells remain diploid (Hedgecock and J. G. White, 1985; Rijnberk et al., 2017). The anterior seam daughters from the asymmetric seam cell division move ventrally, and the P cells reduce in size, lose contact with seam cells and descend into the ventral cord (Fig. 1.5B,C, D and Fig. 1E from Bone et al. (2016)) (Podbilewicz and J. G. White, 1994; Austin and C. Kenyon, 1994). These anterior seam cell daughters express the fusogen *eff-1* that dissolves the apical junctions and allows them to fuse to dorsal hyp7 (Fig. 1.5B, C, D). The posterior seam cell daughters express seam cell-specific markers such as *SCMp::GFP*, *nhr-73* and *egl-18* (Gorrepati, K. W. Thompson, and Eisenmann, 2013; Miyabayashi et al., 1999).

The seam cells H1, V1–V4 and V6, undergo symmetrical proliferative divisions in early L2 stage to increase the seam cell number to 16 per lateral side (depicted by horizontal green bars in the seam cell lineage in Fig. 1.6). During L2, V5 undergoes a unique asymmetric cell division, wherein the anterior cell produces a postdeirid neuroblast (PDE) and the posterior cell maintains a seam fate. Following these divisions, seam cells undergo three rounds of asymmetric cell divisions during the three larval stages (L2, L3 and L4). The anterior daughter cell differentiates to hyp7, and the posterior daughter maintains the seam cell fate except in the case of T, which undergoes a different pattern of asymmetrical cell divisions to produce neuronal cells and hyp7 cells.

In total, there are six distinct seam cell patterns (H0, H1, H2, V1 – V4 including V6, V5 and T) as shown in Fig. 1.6. At the end of postembryonic development, there are sixteen seam



©WormAtlas

Figure 1.5: **First L1 seam cell division in *C. elegans* visualised by *ajm-1::GFP* which marks the apical junctions (reproduced from WormAtlas).** Lateral view. P cells are labelled with letter “P”. (vm) ventral midline; (a) anus. (A) V5 is the first seam cell divide. V1 – V5 have started dividing and stars mark the anterior daughters. H1, H2, V6 and T have not yet divided. V2 – V6 send cytoplasmic processes from the anterior part of the seam cells isolating pairs of P cells. (a) anus. (B) Arrow points to the anterior daughter of V5 which has already fused with hyp7. Arrowheads point to disappearance of apical junctions of anterior seam daughters of V2 and V3 that are in the process of fusing to the dorsal hyp7. Stars depict anterior daughters that are migrating ventrally. H1, H2, V1 and T have divided as well. (C) Lateral view. Anterior daughters have completed fusing to dorsal hyp7. (D) Stars label anterior daughters of V6 and T are in the process of fusing to hyp7. P cells are isolated from their anterior and posterior neighbours. P1/P2 have become smaller and started migrating to the ventral midline.

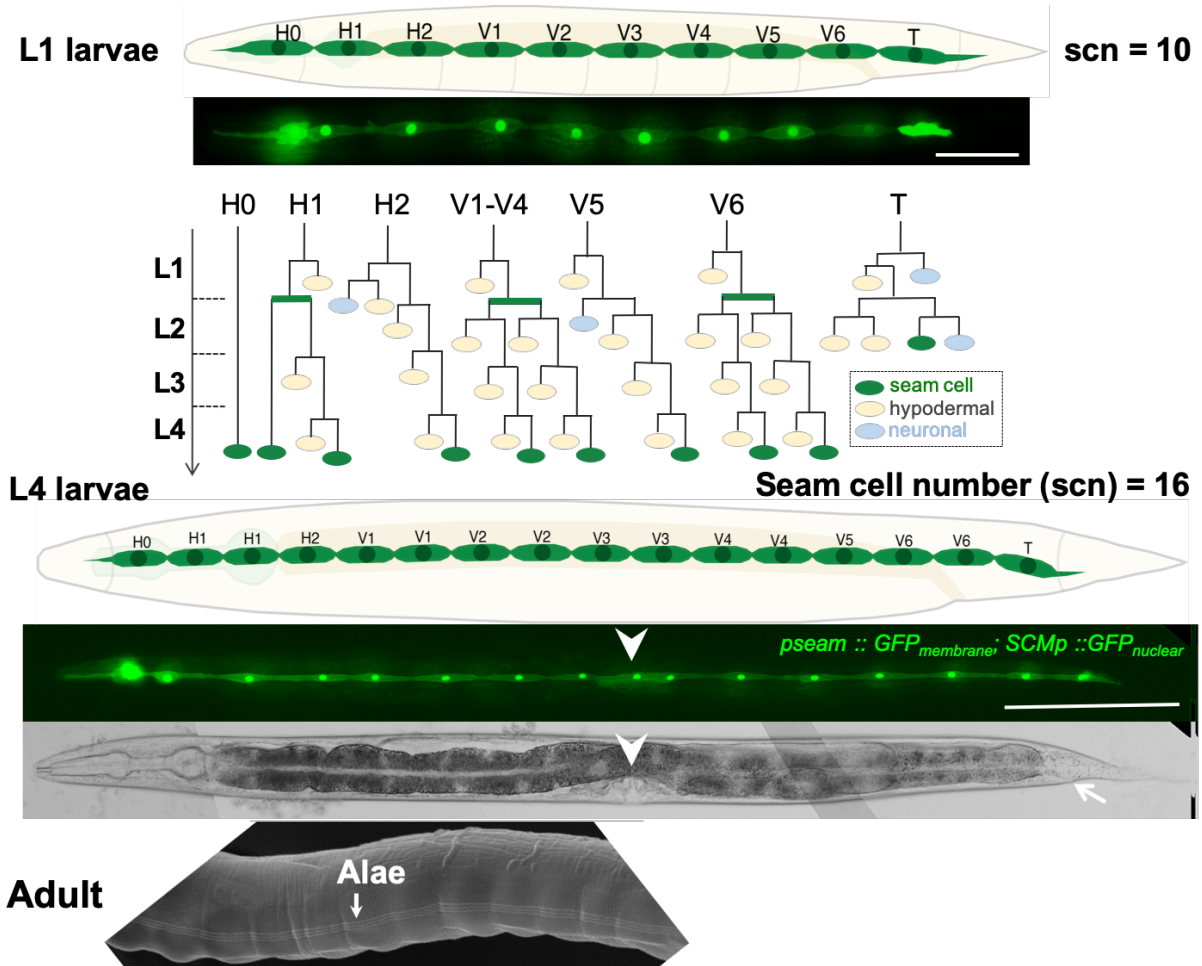


Figure 1.6: *C. elegans* seam cell lineage. Adapted from Katsanos et al. (2017). L1 larvae hatches with 10 seam cells. Seam cells undergo division patterns that can be grouped into 6 (H0, H1, H2, V1 – V4 incl. V6, V5, T) different lineages in *C. elegans*. Seam cells and *hyp7* are coloured green and yellow, respectively. Seam cells divide symmetrically and asymmetrically during the four larval stages. The proliferative symmetric division in L2 is shown by the thick green horizontal bars. At the end of L4 division, *C. elegans* hermaphrodite has 16 seam cells per lateral side at the end of the larval divisions. 8 seam cells (H0, H1a, H1p, V1a, V1p, V2a and V2p) are anterior to vulva, 2 seam cells (V3a and V3p) are over the vulva and 6 seam cells (V4a, V4p, V5, V6a and V6p) are posterior to the vulva. White arrowhead points to vulva. White arrow points to anus in the brightfield image of an L4 animal. After L4, seam cells fuse together and secrete alae (white arrow) in the adult. Scale bars in the fluorescent images are 100 μ m.

cells per lateral side in wild-type animals, and this number is robust to stochastic variation (Mestek Boukhibar and Barkoulas, 2016; Katsanos et al., 2017). After the completion of the last seam cell division in L4, seam cells connect and fuse to form a seam syncytium by expressing another fusogen *aff-1*. The seam syncytium secretes raised cuticular structures called alae in the adult whose function in the adult animal is thought to contribute to structural integrity and chemosensation (Fig. 1.6).

Seam cell division patterns and their positions in the worm are highly invariant (Fig. 1.6), which allows the possibility to detect errors in division patterns based on the position of seam cells. H0 is always present laterally to the first/anterior bulb (metacarpus) of the pharynx and H1a is located laterally to the isthmus of the pharynx. H1p is located posteriorly in relation to the pharynx. V6p and T are located posterior to the anus in L4. V3a and V3p are located closer to each other than they are to V2p or V4a respectively and positioned lateral to the vulva. H2, V1a, V1p, V2a and V2p are anterior to V3a. V4a, V4p, V5, V6a are posterior to V3p. Based on the stereotypical positions of seam cells, they can be binned into three categories relative to the vulva; anterior (H0, H1a, H1p, V1a, V1p, V2a and V2p), mid (V3a and V3p) and posterior (V4a, V4p, V5, V6a, V6p and T) seam cells, which has been used in this thesis as a way to narrow down the cell division errors in mutant backgrounds.

Posterior seam cell daughters lose contact with each other after every seam cell division due to the gaps left by differentiating anterior seam daughter cells that fuse to *hyp7* and move out of seam tissue (Austin and C. Kenyon, 1994; Podbilewicz and J. G. White, 1994). *plx-1*, a plexin, which acts as a receptor for *smp-1* and *smp-2* (transmembrane-type semaphorins) is known to be involved in regulating these cell contacts. In *plx-1* mutants, seam cells make inappropriate dorsoventral contacts with other seam cells, and gaps are found in the seam tissue and alae (Fujii et al., 2002). Seam cells have altered positions or orientations in L1 larvae in *smp-1* and *smp-2* mutants as well, suggesting that the plexin-semaphorin system is essential for correct orientation and positioning of seam cells (Ginzburg, P. J. Roy, and Culotti, 2002). Seam cells make ectopic contacts with other seam cells in L1 larvae in *mab-20* (secreted semaphorin) mutants, suggesting that it may prevent or repel cell extensions or cell exploratory processes (P. J. Roy et al., 2000).

1.4 Genetic control of postembryonic development

Several distinct pathways and transcription factors regulate the division pattern of the seam (reviewed by R. Nimmo and Woollard (2008); R. A. Nimmo and Slack (2009); Joshi et al. (2010); Sawa and Korswagen (2013); Lam and Phillips (2017)). These pathways are briefly summarised below.

1.4.1 Heterochronic pathway regulates the temporal seam cell division patterns

The *lin-4* and *let-7* family of miRNAs belong to the heterochronic pathway and regulate the temporal pattern of seam cell divisions through different larval stages (Lee, Feinbaum, and Ambros, 1993; Wightman, Ha, and Ruvkun, 1993; Reinhart et al., 2000). miRNAs inhibit gene expression post-transcriptionally and act as developmental switches. miRNAs like *lin-4* and *let-7* were first discovered in *C. elegans* to control developmental timing of the seam cell divisions by targeting heterochronic genes such as *lin-14*, *lin-28*, *lin-41*, *hbl-1*, *daf-12*, and *lin-29* (Lee, Feinbaum, and Ambros, 1993; Reinhart et al., 2000). *lin-4* causes L1/L2 and L2/L3 switches by the downregulation of *lin-14* and *lin-28*, respectively (Lee, Feinbaum, and Ambros, 1993; Moss, Lee, and Ambros, 1997). Mutations in *lin-4* or *let-7* lead to the reiteration of larval patterns of cell division, causing seam cell hyperplasia, whereas mutations in their targets *lin-14* and *lin-28* lead to premature terminal differentiation (R. A. Nimmo and Slack, 2009). Therefore, *lin-4* and *let-7* promote differentiation and inhibit self-renewal by downregulating their targets *lin-14* and *lin-28*. In the L4 larval stage, the terminal fusion of seam cells is controlled by the expression of the transcription factor *lin-29*, which activates the expression of a fusogen *aff-1* in seam cells (Friedlander-Shani and Podbilewicz, 2011). *let-7* downregulates *lin-41* expression, which removes the inhibition of *lin-29* (Slack et al., 2000). *let-7*, *lin-4* / *mir-125* miRNA families and other members of the heterochronic pathway like *lin-28* and *lin-41* are also conserved in mammalian development (R. A. Nimmo and Slack, 2009).

1.4.2 Transcription factors involved in seam cell patterning

Many families of transcription factors have been shown to play a role in the seam cell gene regulatory network and show functional conservation with factors involved in mammalian stem cell pathways. First, *C. elegans* homologs of Runx transcription factor (*rnt-1*) and its binding partner CBF β binding partner of Runx (*bro-1*) promote the proliferative/symmetric seam cell division in L2, partly by inhibiting the negative regulator of the cell cycle *cki-1* (R. Nimmo, Antebi, and Woollard, 2005; Hiroshi Kagoshima et al., 2005; Kagoshima et al., 2007). Overexpression of *rnt-1/bro-1* leads to seam cell hyperplasia. Mutations in Runx genes and CBF β are known to cause various leukaemias in humans suggesting a conserved function in stem cell proliferation (Cameron and Neil, 2004). Thus, Runx genes and CBF β play a role in both cell proliferation and differentiation depending on the context (Coffman, 2003; Xia et al., 2007; R. Nimmo and Woollard, 2008). Pbx and Meis transcription factors *ceh-20* and *unc-62* are thought to act upstream of *rnt-1* to repress seam cell proliferation in the anterior differentiating daughter (Hughes et al., 2013).

Second, *ceh-16* (homolog of *Engrailed*) encodes a homeodomain transcription factor that is required for proper specification and differentiation of the lateral seam cells and is also required for the symmetric division of L2 (Huang et al., 2009). *ceh-16* is thought to function in parallel to *rnt-1/bro-1* to promote proliferative divisions in L2 (Joshi et al., 2010). Huang et al. (2009) showed that human ortholog of *engrailed/En2* can rescue the loss-of-function of *C. elegans engrailed/ceh-16*. *En2* rescues *ceh-16* loss-of-function by promoting seam cell proliferation, and overexpression of *En2* or *ceh-16* causes seam cell hyperplasia. *ceh-16* has been shown to maintain seam cell fate by suppressing fusion of seam cells and is also required for expression of seam cell fate markers, such as *egl-18* and seam-specific genes in the embryo (Cassata et al., 2005).

Third, *nhr-25* (ortholog of Fushi tarazu transcription factor 1 and nuclear receptor subfamily 5 group A member 1) encodes a transcription factor that has pleiotropic effects on development, moulting, and reproduction (Chen, Eastburn, and Han, 2004; Ward et al., 2013). The loss of *nhr-25* function in seam cells leads to rounded seam cells and loss of cell-cell contacts. This loss of contacts between seam cells is thought to result in aberrant division patterns causing

gaps in the seam tissue and alae (Silhánková, Jindra, and Asahina, 2005; Chen, Eastburn, and Han, 2004).

Fourth, GATA transcription factors that coordinate the development of diverse tissues are evolutionarily conserved in animals, fungi and plants (Tremblay, Sanchez-Ferras, and Bouchard, 2018). In *C. elegans*, 11 GATA transcription factors play crucial roles in the development of the gut, epidermis and vulva (Block and Shapira, 2015). *elt-1*, is an ortholog of human GATA1 and specifies epidermal fate in *C. elegans*. In seam cells, it directly regulates *bro-1* to promote proliferation and maintains seam cell fate by repressing fusion of the seam cells (Brabin, Appleford, and Woollard, 2011). The GATA factor *elt-3* is expressed exclusively in differentiating anterior seam cell daughters and *hyp7*. A conserved transcription family of *Hes* basic helix-loop-helix (bHLH) transcription factors play important roles in development by regulating cell fate decisions to balance stem cell proliferation and differentiation (Kageyama, Ohtsuka, and Kobayashi, 2007; L. A. Wrischnik and C. J. Kenyon, 1997). In *C. elegans*, variability in seam cell number is increased due to stochastic loss and gain of symmetric cell divisions in animals carrying mutations in *lin-22* (Katsanos et al., 2017). While multiple molecular factors have been identified, but how these interact into specific pathway driving stem cell development and differentiation remains largely unknown.

1.4.3 Wnt/ β -catenin asymmetry regulates the polarity of asymmetric seam cell divisions

The Wnt/ β -catenin asymmetry (W β A) pathway is an important pathway that regulates the asymmetric cell division of seam cells along the anteroposterior axis (Mizumoto and Sawa, 2007a; Kanamori et al., 2008; Gleason and Eisenmann, 2010; Yamamoto, Takeshita, and Sawa, 2011). This noncanonical Wnt pathway is dependent on the divergent β -catenins *wrm-1* and *sys-1* and the conserved β -catenin/*bar-1* (Kidd et al., 2005; Rocheleau, Downs, et al., 1997; Mizumoto and Sawa, 2007b). In the canonical Wnt pathway, in the absence of Wnt ligands, β -catenin is targeted by destruction complex consisting of Adenomatous Polyposis Coli (APC)/*apr-1*, Axin/*pry-1*, glycogen synthase kinase-3 β (GSK-3 β)/*gsk-3* and casein kinase 1 (CK1 α)/*kin-19*. The destruction complex phosphorylates β -catenin targeting it for degradation through the ubiquitin proteasome pathway. In the opposite scenario, upon Wnt receptor

activation, β -catenin is stabilised, and it enters the nucleus, and along with transcription factor T-cell factor/lymphoid enhancer factor (TCF/LEF)/*pop-1* activates Wnt target genes (reviewed by Sawa and Korswagen (2013)).

In contrast to the canonical Wnt pathway, the noncanonical Wnt pathway relies on the asymmetric distribution of Wnt pathway components rather than solely on the stabilisation of β -catenin (Mizumoto and Sawa, 2007b; Mizumoto and Sawa, 2007a). There are multiple lines of evidence for an intrinsic mechanism for generating polarity in seam cells. First, the seam cells remain polarised (undergo asymmetric cell division) along the anteroposterior axis despite the lack of Wnt ligands suggesting a permissive role, not an instructive role for Wnt ligands (Yamamoto, Takeshita, and Sawa, 2011). Five Wnt ligands (*cwn-2*, *cwn-1*, *mom-2*, *egl-20*, *lin-44*) are expressed along the anteroposterior axis in partially overlapping regions (Harterink et al., 2011). Second, loss of Wnt receptors Frizzled/*lin-17*/*mom-5* and Ror/*cam-1* convert asymmetric divisions to symmetric divisions suggesting that they generate polarity (Yamamoto, Takeshita, and Sawa, 2011). Third, many Wnt components are asymmetrically localised. The Wnt receptors Frizzled/LIN-17 and dishevelled/DSH-1 are asymmetrically localised to the posterior cortex (Takeshita and Sawa, 2005; Mizumoto and Sawa, 2007a). Negative $W\beta A$ regulators (APR-1 and PRY-1) are asymmetrically localised to anterior cortex, whereas positive $W\beta A$ regulators (Frizzled/MOM-5, Dishevelled/DSH-2 and Dishevelled/MIG-5) are asymmetrically localised to the posterior cortex (1.7). Loss of negative regulators of $W\beta A$ can lead to symmetrisation of asymmetric seam cell divisions. Knockdown of $CK1\alpha$ /*kin-19* and APC/*apr-1* or Axin/*pry-1* leads to increase in seam cell number (Banerjee et al., 2010; Gleason and Eisenmann, 2010). Wnt ligands may regulate the anterior cortical localisation of negative regulators APR-1 and PRY-1 as lack of EGL-20 abrogates this asymmetry (Mizumoto and Sawa, 2007a). APR-1 becomes expressed in the posterior cortex in Dishevelled/*mig-5* mutants suggesting that *mig-5* regulates APR-1 (Baldwin, Clemons, and Phillips, 2016).

TCF/POP-1 acts as a transcriptional repressor when present at a higher level in the nucleus and as a transcriptional activator when present at a lower level, depending also on other factors that contextualise POP-1 function (Shetty et al., 2005). Depletion of *pop-1* leads to an increase in seam cell number at the expense of hyp7 cell fate (Gleason and Eisenmann, 2010). Upon $W\beta A$

activation, the ratio of SYS-1/POP-1 is higher in the posterior seam cell daughter and lower in the anterior seam cell daughter. When SYS/POP-1 ratio is high, SYS-1 binds to POP-1 and together they activate Wnt target genes (Banerjee et al., 2010; Gleason and Eisenmann, 2010; Mizumoto and Sawa, 2007a). During telophase, asymmetry in a seam cell is established by APR-1/APC, which promotes nuclear export of WRM-1/ β -catenin by stabilising microtubules in the anterior cortex (Sugioka, Mizumoto, and Sawa, 2011). LIT-1/NLK and WRM-1/ β -catenin are localised preferentially to the anterior cortex during cell division and to the posterior nucleus after cell division in the seam cells (Takeshita and Sawa, 2005; Kanamori et al., 2008). The asymmetrical distribution of SYS-1/ β -catenin is regulated by the destruction complex as demonstrated by symmetrical distribution of SYS-1 upon knockdown of *kin-19* or *apr-1* (Baldwin and Phillips, 2014). In the posterior nucleus, WRM-1/LIT-1 complex phosphorylates POP-1, which leads to the displacement of transcriptional repressor POP-1 from the nucleus increasing the SYS-1/POP-1 ratio, which ultimately leads to transcription of Wnt target genes (Rocheleau, Yasuda, et al., 1999; Lo et al., 2004).

W β A pathway activates *egl-18*, a GATA transcription factor, which specifies seam cell fate during asymmetric seam cell divisions. POP-1 binds to a site in the promoter of *egl-18* in vitro that is required for seam cell expression of *egl-18* (Gorrepati, K. W. Thompson, and Eisenmann, 2013). *egl-18* and its paralog *elt-6* are expressed in the seam after both symmetric and asymmetric cell divisions and act redundantly to regulate cell fates and fusion in the vulva and seam cells (Koh and Rothman, 2001; Koh, Peyrot, et al., 2002; Eisenmann and S. K. Kim, 2000). Elimination of both *egl-18* and *elt-6* results in misexpression of the hypodermal marker *elt-3* in seam cells and fusion with the hypodermis and developmental arrest of L1 larvae (Koh and Rothman, 2001). In *egl-18* mutants, there is a loss of seam cells resulting in fewer seam cells and lack of functional vulva. Furthermore, the proliferation of seam cells through overactivation of Wnt signalling by *pop-1* RNAi is dependent on *egl-18* and animals with overactivated Wnt signalling are sensitised to the loss of *egl-18* (Gorrepati, K. W. Thompson, and Eisenmann, 2013). Loss or silencing of GATA transcription factors has been observed in many aggressive breast, colorectal and lung cancers (Zheng and Blobel, 2011) suggesting a conserved role in proliferation.

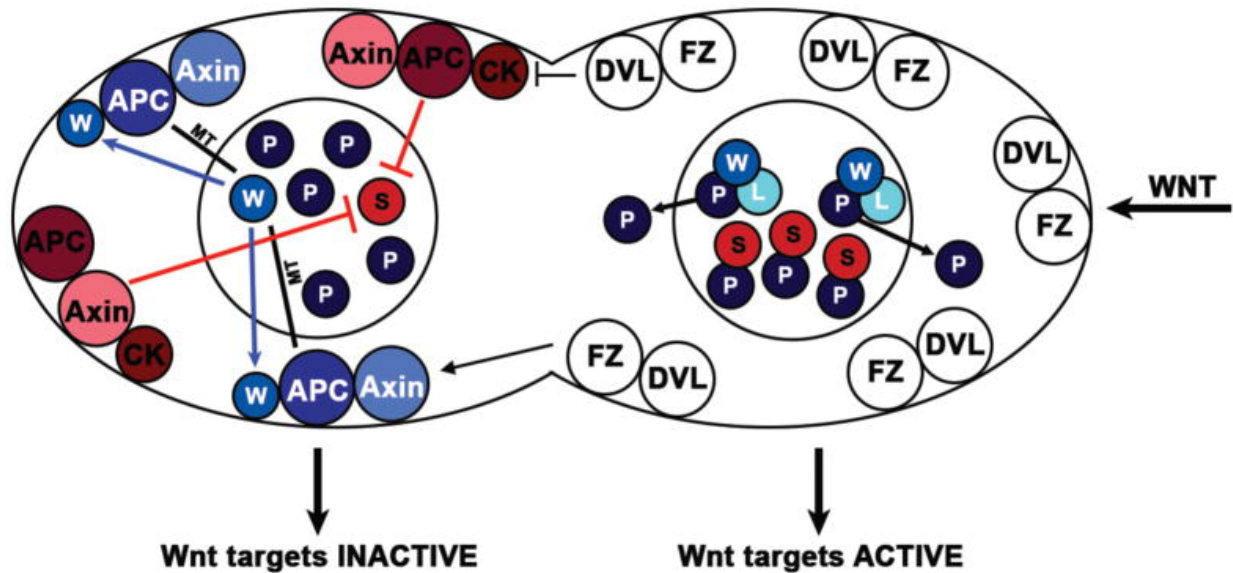


Figure 1.7: A model of asymmetric cell division controlled by $W\beta A$ pathway (reproduced from Lam and Phillips (2017) and originally adapted with permission from Baldwin, Clemons, and Phillips (2016)). Note the asymmetric distribution of negative Wnt regulators APC and Axin/PRY-1 to the posterior cortex and positive $W\beta A$ regulators Frizzled/MOM-5 (FZ), Dishevelled/DSH-2 (DVL) to the anterior cortex. Wnt components from β -catenin/WRM-1 (W) regulating and β -catenin/SYS-1 (S) regulating pathways are colour-coded blue and red as, respectively. Following cell division, the two separate pathways are balanced by Dishevelled and they generate transcriptional asymmetry through two pools of APC. Frizzled and Dishevelled function in the posterior cortex restrict cortical activity of localisation of APC and Axin. In the anterior cortex, microtubules (MTs) are stabilised by one pool of APC in the anterior cortex to promote WRM-1 (W) export out of the nucleus, which leads to high levels of TCF/POP-1 (P). SYS-1 (S) is degraded by another pool of APC with $CKI\alpha$ /KIN-19 (CK). This leads to low SYS-1/POP-1 ratio and Wnt target genes are repressed. In the posterior daughter, WRM-1 activates LIT-1 (L) and together they phosphorylate POP-1 (P). As a consequence, SYS-1/POP-1 ratio is high and together they activate Wnt target genes.

Finally, the heterochronic pathway interacts with the $W\beta A$ pathway. For example, *lit-1* encoding Nemo-like kinase in the $W\beta A$ pathway also regulates the temporal fate of seam cells, perhaps by silencing genes in the heterochronic pathway (Ren and H. Zhang, 2010). More recently it was shown that in the absence of Axin/PRY-1, *lin-4* and *let-7* miRNA family members were upregulated in both *C. elegans* and *C. briggsae* suggesting a conserved role in maintaining wild-type expression of heterochronic miRNAs (Mallick, Ranawade, and Gupta, 2019). Transcription factors like *lin-22* have been shown to antagonise Wnt signalling in seam cells as loss of *lin-22* leads to stochastic activation of Wnt signalling (Katsanos et al., 2017). *rnt-1/bro-1* pathway is thought to act parallel to or downstream of $W\beta A$ pathway because double knockdown of *rnt-1/bro-1* and *pop-1* reduces the number of extra seam cells compared to knockdown of *pop-1* alone (Gleason and Eisenmann, 2010).

1.4.4 Role of cell fusion in seam cell development

In *C. elegans*, one-third of the somatic cells generated during development fuse (Podbilewicz and J. G. White, 1994; Podbilewicz, 2000). As with the cell lineage, the pattern of the cell fusions is highly invariant, and 300 cells reside in 44 multinucleated syncytia (Shemer and Podbilewicz, 2000; Oren-Suissa and Podbilewicz, 2007; Alper and Podbilewicz, 2008). The epidermis is made of 8 syncytia containing 186 nuclei (Hedgecock and J. G. White, 1985). Out of these, *hyp7* is the largest syncytium in the epidermis containing 139 nuclei (Oren-Suissa and Podbilewicz, 2010). Two fusogens, *eff-1* and its paralog *aff-1*, encode nematode specific type I transmembrane proteins that are required for most of the cell fusion events in *C. elegans* (Mohler et al., 2002; Sapir et al., 2007). EFF-1 has structural homology to class II viral fusogens. No homologous genes of *eff-1* or *aff-1* have been identified outside the phylum Nematoda (Sapir et al., 2007; Shemer and Podbilewicz, 2003; Mohler et al., 2002). Despite this, cell fusion mediated by different viral derived fusogens has a conserved role in development, for example playing a role in muscle fibre formation and fusion of epithelial cells in the placenta in humans (Segev, Avinoam, and Podbilewicz, 2018).

Fusion seems to originate close to the apical junction and expand basally, which can be visualised by the disappearance of apical junctions labelled with *AJM-1::GFP* in the seam (Fig. 1.5B,C,D). However, this is not the case in all cell fusions. For example, dorsal and ventral cells

fuse in the absence of apical junctions in the contact zone and fusion of myoepithelial cells in the pharynx occurs without disassembling apical junctions (Podbilewicz and J. G. White, 1994; Shemer, Suissa, et al., 2004).

Most of the dorsal hyp7 cells fusion along with some ventral cells fuse due to EFF-1 expression during embryonic elongation. The loss of *eff-1* results in short and dumpy animals due to failure to elongate in embryogenesis caused by the lack of fusion of dorsal hyp7 cells (Shinn-Thomas et al., 2016). Ectopic expression of *eff-1* and *aff-1* in cells where it is not typically expressed causes fusion between cells not normally fated to fuse and causes a lethal phenotype (Shemer, Suissa, et al., 2004; Campo et al., 2005; Sapir et al., 2007). Several inter tissue fusion-barriers which prevent inappropriate fusion have been observed, which correspond to positions of basal laminae to surround different tissues such as pharynx, body wall muscle and hypodermis (Campo et al., 2005). Also, there is tight transcriptional regulation of these fusogens in a tissue-specific manner to avoid inappropriate fusion as *aff-1* and *eff-1* are potent fusogens. In the seam GATA transcription factors *elt-1* and *egl-18* are known to repress *eff-1* in posterior seam cells as seam cells fuse to hyp7 upon knockdown of *elt-1* or *egl-18* (Koh and Rothman, 2001; Brabin, Appleford, and Woollard, 2011). *eff-1* is ectopically expressed in lateral seam cells in *ceh-16* mutants and upon knockdown of *ceh-16* suggesting that it represses *eff-1* in the embryo (Cassata et al., 2005).

Both fusing cells must express EFF-1 in order to fuse. EFF-1 expression in hypodermal cells is dynamic and is actively removed from the plasma membrane through a DYN-1 (dynamin) and RAB-5 dependent endocytosis. Downregulation of *dyn-1* or *rab-5* leads to accumulation of EFF-1 on the membranes of cells causing precocious cell fusion, suggesting that endocytosis negatively regulates cell fusion (Smurova and Podbilewicz, 2016). It was observed that EFF-1 and F-actin are enriched at the cortex of anterior seam cell daughters. In addition, VAB-10 (spectraplaklin) was found to link EFF-1 to the actin cytoskeleton, and this was reinforced by EFF-1, causing a feedback loop. Furthermore, WASP and Arp2/3 dependent actin polymerisation is required for recruitment of EFF-1 to fusion sites as is evidenced by the delay in cell fusion in conditional mutants of WASP and Arp2/3 (Yang et al., 2017).

In the epidermis, *eff-1* and *aff-1* are utilised independently. *eff-1* is essential for fusion

of the anterior seam cell daughters after asymmetric cell division to the hypodermis (*hyp7* syncytium) during larval development, while *aff-1* is required for the fusion of seam cells in L4 larval stage, which is associated with terminal differentiation. An example of tissue where both *eff-1* and *aff-1* are required is the formation of the wild-type tail spike. A pair of tail spike cells fuse due to action of AFF-1 in the *C. elegans* embryo and produce a microtubule-filled process towards the tail tip (Chiorazzi et al., 2013). After the formation of the tail tip, the binucleate cell undergoes apoptosis by the transcriptional induction of *ced-3* caspase (Sulston, Schierenberg, et al., 1983; Chiorazzi et al., 2013). EFF-1 is required for the clearing of the cell-process through phagocytosis (Ghose et al., 2018). Failure of this phagocytosis leads to 100% of *eff-1* mutant hermaphrodites displaying a bulbous instead of the whip-like tail.

In all asymmetric seam cell divisions, anterior daughters lose expression of *SCMp::GFP*, and start expressing EFF-1 and fuse to *hyp7* shortly after being born (Gattegno et al., 2007; Campo et al., 2005). EFF-1 was enriched at fusion sites within $155 \text{ min} \pm 20 \text{ min}$ after the anterior seam cell daughter was born (Yang et al., 2017). Upon the breakdown of the cell membrane during cell fusion, it is thought that differentiation signals from the surrounding *hyp7* enter the anterior seam cell daughters causing them to switch to expression of the epidermal marker *dpy-7::yfp* (Brabin, Appleford, and Woollard, 2011; Brabin and Woollard, 2012).

In *eff-1* mutants, anterior seam cell daughters do not fuse to *hyp7*, fail to form a syncytia and persist laterally between posterior seam cells throughout larval development (Mohler et al., 2002). They maintain apical junction marker *AJM-1::GFP* and fail to form a fragmented epidermis (Mohler et al., 2002; Podbilewicz, 2006; Brabin, Appleford, and Woollard, 2011). The seam daughter cells are thought to not reconnect properly with their neighbouring seam cells and, as a result have breaks in the seam line and alae (Brabin, Appleford, and Woollard, 2011; Shemer, Suissa, et al., 2004). It has been proposed that anterior seam daughters may not differentiate in the absence of EFF-1 and are stuck in developmental limbo suggesting an important role in anterior seam cell daughter differentiation for *eff-1* (Brabin and Woollard, 2012).

1.5 *C. elegans* seam cell development as a model system for studying developmental robustness

C. elegans is an excellent model for experimental studies of developmental robustness and discovering CGV. Its development is highly invariant, and its post-embryonic lineage has been fully mapped (Sulston and Horvitz, 1977). *C. elegans* are self-fertilizing hermaphrodites with a short generation time and can be maintained in large numbers, which increase the statistical power for quantitative genetics. *C. elegans* is easy to culture and can be maintained under constant laboratory conditions to reduce the effect of environment on robustness. The animals are isogenic, which is ideal for minimising the effect of background genetic variation in studies of robustness.

Furthermore, there are genetically distinct *C. elegans* wild isolates with annotated whole genome sequences, which can be used to dissect the genetic basis of traits. In addition, *C. elegans* can be frozen and thawed periodically to avoid laboratory adaptation which can confound the effect of external perturbations. Therefore, stochastic noise from gene expression and introduced genetic mutations are the most common sources of variation in the experimental study of robustness in *C. elegans* (Maduro, 2015). In this thesis, I use seam cells as a model for studying developmental robustness. Seam cells can easily be labelled with fluorescent markers and visualised and counted under a fluorescent microscope. Seam cells have a predictable lineage, and there is a reproducible final number of 16 seam cells at the end of development. As discussed above, these lateral epidermal cells are stem cell-like in their ability to self-renew and produce differentiated neural and epidermal cells, and the components in the seam cell gene network have human orthologs. Several signalling pathways (Wnt/Wingless, TGF- β , RTK/Ras/MAPK, Insulin and Notch) and developmental regulators (Hox) are conserved between *C. elegans* and *Homo sapiens*, which makes the findings from *C. elegans* development potentially relevant to understanding human disease as well (Shaye and Greenwald, 2011). Furthermore, the genetic tractability of *C. elegans* combined with a completely sequenced and annotated genome and a toolbox of experimental techniques like RNAi, CRISPR-Cas9 genome editing, single molecule mRNA fluorescent in situ hybridisation (smFISH) and transgenesis makes it a particularly

powerful model organism.

1.6 Aims of this research

The overall objective of my doctoral research was to understand the mechanisms and consequences of developmental robustness in multicellular organisms. My thesis focused on the highly invariant seam cell number in the model organism *C. elegans*. Identification of genes promoting robustness and their characterisation represents the first step towards understanding mechanisms of robustness. Robustness in development allows genetic variation to accumulate. Therefore, my work expanded into detecting conditionally neutral genetic variation, also known as cryptic genetic variation using quantitative genetic approaches. The ultimate goal was to identify the molecular nature and genetic architecture of cryptic genetic variation. This dissertation provides a framework for investigating the mechanisms and consequences of robustness in a highly tractable multicellular developmental model.

In Chapter 3, I present the results from a phenotypic variance-based forward genetic screen aimed at identifying robustness conferring genes. I identified *eff-1* as a modulator of seam cell number variance, which I use throughout this thesis as a readout of seam cell patterning robustness. I also studied the phenotypic consequences of the loss of *eff-1* function on seam cell morphology and epidermal cell differentiation. Based on long-term time-lapse imaging and molecular genetics experiments, I discuss the developmental basis of the increase in variance of seam cell number. Chapter 3 highlights how the lack of *eff-1* does not directly affect seam cell fate and how the presence of it helps buffer seam cell number.

I describe the effect of genetic variation, which is naturally present in wild isolates of *C. elegans*, on seam cell development in Chapter 4. I first used environmental perturbations, such as higher growth temperature, to reveal cryptic genetic variation affecting seam cell development. I also employed a combination of CRISPR-Cas9 mediated genome editing and genetic introgressions to produce mutations in known seam cell regulators to explore how the genetic background may influence the outcome of these mutations. My results presented in this chapter highlight both genotype by genotype ($G \times G$) and genotype by environment ($G \times E$) interactions affecting seam cell development.

The aim of the work described in Chapter 5 was to map the genetic basis of the differential expressivity of *egl-18(ga97)* mutation between N2 (Bristol) and CB4856 (Hawaii) strains using a quantitative genetics approach. I present the identification of multiple quantitative trait loci affecting *egl-18(ga97)* mutation expressivity. By producing near isogenic lines, I studied the contributions of individual quantitative trait loci or in combination to seam cell number. I also discuss the genetic architecture underlying seam cell gene network. I was able to narrow down the large genomic intervals in the quantitative trait loci containing thousands of genes to smaller genomic intervals containing fewer genes. This chapter proposes for further investigation specific candidate genes potentially underlying the quantitative difference, based on RNAi screens and the published literature.

Finally, in the general discussion (Chapter 6), I bring these findings together to suggest that there are intercellular and intracellular process that underlie robustness of *C. elegans* seam cell development. I argue that *eff-1* contributes to developmental robustness of seam cell patterning and fusion is not required for cell differentiation. I discuss the cryptic genetic variation affecting seam cell development discovered through different perturbations and propose future experiments to map it. Finally, I propose a developmental model by which candidate genes may buffer seam cell number in *egl-18* loss-of-function mutants.

Chapter 2

Materials and Methods

2.1 General methods used in *C. elegans*

2.1.1 Maintenance

C. elegans was maintained monoxenically on a lawn of *Escherichia coli* strain OP50, an uracil auxotroph whose growth is limited, which allows for easier observation of *C. elegans*. The OP50 was seeded on Nematode Growth Medium (NGM) in 60 mm diameter petri dishes (Brenner, 1974). Animals were either picked using a pick made of platinum wire fused at the end of a glass pipette or by transferring a piece of agar containing animals with a scalpel. Both the platinum wire and the scalpel were sterilised under a flame before and after each use to avoid contamination. All *C. elegans* strains, unless stated otherwise, were propagated during the course of the experiments at 20 °C in an incubator. All the buffers, M9 buffer, freezing and bleaching solutions were prepared according to standard protocols (Stiernagle, 2006). *C. elegans* larvae were synchronised by bleaching gravid hermaphrodites and washing the eggs twice with M9 buffer and placed on NGM plates seeded with OP50 bacteria. *C. elegans* strains were obtained from Caenorhabditis Genetics Center (CGC), a *C. elegans* repository in St. Paul, MN, USA. All strains used (listed in Appendix A.1) are in the N2 Bristol background unless stated otherwise.

2.1.2 Cryopreservation

A strain that needed to be frozen was synchronised by transferring a piece of agar containing many animals onto 6 × petri dishes (55 mm diameter) or 3 × petri dishes (90 mm diameter). Based on the growth rate of the strain, the plates were monitored over the next 3 d – 7 d to obtain a saturated plate of L1s. These plates were then washed with M9 buffer into 15 ml centrifuge tubes and spun at 3000 RPM for 2 min. The supernatant was removed with a plastic pipette and discarded. This wash was repeated at least twice to remove the bacteria. After the final wash, the animals were resuspended in 3 ml of M9 buffer and 3 ml of freezing solution was added and mixed. This solution was then aliquoted into three cryotubes and placed into a styrofoam box and placed in a –80 °C freezer. After at least one week, one of cryotubes was transferred on ice to the lab bench for a test-thaw. A small amount of the frozen solution was transferred onto a 55 mm diameter petri dish using a sterile spatula and the cryotube was

transferred back to the -80°C freezer. The petri dish was monitored over the next couple of days for presence of live animals to make sure that the freezing process was successful. If the test-thaw was successful, two cryotubes were stored in liquid nitrogen and one cryotube was left at -80°C . If the test-thaw was unsuccessful, the freezing process was repeated from the beginning until the test-thaw was successful.

2.1.3 Transgenesis by microinjection

A 10 μl injection mix was prepared with the plasmids of interest at $5\text{ ng } \mu\text{l}^{-1} - 50\text{ ng } \mu\text{l}^{-1}$, a co-injection marker at $5\text{ ng } \mu\text{l}^{-1} - 20\text{ ng } \mu\text{l}^{-1}$ and a plasmid BJ36 to a final concentration of at least $100\text{ ng } \mu\text{l}^{-1}$. The injection mix was mixed by pipetting up and down and then centrifuged at maximum RPM for 5 min. 2 μl of the injection mix was loaded into the injection needle. Multiple young adults were immobilised by immersing them in halocarbon oil 700 (Sigma Aldrich, Inc.) on a 2% dried agarose pad for injection. Microinjection was performed according to the protocol outlined in Evans (2006). Briefly, for each animal, the injection needle was inserted into one of the two syncytial gonad arms and the injection mixture was released until the gonad swelled up. The needle was then gently retracted from the animal. The injected animals were rescued by adding M9 buffer on top of the halocarbon oil and transferring the worms carefully onto a new NGM plate. The animals were left to recover for a couple of hours and moved onto individual NGM plates. The plates were screened 2 d after the injection for the presence of F1 progeny containing the co-injection marker using a Zeiss dissecting microscope (AXIO Zoom.V16). These transgenic F1 progeny from a single injected P0 were transferred onto a new NGM plate to establish a transgenic line. Each P0 that produced transgenic F2s was considered a single line.

2.1.4 RNA interference (RNAi) by feeding

RNAi was performed by feeding bacteria expressing double-stranded RNA (dsRNA) corresponding to the targeted gene. RNAi bacterial clones for the genes of interest were streaked onto a Lysogeny broth (LB) agar plate with $50\text{ } \mu\text{g ml}^{-1}$ ampicillin and $12.5\text{ } \mu\text{g ml}^{-1}$ tetracycline mostly from the Ahringer library (Kamath and Julie Ahringer, 2003) and occasionally from Vidal library (Rual et al., 2004) or cloned into an empty vector (L4440) (see section

2.3.4). The clones were grown overnight in 9 ml liquid LB medium with $50 \mu\text{g ml}^{-1}$ ampicillin and $12.5 \mu\text{g ml}^{-1}$ tetracycline. 3 ml of the culture was used to extract plasmids using PureYield™ Plasmid Miniprep kit (Promega). The plasmids were sequenced using the standard vector primer M13 uni (-21) (listed in A.2) by Sanger sequencing (Eurofins Genomics. <https://www.eurofinsgenomics.eu>). The sequence was BLASTed against *C. elegans* genome PRJNA13758 on https://wormbase.org/tools/blast_blat to confirm that the RNAi clone targeted the gene of interest. The bacterial cultures were seeded onto RNAi plates and allowed to dry for 2 d – 4 d before use. RNAi plates were prepared by autoclaving NGM as in (Stiernagle, 2006) and adding $25 \mu\text{g ml}^{-1}$ ampicillin, $12.5 \mu\text{g ml}^{-1}$ tetracycline and 1 mM filter-sterilised Isopropyl β -D-1-thiogalactopyranoside (IPTG). RNAi plates were used within a month of preparation. To test the effect of knockdown of a gene of interest on the postembryonic development, 3 – 7 animals were bleached in a spot of bleach on the RNAi plate. F1 animals were phenotyped 2 d after the bleaching step. To test the effect of knockdown of gene of interest throughout development, $4 \times$ L4 animals were picked onto the RNAi plates and allowed to lay eggs. F1 animals and in case of egg-laying defective (egl) mutants F2 animals were phenotyped after 2 d or 5 d respectively.

2.1.5 Genetics

Males for crosses were generated by subjecting 5 – 10 L4 hermaphrodites to heat shock at 30°C for 3 h and allowing the hermaphrodites to develop and lay eggs at 20°C . The progeny was screened after 2 d to recover few males in the population. These males were used to set up additional maintenance crosses with 3 – 4 3 d adult hermaphrodites. Most genetic crosses were set up with a 1:4 ratio of hermaphrodites to males to ensure successful mating encounters.

2.1.6 Temperature treatment on seam cell development

Gravid hermaphrodites were spot bleached on plates and the embryos were allowed to develop postembryonically at 25°C . In the case of wild-isolate XZ1516 which is temperature sensitive and embryos did not develop upon bleaching, 4×30 synchronised healthy L4 animals were allowed to lay eggs for 8 h at 15°C and two plates were transferred to 20°C and 25°C .

2.1.7 Chemical mutagenesis, mutant screening and mapping

Synchronised L4 larvae of JR667 strain were incubated in 50 mM Ethyl methanesulfonate (EMS) in M9 buffer for 4 h (Brenner, 1974). EMS was performed in collaboration with Dr. Lamia Mestek Boukhibar, a postdoctoral scholar in the lab. 100 mutagenised P0s were singled out and they were allowed to recover for 2 d. The P0s were transferred every day for 2 d from the day they started laying eggs. The F1s were washed off the plates every day for 2 d from the day they started laying eggs. F2s were screened for aberrant seam cell number (scn <16 or >16) under a fluorescent dissecting microscope (AXIO Zoom.V16). Worms with aberrant scn were placed on individual NGM plates and the scn phenotype was validated in the population of F3s. Lines from singled F2s that showed two-sided errors were chosen for mapping the causative mutation(s). The validated mutant hermaphrodites in the Bristol (N2) genetic background were crossed to males of the polymorphic Hawaiian strain (CB4856). The F1 hermaphrodites from a successful cross were allowed to self. F2 animals were screened for aberrant seam cell phenotype and placed individually on NGM plates. F3 animals were phenotyped for scn and once they have consumed all the OP50, the worms are washed off in M9 buffer and stored as a pellet in -20°C freezer. The worm pellets from F3 animals that showed the mutant phenotype (i.e., <16 and >16) were pooled and their DNA was extracted according to the protocol in section 2.3.7 followed by whole genome sequencing (WGS) (see section 2.4) to find the causative mutation.

2.2 Microscopy and image analysis

2.2.1 Scanning electron microscopy (SEM) microscopy

Day-1 adult *eff-1* animals were washed twice with M9 buffer and fixed in 4% glutaraldehyde in M9 fixative solution for 3 h. Fixed animals were washed twice with M9 buffer. Fixed animals were dehydrated by incubating for 30 min in a series of solutions with increasing ethanol concentration (15% – 100%). Fixed samples were dried using a critical point dryer (K850, ProSciTech) and coated with gold/palladium for 2×90 s using the SC7620 Mini Sputter Coater (Quorum technologies). The samples were imaged on a JEOL JSM-6390 scanning electron microscope using 5 kV – 25 kV acceleration voltage. SEM images wild-type and *eff-1* animals

were acquired by my supervisor Dr. Michalis Barkoulas.

2.2.2 Counting seam cell number

At least 40 animals were anaesthetised using 100 μ M sodium azide on a fresh 2% agarose pad. All animals were scored for seam cell number (scn) at young adult stage or at late L4 stage if the animals were egl or sick. Number of seam cells were scored on the side closest to the objective. Number of seam cells in 30 animals or more were scored per strain on an inverted fluorescence microscope using a 40 \times objective on a Zeiss compound microscope (AxioScope A1).

2.2.3 Lineaging seam cells

Synchronised animals of wild-type and mutant at different developmental time points as described below were fixed in 4% paraformaldehyde in PBS after washing them twice with PBS and stored at 4°C.

- (i) 10 h for L1 asymmetric cell division
- (ii) 17 h for L2 symmetric cell divisions
- (iii) 27 h for L2 asymmetric cell divisions
- (iv) 31 h for L3 asymmetric cell divisions
- (v) 40 h for L4 asymmetric cell divisions

Wild-type and mutant animals at the same time point were mounted on a fresh 2% agarose pad and seam cell division patterns were observed for errors in the seam cell division pattern on a Nikon Ti Eclipse epifluorescence inverted microscope using a 40 \times objective .

2.2.4 Long-term time-lapse microscopy of *eff-1*

Animals were imaged according to Gritti et al. (2016). 11 *eff-1(icb4)* animals were imaged in microchambers by Dr. Michael Fasseas, a postdoctoral scholar in the lab in collaboration with the van Zon lab at the AMOLF in Amsterdam, Netherlands. Lineaging of seam cells in animals was done by looking at frame where divisions took place and confirming with additional frames when necessary by a research assistant Ritobrata Ghose in the lab using custom Python pipeline developed by Gritti et al. (2016) (https://github.com/jvzonlab/timelapse-natcomm-2016/tree/master/Seam_cell_lineage_analysis). I confirmed the errors by going through the

specific lineages through as many frames as needed. 11 animals were imaged through development. 2 animals were excluded from lineaging because of bad imaging quality and one of the animals did not develop properly due to microchambers. 18 lineages were produced by lineaging both the lateral sides of remaining 9 animals. Imaging of one animal (C013) was only complete till L3 due to imaging failure. The frequencies of errors in L4 were adjusted according to this.

2.2.5 Single molecule mRNA fluorescent in situ hybridisation (sm-FISH)

For synchronisation of larvae, gravid hermaphrodites were bleached and eggs were allowed to hatch and develop at 20 °C for 27 h before fixation. smFISH was performed using custom-made probes labelled with Cy5 for *eff-1*, *nhr-73* and *elt-6* (listed in Appendix A.3). Z-stacks with 17 slices, each of 0.7 μm , were acquired for each larvae with a 100 \times oil immersion objective using an Andor iKon M 934, 1024 \times 1024 CCD camera system on a Nikon Ti Eclipse epifluorescence inverted microscope. Each slice in the Z-stack was acquired by exposing to three different channels: DAPI, Far red and GFP for 200 ms, 3 s and 300 ms respectively using the Nikon image acquisition software. The DAPI and GFP images helped identify the seam nuclei for quantification of mRNA spots in Far red image. First, a region of interest (ROI) was drawn around a seam cell of interest using a custom script in MATLAB[®]. Second, mRNA spots were identified manually and quantified by another custom script in MATLAB[®]. One-way or two-way ANOVA as appropriate for the experiment was performed on mRNA molecules as the response variable in R 3.2.0.

2.2.6 Seam-cell shape analysis

Images were acquired using a oil immersion 40 \times objective with a CoolSNAP HQ Monochrome camera (Photometrics,USA) mounted on an inverted Leica DM-IRBE microscope (Leica Microsystems, Germany). The animals were bleached and imaged 17 h after hatching. Images were acquired with exposure times of 2000 ms for the GFP channel and 4 ms for the brightfield channel with no pixel binning with MetaMorph imaging software. Animals were straightened and segmented using a custom pipeline put together in Fiji by an undergraduate student Fu

Xiang Quah (Schindelin et al., 2012). Cell shape parameters in Fiji such as major, minor, perimeter and area were used to perform principal component analysis on individual cells at the end of L1 asymmetric division.

2.2.7 Measurement of angles between the seam cells during divisions

The angle between cells after cell division was calculated as shown in Fig. 3.6. The posterior cell was used as the reference point. A right-angled triangle could be imagined whose hypotenuse joins the centroids of the two cells. The angle was converted to degrees. Angles between cells after L2 symmetric and asymmetric cell division were quantified as shown in Fig. 3.6C and Fig. 3.6G. Angles for anterior-posterior pair in L2 asymmetric cell division were measured for H1, H2, V1 – V4, V6 and T seam cells. I performed one-way analysis of variance (ANOVA) on magnitude of angles between the strains. The genotype was considered the explanatory variable and the angle as the dependent variable.

2.2.8 Confocal microscopy

Animals of the desired stage were anaesthetised using 100 μ M sodium azide on a fresh 2% agarose pad and imaged using a Leica SP5 inverted confocal microscope with the appropriate laser configuration. A constant number of slices were imaged for MBA804 (N2) and MBA846 (CB4856) carrying the *POPHHOP* transgene *huIs154 [Pes10::TCFenh::NLS-GFP; dpy-20(+)]*.

2.2.9 *POPHHOP* marker intensity analysis

The confocal images were extracted with Fiji and the sum total fluorescence of all the slices for a given animal was extracted. An region of interest (roi) was drawn from the vulva to the tail tip containing intestinal, vulval and seam cells expressing *POPHHOP* marker using the polygon tool in Fiji on the bright field image. The roi was copied and pasted on the background area without the worm. The following parameters were selected in the set measurements section of analyse menu in Fiji: area, integrated density and mean grey value. The parameters were extracted from the fluorescent image by selecting measure from the analyse menu. The corrected total cell fluorescence (CTCF) was calculated as follows

$$\text{CTCF} = \text{Integrated Density} - (\text{Area of roi} \times \text{Mean fluorescence of background readings}) \quad (2.1)$$

2.3 Molecular methods

2.3.1 Genomic DNA extraction from a single worm and polymerase chain reaction (PCR)

For DNA extraction from a single worm, a single worm was placed in 5 μl worm lysis (SWL) buffer containing 0.2 $\mu\text{g } \mu\text{l}^{-1}$ of Proteinase K in a 0.2 ml PCR tube with attached lids. The worm lysis buffer was then incubated at 65 $^{\circ}\text{C}$ for 1 h to digest proteins by proteinase K enzyme followed by another incubation step at 95 $^{\circ}\text{C}$ for 15 min to inactivate Proteinase K. 1 μl of this solution was used as template in a 50 μl PCR reaction. For pooled worm lysis, 2 – 50 animals were placed in 5 μl – 10 μl worm lysis buffer and incubated as above. 1 μl and 0.5 μl of this solution respectively was used as template in a 50 μl PCR reaction. Gotaq[®] DNA Polymerase from Promega was used for diagnostic or troubleshooting PCRs where the amplicon size was lesser than 3 kb with extension time of 1 kb min^{-1} . Most primers had an annealing temperature 56 $^{\circ}\text{C}$ unless mentioned otherwise. PCRs where the amplicon size was greater than 3 kb or when the amplicon was used for downstream cloning application, Phusion[®] High-Fidelity DNA Polymerase from New England Biolabs was used. The extension time of 1 kb min^{-1} was used according to the size of the amplicon.

2.3.2 Sanger sequencing of targeted DNA sequences

The DNA sequence that needed to be sequenced was amplified by PCR using appropriate primers (listed in Appendix A.2). The PCR product was purified using Wizard[®] SV Gel and PCR Clean-UP System according to the protocol from Promega. Plasmids or PCR products that needed to be sequenced were sent for Sanger sequencing to Eurofins Genomics (<https://www.eurofinsgenomics.eu>) with the appropriate primers. The DNA sequence was downloaded and aligned on benchling or on ape.

2.3.3 Cloning *egl-18*

egl-18 was cloned into a vector with seam cell specific promoter and *unc-54* 3' UTR, pIR5 (Katsanos et al., 2017) to produce *pseam::egl-18::unc-54*. Primers were designed to amplify the coding sequence of *egl-18* from CHROMSOME_IV: 1913472..1917360 (listed in Appendix A.2). The following sequences were added to the forward and reverse primers for cloning *egl-18* into pIR5.

(i) Forward primer: 5'- TTGCTTGGAGGGTACCGAGTTTAAACATTT.... -3'

(ii) Reverse primer: 5'- GTAATTGGACTTAGAAGTCAGAGGCAATTT.... -3'

The primers were used to amplify the genomic fragment from *egl-18* using worm lysis from N2 as a template. The PCR product was purified using Wizard[®] SV Gel and PCR Clean-UP System according to the protocol from Promega. 1 µg of empty vector pIR5 was digested with FastDigest enzyme SmaI and the digest was run on a 1% agarose gel with SYBR[™] Safe DNA Gel Stain (ThermoFisher Scientific). The linearised vector fragment of size 4027 bp was extracted from the gel using Wizard[®] SV Gel and PCR Clean-UP System according to the protocol from Promega. 100 ng of the linearised vector and equimolar amount of the PCR product were added to 10 µl of Gibson master mix and the volume was made up to 20 µl with H₂O. The reaction mix was incubated at 50 °C for 60 min and then placed on ice. 10 µl of the reaction mix was transformed using DH5α bacteria. Two bacterial clones were verified using Sanger sequencing (Eurofins Genomics. <https://www.eurofinsgenomics.eu>). The resulting plasmid *pSK5[pseam::egl-18::unc-54 3' UTR]* was injected at 20 ng µl⁻¹ with *myo-2::dsRed* at 5 ng µl⁻¹ as co-injection marker.

egl-18 was cloned into a vector with hypodermis promoter and *unc-54* 3' UTR, pIR6 (size = 3640 bp) to produce *dpy-7p::egl-18::unc-54* according to the same cloning protocol as described above with the difference of the following sequences being added to the forward and reverse primers for cloning *egl-18* into pIR6.

(i) Forward primer: 5'- ACATTTTGTTCAGATAAGTTTAAACATTT.... -3'

(ii) Reverse primer: 5'- GTAATTGGACTTAGAAGTCAGAGGCAATTT -3'

The resulting plasmid *pSK6[dpy-7p::egl-18::unc-54 3' UTR]* was injected in 20 ng µl⁻¹ with

myo-2::dsRed at $5 \text{ ng } \mu\text{l}^{-1}$ as co-injection marker.

2.3.4 Cloning RNAi constructs

Primers were designed to amplify a fragment spanning at least a couple of exons with the introns in between them for the gene of interest (listed in Appendix A.2) . The following sequences were added to the forward and reverse primers for cloning.

- (i) Forward primer: 5'- AGACCGGCAGATCTGATATCATCGATG.... - 3'
- (ii) Reverse primer: 5'- TCGACGGTATCGATAAGCTTGATATCG.... - 3'

The primers were used to amplify the genomic fragment from the gene of interest using worm lysis of N2 as a template. The PCR product was purified using Wizard[®] SV Gel and PCR Clean-UP System according to the protocol from Promega. $1 \mu\text{g}$ of the RNAi empty vector L4440 was digested with FastDigest enzyme EcoRI and the digest was run on a 1% agarose gel with SYBR[™] Safe DNA Gel Stain (ThermoFisher Scientific). The linearised vector fragment of size 2790 bp was extracted from the gel using Wizard[®] SV Gel and PCR Clean-UP System according to the protocol from Promega. 50 ng of the linearised vector and equimolar amount of the PCR product were added to 5 μl of Gibson master mix and the volume was made up to 10 μl with H₂O. The reaction mix was incubated at 50 °C for 60 min and then placed on ice. 5 μl of the reaction mix was transformed using HT115 bacteria.

2.3.5 Design and cloning of single-guide RNA for Clustered Regularly Interspaced short Palindromic Repeats (CRISPR)-Cas9 mediated genome editing

The co-CRISPR strategy from Arribere et al. (2014) was used to edit the following genes: *bro-1*, *nhr-25*, *rnt-1*, *egl-18*, *lin-22* and *eff-1*. An sgRNA targeting the following sequences in the first, second or third exon of the gene of interest was cloned into *pU6::unc-119* sgRNA vector by replacing the *unc-119* sgRNA as previously described (Friedland et al., 2013). sgRNA constructs targeting *bro-1*, *rnt-1*, *nhr-25* and *egl-18* were produced with the help of students Julia Spindel and Judy Ghalayini.

- (i) *rnt-1* sgRNA (third exon): 5'- AGCAAAAGTGCATCGACAAG - 3'
- (ii) *nhr-25* sgRNA (first exon): 5'- GTTTGTGGTGATCGAGTCTC - 3'

- (iii) *egl-18* sgRNA (second exon): 5'- AATGATGCAATTATTATCAA - 3'
- (iv) *egl-18* sgRNA (second exon): 5'- GGAGCGATCCGATATCCCGA - 3'
- (v) *bro-1* sgRNA (second exon): 5'- AATCAATATACCTGTCAAGT - 3'
- (vi) *lin-22* sgRNA (first exon): 5'- ACTGAAATTGAATCCGATGG - 3'
- (vii) *eff-1* sgRNA (first exon): 5'- GGTGTCTTGGAACAGTGTGG - 3'

icb44 is a deletion of 8 bp (GGAATCAATATACCT- - - - -GGAATGGT), *icb45* is a in-frame deletion of 9 bp (GGAATCAATATA- - - - -TTGGAATGGT), *icb46* is a in-frame 9 bp indel (GGAATCAATATACCTGTCA- - - - - GTTGGAATGGT), *icb49* is an in-frame deletion of 9 bp (GGAATCAATATACCTG- - - - -AATGGT) in the second exon of *bro-1*. *icb50* is a deletions of a single base pair (ATTGAATCCG-TGGTGGAAATCTC), *icb49* is a 5 bp deletion (ATTGAAT- - - -GTTGGAATCTC) and *icb52* is 9 bp in-frame deletion (CTGAAATTG- - - - -GTGGAATCTC) in the first exon of *lin-22*.

2.3.6 co-CRISPR and mutant screening

The injection mix contained *peft3::cas9* at 50 ng μl^{-1} , *pU6::dpy-10 sgRNA* at 25 ng μl^{-1} , *pU6::gene of interest sgRNA* at 25 ng μl^{-1} , repair oligo template for *dpy-10* at 10 pmol μl^{-1} , repair oligo template for *gene of interest* at 10 pmol μl^{-1} and *myo2::dsRed* at 5 ng μl^{-1} . F1 animals showing morphological phenotypes indicative of modifications at the *dpy-10* locus were examined for the presence of multiple PDE neurons. PCR was performed on F2 animals by using primers for the gene of interest (listed in Appendix A.2), and the amplified fragment was sequenced to find the nature of the de novo mutation. For injections targeting *lin-22*, wild-isolates carrying *vtIs1 [dat-1p::GFP + rol-6]* and *wIs51 [SCMp::GFP]* were used. I did not use a repair template for injections targeting *eff-1* and *lin-22*.

2.3.7 Genomic DNA extraction from large numbers of worms

Worms whose DNA was to be extracted were propagated in large numbers on a 55 mm diameter petri dish. Once they had consumed all the OP50 bacteria on the petri dish, they were washed off the plate with M9 buffer into a 1.5 ml centrifuge tube. The worms in M9 buffer were kept on ice for 1 h – 3 h to pellet the worms and remove bacteria. The worms were washed at least twice with M9 to remove bacteria. At the end of the final wash, M9 was removed as much as possible

without disturbing the worm pellet. The worm pellet could be frozen at -20°C until DNA extraction. DNA was extracted using Genra Puregene Kit (Qiagen[®]). DNA concentration and quality was quantified using a NanoDrop (Spectrophotometer ND-1000). 1 μg of DNA was also run on a 1% agarose gel to make sure that the DNA was not sheared and intact.

2.4 Whole genome sequencing

Good quality DNA of at least 1 μg was sent to GATC Biotech (Eurofins Genomics <https://www.eurofinsgenomics.eu> for sequencing on an Illumina[®] HiSeq platform (read length = 2×125), to obtain at least 10 million read pairs ($25 \times$ coverage). The whole genome sequencing data in fastq format was run through the CloudMap Hawaiian variant mapping pipeline on a local galaxy server maintained in the lab (Minevich et al., 2012) with the help of a postdoctoral scholar Dr. Michael Fasseas in the lab. The regions of the chromosome with linkage to the causal mutation will not have polymorphisms associated with CB4856 but will resemble the N2 strain. Thus the causative mutation could be found in the preserved N2 genomic regions by looking at mutations in annotated gene features.

2.5 Quantitative genetics methods

2.5.1 Introgression of the seam cell marker *SCMp::GFP* into wild-isolates

Seam cell marker *wIs51 [SCMp::GFP]* along with *vtIs1[dat-1p::GFP + rol-6]* located on chromosome V were introgressed into wild-type isolates (JU775, JU2519, CB4856) in a two-step cross repeated five times to produce $10 \times$ backcrossed strains. In the first step, hermaphrodites of the strain MBA95 carrying *lin-22(ot267)*, *wIs51* and *vtIs1* were mated with males from the wild-type isolate. In the second step, F1 males from the previous cross were crossed to wild-type hermaphrodites. The F1 hermaphrodites carrying both *wIs51* and *vtIs1* from the previous cross were allowed to self and homozygous progeny for the marker were considered backcrossed twice. The number of *dat-1p::GFP* neurons were observed in the $10 \times$ backcrossed wild-type animals to make sure that *lin-22(ot269)* was not segregating in the population. *lin-22(ot269)* increases the number of *dat-1p::GFP* neurons from one in wild-type to an average of three in

the population (Katsanos et al., 2017; Doitsidou and Hobert, 2019). Based on previous observations in the lab, the two transgenes (*wIs51* and *vtIs1*) were far enough on chromosome V to get recombinants carrying only one of them when crossed. A final cross was set up with males from wild-isolates and hermaphrodites from $10 \times$ backcrossed wild-isolates carrying both the transgenes *wIs51* and *vtIs1*. The F1 hermaphrodites from a successful cross (presence of approx. 50% males) were allowed to self and recombinant animals which only carried *wIs51* were picked and propagated. Michalis Barkoulas introgressed *wIs51* transgene into two wild-isolates (JU2007 and XZ1516) by crossing wild-type males to JR667 hermaphrodites. F1 males carrying the transgene were crossed to wild-type hermaphrodites. This process was repeated ten times to produce $10 \times$ backcrossed wild-type strains (listed in Appendix A.1).

2.5.2 Introgression of mutations in a GATA transcription factor (*egl-18*), and a fusogen (*eff-1*) into wild-isolates

Mutations in *egl-18* located on chromosome IV were introgressed into wild-type strains MBA256 and MBA19 (CB4856 and JU2007 respectively carrying *wIs51* transgene) in a two-step cross. In the first step, MBA256 and MBA19 males were crossed to *egl-18(ls)* hermaphrodites carrying the *wIs51* transgene. In the second step, F1 males from the previous cross were crossed to wild-type hermaphrodites. F1 hermaphrodites from a successful cross (presence of approx. 50 percent males) were allowed to self. Animals that were egg-laying defective were picked up and the two-step cross was repeated five times to produce $10 \times$ backcrossed strains. The process was same for introgression of *eff-1(icb4)* into MBA19 background (listed in Appendix A.1).

2.5.3 Generation of recombinant inbred lines (RILs) containing *egl-18(ga97)* mutation

Recombinant inbred lines (RILs) were created by crossing hermaphrodites from strain MBA256 (*icbIR2(V,N2>CB4856)*, *wIs51 [SCMp::GFP] V*) and males from strain MBA231 (*icbIR2(V,N2>CB4856)*, *vtIs1[dat-1::gfp] V*, *wIs51 [SCMp::GFP] V*). F1 males from the cross were crossed to hermaphrodites from strain MBA290 (*egl-18(ga97) IV*; *wIs51 [SCMp::GFP] V*). Multiple F1 hermaphrodites which were egg-laying defective and carry the transgene *vtIs1 [dat-1::gfp] V*, *wIs51 [SCMp::GFP]*

V were picked as they were the cross progeny and allowed to self. These egg-laying defective gravid F1s were bleached to release the eggs. 117 F2s which carried only *wIs51/SCMp::GFP/V* transgene were picked onto single NGM plates and allowed to self. One hermaphrodite was transferred per generation for 10 – 14 generations to establish 116 RILs (Fig. 5.1). RIL-17 did not propagate during the selfing process. All the RILs were maintained on the same batch of NGM plates at 20°C to avoid changes in allele segregation frequencies caused by selection pressure due to environmental conditions.

2.5.4 Phenotyping the RILs for seam cell number (scn) and pooling strategy for bulk segregant analysis

RILs were phenotyped over a week as described in section 2.2.2. RIL populations grew at different rates as they were egg-laying defective as they carried a mutation *ga97* in the *egl-18* gene. 116 RILs were scored for scn at least twice and 38 RILs were scored thrice to be confident about their scn phenotype. We pooled the RILs into two extreme groups (22 RILs in the low-scn bulk and 24 RILs in the high-scn bulk) for QTL mapping. To understand the best pooling strategy for genotyping, we separated each of the extreme groups into a stringent group (10 RILs with the lowest or the highest scn in each bulk) and less stringent group containing all the RILs in that bulk. The 116 RILs including those that were selected for QTL mapping and their scn values are listed in the appendix (Appendix B.3). A few lines were phenotyped by Dimitrios Katsanos, a Phd student in the lab and by Sophie Gilbert, a postdoctoral scholar in the lab.

2.5.5 Quantitative trait locus QTL mapping

We used a bulk segregant analysis approach to discover quantitative trait loci (QTLs) associated with the scn phenotype (Michelmore, Paran, and Kesseli, 1991; Frézal et al., 2018). DNA was extracted from the two low and two high bulks according to protocol details in section 2.3.7 and sent for whole genome sequencing (see section 2.4).

2.5.6 Bioinformatic analysis of whole genome sequencing data

We used the CloudMap Hawaiian variant mapping pipeline used for mutant discovery (Minevich et al., 2012) for discovering genomic regions associated with the scn phenotype (see section 2.4). Using VCF tools on the locally installed galaxy server, we derived the genotype at known SNP positions for low-scn and high-scn bulks by comparing to the reference genome ce10. The genotype for low-scn and high-scn was extracted into a single vcf file from the whole genome data using vcf combine on galaxy. The combined vcf was converted to tabular format using GATK tools on a bash terminal on a mac. Frequency of CB4856-snp at each SNP position was calculated as the ratio of read counts with CB4856 SNP divided by the total number of reads at this SNP and the same method was used to get N2 SNP frequency.

2.5.7 Bulk segregant analysis

Single nucleotide polymorphisms (SNPs) between the parental strains N2 and CB4856 throughout the genome can be used as markers. For the low-scn and high-scn bulk, the SNP positions where the genome quality (as determined by GQ scores <40) for the sequencing was low were discarded (Wall et al., 2014; Song, L. Li, and G. Zhang, 2016) using the QTLseqr package in R. SNPs where the total read depth was lower than 22 in the low-scn bulk and lower than 24 in the high-scn bulk were discarded following the protocol in Frézal et al. (2018). Log-odds ratios were not calculated for the stringent condition as the high-bulk sample had low read depth.

Under the null hypothesis that there are no QTLs segregating anywhere in the genome, genomic regions do not affect the scn and the SNP frequencies of N2 and CB4856 are expected to be 50% in both low-scn and high-scn bulks. In contrast, under the alternative hypothesis that there are QTLs that affect scn, genomic regions responsible for high scn are likely to be enriched for the CB4856 SNPs in the high-scn bulk and regions of the genome responsible for low scn will be enriched for the N2 SNPs in the low-scn bulk. Low-scn bulk (l) and high-scn bulk (h) should therefore have diverging frequency of SNPs for the genomic regions affecting scn. We used Log-odds ratio, which is the logarithm (base 10) likelihood ratio as a test statistic to evaluate if the deviations observed in SNP frequencies between the low-scn and high-scn bulk were statistically significant than expected from a null distribution.

We calculated observed log-odds ratio using the following formula for previously defined genomic intervals:

$$\text{log-odds ratio} = \log_{10} \left(\frac{\frac{l}{n_1 - l}}{\frac{h}{n_h - h}} \right) \quad (2.2)$$

where n_l and n_h are number of RIL lines pooled for low-scN and high-scN bulks, respectively. The log-odds ratios under the null hypothesis were calculated where SNP frequencies for both the bulks were generated by 1 million simulated Bernoulli trials with $p=q=0.5$. From these log-odds ratios, we found the two-tailed genome-wide threshold at a significance level ($\alpha = 0.05$). The observed log-odds ratio was plotted against the genome location, and was compared to the genome-wide thresholds. The presence of a QTL was inferred when the log-odds ratio exceeded the threshold.

2.6 Molecular genetics

2.6.1 Designing genetic markers based on single nucleotide polymorphisms (SNPs) in the CB4856 genome

Single nucleotide polymorphisms (SNPs) between the strains N2 and CB4856 throughout the genome are known and can be used as markers to distinguish between the strains. There are a subset of snps which modify restriction enzyme recognition sites called snip-SNPs and are detected as restriction fragment length polymorphisms (RFLPs) (Wicks et al., 2001). Primers (listed in Appendix A.4 and Appendix A.5) were designed on benchling to amplify a region of the genome carrying the SNP or snip-SNP. The PCR product was purified using Wizard[®] SV Gel and PCR Clean-UP System according to the protocol from Promega. In the case of regular SNPs, the purified PCR product was sequenced using the forward or reverse primer by Sanger sequencing (Eurofins Genomics. <https://www.eurofinsgenomics.eu>). However, for snip-SNPs, 1 μ g of the purified PCR product was digested with the appropriate FastDigest restriction enzyme for up to 4 h and run on a 1.5% agarose gel.

2.6.2 Designing genetic markers based on deletions/insertions in the CB4856 genome

In addition to SNPs, there are many insertions and deletions in the CB4856 genome when compared to the reference N2 genome. These deletions can be used as markers to distinguish between the strains. Most deletions of 50 kb – 200 kb in genomic region of interest were primarily visualised using the variant browser on the *C. elegans* Natural Diversity Resource (CeNDR) platform (Cook et al. (2017), <http://www.elegansvariation.org/>) by aligning the BAM files of N2 and CB4856. Additional deletions/insertions in the genomic region of interest were found in the supplementary table (O. A. Thompson et al. (2015), File_S3_Celegans_N2_CB4856_INDELS_features.txt). The genome releases PRJNA13758 and PRJNA275000 of N2 and CB4856, respectively were downloaded from ftp server on wormbase (<ftp.wormbase.org>). The sequence flanking the deletion was retrieved using samtools from both N2 and CB4856 and annotated (H. Li et al., 2009). Primers (listed in Appendix A.6) were designed on benchling to produce a <1.5 kb amplicon. The primers were BLASTed against *C. elegans* N2 genome PRJNA13758 and CB4856 genome PRJNA275000 on https://wormbase.org/tools/blast_blat to make sure that the primers did not have polymorphisms. In addition, in silico PCR was performed with the designed primers on the e-PCR Search on wormbase (Schuler (1997), <https://wormbase.org/tools/epcr>). The primers were validated by a PCR with Gotaq[®] DNA Polymerase from Promega and the size differences between the amplicons were visualised on a 2% agarose gel with SYBR[™] Safe DNA Gel Stain (ThermoFisher Scientific). MassRuler Express Forward DNA Ladder Mix (ThermoFisher Scientific) was used as reference on DNA gel electrophoresis.

2.6.3 Generation of near isogenic lines (NILs) containing quantitative trait loci

Using the markers designed in section 2.6.2, three QTLs on chromosomes II, III and X from RIL-28 were introgressed into N2 background. First, wild-type strains carrying a single QTL were derived. Males from JR667 were crossed to egg-laying defective hermaphrodites from RIL-28, the F1 males were crossed to JR667 hermaphrodites. F1 hermaphrodites from a successful cross

were singled out and allowed to self. 8 F2 hermaphrodites were singled out. DNA was extracted from 8 single F2 hermaphrodites according to protocol in section 2.3.1 after 2 d – 3 d. First, one PCR per chromosome was performed using markers (listed in Appendix A.6) and run on a 2 % agarose gel with SYBR™ Safe DNA Gel Stain (ThermoFisher Scientific). Additional PCRs for markers for the QTLs were performed on the F2 hermaphrodites that were homozygous for CB4856 fragment to ensure that the full QTL was present. A recombinant F2 carrying a smaller fragment of QTL on CHROMOSOME_II was discovered and the progeny were homozygosed and backcrossed to N2. F2 hermaphrodites carrying a single QTL was backcrossed four times to get 10 × backcrossed NILs using the same two-step cross scheme outlined in section 2.5.1.

To get NILs carrying the *egl-18(ga97)* mutation, males from the NILs carrying individual QTLs were crossed to *egl-18(ga97)* hermaphrodites. Multiple wild-type F1 hermaphrodites were picked into a plate and allowed to self. Eight egg-laying defective F2 hermaphrodites were singled out and PCRs were performed on markers to get homozygous QTL. To get *egl-18(ga97)* strains carrying two QTLs, males from the previous cross (chr. II) were crossed to wild-type hermaphrodites carrying QTL on another chromosome (e.g. chr. III). Presence of two QTL on chromosomes II and III in singled out F1s was detected by PCR using markers. These QTLs were homozygosed in F2 hermaphrodites.

To get *egl-18(ga97)* strains carrying three QTLs, males from JR667 were crossed to an *egl-18(ga98)* hermaphrodites carrying two QTLs (on CHROMOSOME_II and CHROMOSOME_III). F1 males were crossed to *egl-18(ga98)* strain carrying a third QTL (on CHROMOSOME_X). Singled wild-type F1 hermaphrodites were screened for animals that carried all three QTLs and homozygosed. The F2 hermaphrodite which was homozygous for the three QTLs was found to be carrying a smaller fragment of QTL on CHROMOSOME_II. A wild-type strain carrying all the three QTLs was made from the *egl-18* strain carrying all three QTLs by crossing hermaphrodites to JR667 males and screening F1s for the presence of QTLs and homozygosing QTLs. After this cross, it was discovered with new deletion markers that the wild-type strain carrying all three QTLs had a narrowed QTL on CHROMOSOME_III which was due to the lack of resolution at the time of the cross.

Recombinants with smaller QTL intervals recovered during the crosses were homozygosed

and kept as they were of interest to find the causative genetic variation for the differential scn phenotype. All the strains (listed in Appendix A.1) were frozen and maintained for phenotyping scn.

2.6.4 Generating smaller QTLs using genetic crosses and screening

Most of the smaller interval QTLs were generated by saving recombinants produced by chance in previous crosses. However, to break a QTL (*icbIR21*) on CHROMOSOME _II into a smaller fragment, we set up a cross with hermaphrodites carrying *icbIR21* and JR667 males. We picked 196 F1 hermaphrodites and let them self. DNA was extracted from 196 single F2 hermaphrodites according to protocol in section 2.3.1 after 3 d. Two PCRs for two deletion markers were performed to find recombinants. We allowed multiple F1 recombinants to self and we confirmed their QTL intervals with additional markers.

2.7 Statistical analysis

All statistical analysis was done in R 3.2.0 using a sublime text editor. One-way or Two-way ANOVA was conducted depending on the experiments to test for differences in the mean seam cell number between strains and different temperature conditions. In ANOVA, the seam cell number is considered the response variable, strain and temperature are considered fixed explanatory variables. In a one-way ANOVA, when there was a significant effect of strain/RNAi on seam cell number, post hoc Tukey HSD tests were conducted when all pair-wise means were to be compared or post hoc Dunnett's test was conducted if all samples were to be compared against a control. Levene's median test was used to test for differences in variance in seam cell number between strains.

Chapter 3

Identification of a novel role for fusogen
eff-1 in the robustness of seam cell
number

3.1 Introduction

Development of multicellular organisms from single-cell embryos requires synchronisation of gene expression across spatial and temporal scales to produce functional organisms in a changing environment (Maduro, 2015). However, phenotypic variation is low in development. Therefore, development is robust to internal perturbations such as noise in gene expression, mutations and standing genetic variation in the population and external perturbations such as nutrition and temperature (Félix and Wagner, 2008; Masel and Siegal, 2009). An unresolved question in developmental biology is what are the mechanisms that underlie developmental robustness. For example, are robustness genes specific to a trait and type of variation? Are they network hub genes like cell-cycle regulators and chromatin remodellers?

Systematic screens to discover genes conferring robustness have only been done in unicellular organisms like *S. cerevisiae* (Levy and Siegal, 2008) but not in multicellular organisms. *C. elegans* development is an ideal model to study developmental robustness because its entire embryonic cell lineage and postembryonic lineage is known (Sulston, Schierenberg, et al., 1983; Sulston and Horvitz, 1977). Also, it can be maintained isogenically, which allows us to disentangle the effects of genetics and environment on development. Seam cells are epidermal cells that undergo stem cell-like division patterns during postembryonic development to produce epidermal and neuronal cells. Postembryonic seam cell patterning is mostly invariant between animals (Sulston and Horvitz, 1977; Podbilewicz and J. G. White, 1994). The *C. elegans* L1 larvae is born with 10 seam cells per lateral side. These cells undergo stereotyped symmetric and asymmetric cell divisions during larval development to produce 16 seam cells in the adult.

Seam cells in an L1 larva go through first asymmetric division to produce a hypodermal cells that fuse to *hyp7* and seam cells that continue to divide in the following larval stages. In the early L2 stage, H1, V1 – V4 and V6 seam cells undergo symmetrical proliferative divisions to produce 16 seam cells per lateral side. During the following larval stages in L2, L3 and L4, these 16 seams cells undergo asymmetric divisions. The anterior seam cell daughter differentiates into an epidermal, or neuronal cell, and the posterior daughter maintains the stem cell fate. There are 16 seam cells per lateral side in wild-type animals at the end of development in L4, and this number is robust to stochastic variation (Mestek Boukhibar and Barkoulas, 2016).

eff-1 encodes a well-characterised nematode-specific fusogen that is required for most of the cell fusion events in *C. elegans* (Mohler et al., 2002). In wild-type *C. elegans*, one-third of all somatic cells born fuse throughout development (Podbilewicz and J. G. White, 1994). The largest organ, the epidermis is made of 8 syncytia containing 186 nuclei, the largest syncytium being hyp7 (Hedgecock and J. G. White, 1985). Anterior seam cell daughters produced by asymmetric division of lateral seam cells fuse to hyp7 syncytium during larval development while posterior seam daughters maintain seam cell fate. These cells lose expression of *SCMp::GFP* and start expressing *eff-1* after division, which induces the fusion pores in the cell membrane (Podbilewicz, Leikina, et al., 2006). Upon completion of fusion of anterior seam cell daughters to hyp7, the nucleus moves out from the seam tissue. Following which the posterior seam cell daughters reconnect with neighbouring seam cells (Podbilewicz and J. G. White, 1994; Austin and C. Kenyon, 1994).

In this chapter, I aim to discover the genes that buffer seam cell number (SCN) through an unbiased forward genetic screen. Many mutagenesis screens have identified mutations in genes that affect seam cell patterning (Mohler et al., 2002), seam cell number (Huang et al., 2009) or both (Wildwater et al., 2011). However, these screens have focused on isolating mutations affecting the average seam cell number. While these types of screens may identify genes that are involved in the core developmental pathway, they may not find genes that may be involved in reducing phenotypic variation in development. Recently, we performed a genetic screen based on increase in SCN variability, and discovered a new role for *Hes*-related bHLH transcription factor *lin-22* in buffering SCN variability (Katsanos et al., 2017). I used this strategy to recover a mutation *icb4* in the fusogen *eff-1* that affects SCN variability. I found that other alleles of *eff-1* also show variable SCN and it is not specific to the *icb4* allele we recovered.

We discovered that *eff-1* affects SCN variability in a completely unbiased way in the genetic screen for variable seam cell number (VSC) mutants. Consistently, previous studies found *eff-1* mutants to have wild-type SCN on average (Huang et al., 2009; Katsanos et al., 2017). However, *eff-1* has not been studied in the context of seam cell patterning but as a gene that facilitates differentiation of the anterior seam cell daughters, which express *EFF-1*. We show through long-term time-lapse lineaging that *eff-1* mutants are not wild-type in seam patterning even though

they may have a wild-type SCN on average. We characterise the seam cells in great detail in the *eff-1* mutant to investigate the mechanism by which *eff-1* affects seam cell patterning. I present evidence in this chapter for a new role for *eff-1* in reducing SCN variability.

3.2 Results

3.2.1 Forward genetic screen for mutants with variable seam cell number identifies MBA21

To discover genes that buffer SCN, we recovered mutants with increased variance of SCN compared to wild-type following the strategy outlined in Katsanos et al. (2017). Briefly, we isolated F2 animals with deviation from wild-type SCN ($SCN > 16$ or < 16) from mutagenised *C. elegans* strain JR667 carrying the seam cell marker (*wIs51 [SCMp::GFP]*). These F2 mutants were allowed to self and SCN was scored in a population of F3 animals. Mutants that showed two-sided errors of SCN, i.e., animals in the population have higher and lower SCN compared to wild-type, termed variable seam cell number (VSC) mutants were kept for further analysis. VSC mutants display an increase in variance of SCN (measured by SD) with minimal change in average SCN (Fig. 3.1A). One of the five mutants satisfying this criterion was MBA21 (Fig. 3.1B). MBA21 showed a significant increase in SCN variance compared to wild-type but not in the average SCN ($WT = 16 \pm 0.26$ SD versus $MBA21 = 16.47 \pm 1.22$ SD, $p_{variance} = 0.01$, $p_{average} = 0.75$). MBA21 was picked because it had a dumpy (Dpy) phenotype in addition to VSC phenotype. We found that Dpy and VSC phenotype co-segregated (Fig. 3.1C). So, we argued that this would be easier to map. To map the causative mutation for VSC phenotype in MBA21, we crossed MBA21 to a polymorphic strain CB4856 (Hawaiian strain) and picked F2 recombinants based on their Dpy phenotype following the mapping strategy used in Doitsidou, Poole, et al. (2010). We pooled progeny of these F2 recombinants and used whole genome sequencing to identify the mutation in MBA21.

3.2.2 Mutation in MBA21 strain maps to the fusogen *eff-1*

In order to identify the causative mutation from the whole genome sequencing data, we used the CloudMap pipeline (Minevich et al., 2012). Briefly, the sequencing reads were aligned to

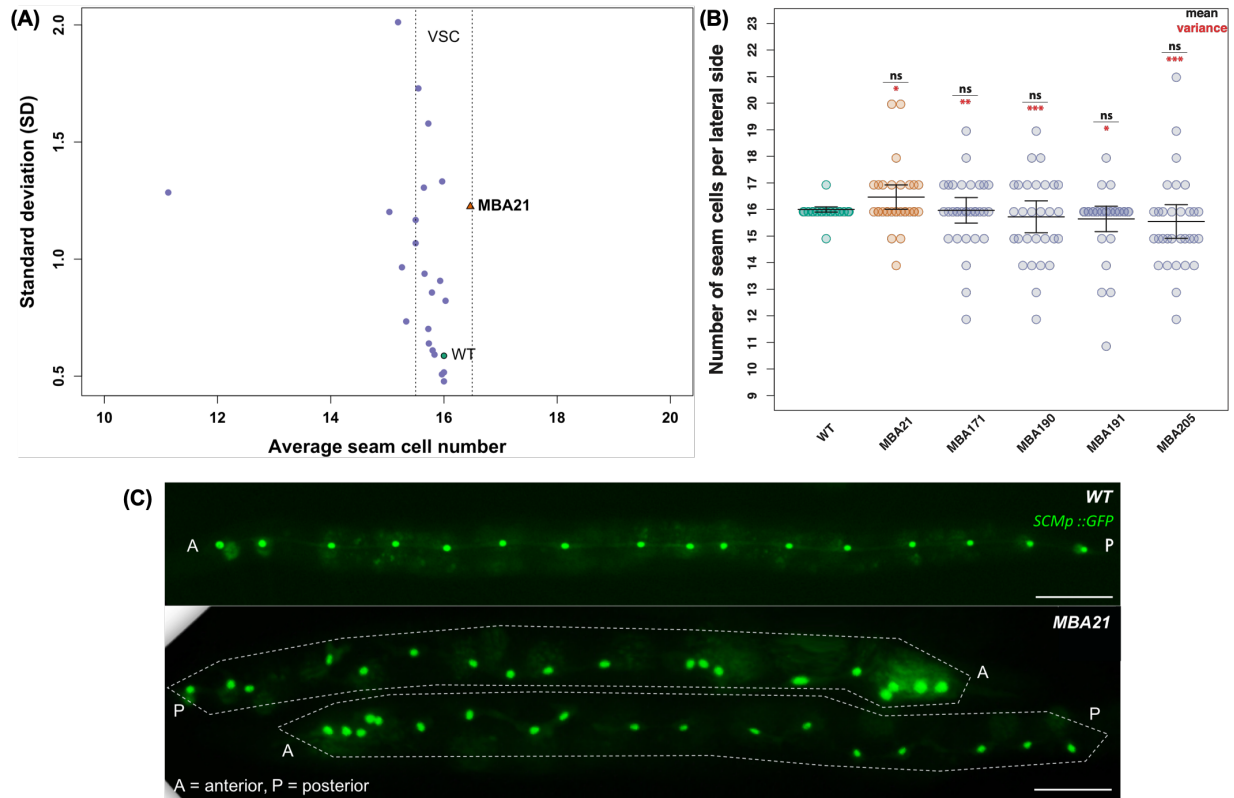


Figure 3.1: **Seam cell number mutants recovered from mutagenesis screen focusing on developmental variability.** (A) Relationship between average SCN and SD. Each point represents a mutant from the EMS screen. Parental wild-type strain (JR667) and mutant strain (MBA21) are coloured green and orange, respectively. Mutants with same average SCN but higher SD compared to wild-type (fall inside the vertical dashed lines) are considered variable seam cell (VSC) mutants. $26 \leq n \leq 37$ per strain. (B) SCN in selected VSC mutants. One-way ANOVA showed that there was no statistically significant effect of strain on seam cell number ($F(5, 177) = 1.87$, $p = 0.1$). $29 \leq n \leq 32$ per strain. Error bars indicate average SCN \pm 95% confidence intervals. Black stars/letters show statistically significant changes in the average SCN by post hoc Tukey's HSD test, and red stars depict changes in variance with a Levene's median test compared to wild-type strain. In both cases, *** corresponds to $p < 1 \times 10^{-4}$ and ns corresponds to "not significant". (C) Representative images of wild-type and MBA21 mutant. *C. elegans* have 16 *SCMp::GFP* positive seam cells at the end of L4 division. Note the dumpy phenotype, and uneven distribution of seam cells in the mutant. MBA21 has more cells in the anterior part of the worm compared to wild-type. A = anterior, P = posterior. Scale bars are 50 μm

the reference *C. elegans* genome and variants (nucleotide differences present in mapping strain compared to the reference genome) at $\approx 10^5$ SNP loci were detected. Allele frequencies were calculated at known SNP positions as the number of sequencing reads containing the variant allele (Hawaiian) divided by the total number of reads. Genomic region containing the causal mutation will be devoid of variants and reveals the mapping interval and have an allele frequency of zero. Mutation in MBA21 mapped to ≈ 6 Mb – ≈ 10 Mb region in the middle of chromosome II (Fig. 3.2). EMS mutants of course carry many other mutations in the genome in addition to the causative mutation(s) for the phenotype of interest. In order to narrow down the causative mutation, we filtered the list of mutations in ≈ 6 Mb – ≈ 10 Mb region of chromosome II for those affecting protein coding regions of genes (Table. 3.1). We found a C – T transition in the third exon of the *eff-1* gene that results in a premature stop codon (Q148STOP). *eff-1* encodes a fusogen that is required for most of the cell fusions that occur in the wild-type *C. elegans* development (Mohler et al., 2002). MBA21 is Dpy and phenotypically resembles the published *eff-1* mutant phenotype suggesting that the VSC mutant may be a loss-of-function mutant of *eff-1*. We call this new allele *icb4* and will be referred as such for the rest of the thesis.

Chr	Position	Ref	Change	Type	Hom/Het	Quality	Coverage	Transcript ID	Gene	Exon	Effect	old AA/new AA	Old codon/New codon	Codon Num (CDS)
II	6447343	C	T	SNP	Hom	1678.77	57	K06A1.4.1	nhr-22	4	NON SYNONYMOUS CODING	D/N	Gat/Aat	338
II	6669191	*	-AAAG	DEL	Het	198.73	42	T19D12.1	T19D12.1	5	FRAME SHIFT	-/-	-/-	679
II	6669202	*	+CC	INS	Het	49.73	49	T19D12.1	T19D12.1	5	FRAME SHIFT	-/?	-/CC	682
II	6735093	C	T	SNP	Hom	1205.77	39	T14B4.3	T14B4.3	2	NON SYNONYMOUS CODING	A/V	gCt/gTt	48
II	6736494	*	+C	INS	Hom	1940.73	49	T14B4.2	T14B4.2	2	FRAME SHIFT	-/?	-/C	94
II	6758434	C	T	SNP	Hom	1385.77	53	F41G3.3	F41G3.3	5	NON SYNONYMOUS CODING	P/S	Cca/Tea	447
II	6787427	C	T	SNP	Hom	1394.77	49	T13C2.6b	T13C2.6	13	NON SYNONYMOUS CODING	A/V	gCc/gTc	837
II	7123673	T	G	SNP	Het	31.77	42	F10E7.2	F10E7.2	3	NON SYNONYMOUS CODING	C/W	tgT/tgG	59
II	7167657	G	T	SNP	Hom	1507.77	50	C27H5.5	col-36	2	NON SYNONYMOUS CODING	P/Q	cCa/cAa	301
II	7488330	C	T	SNP	Hom	1205.77	35	D1022.9	D1022.9	5	NON SYNONYMOUS CODING	R/K	aGa/aAa	245
II	7865439	C	T	SNP	Hom	1318.77	41	T09A5.11.3	ostb-1	3	NON SYNONYMOUS CODING	H/Y	Cac/Tac	291
II	8347260	C	T	SNP	Hom	795.78	29	C26D10.5d	<i>eff-1</i>	3	STOP GAINED	Q/*	Cag/Tag	148
II	9104894	C	T	SNP	Hom	1331.77	42	T24H10.5	T24H10.5	5	NON SYNONYMOUS CODING	A/V	gCa/gTa	289
II	9893049	C	T	SNP	Hom	1414.77	46	C08H9.16	C08H9.16	1	NON SYNONYMOUS CODING	T/I	aCt/aTt	107

Table 3.1: **Mutation in MBA21 maps to third exon of *eff-1*.** List of mutations in protein coding region of candidate genes in 6 Mb – 10 Mb region of chromosome II. A C – T transition resulting in a premature stop codon was found in *eff-1*, a gene encoding a fusogen.

3.2.3 *icb4* fails to complement a mutant allele of *eff-1*

In order to confirm that *icb4* is indeed a new allele of *eff-1*, I performed a genetic complementation test with a known mutant allele (*hy21*) of *eff-1*. F1 hermaphrodites carrying *icb4* and *hy21* in trans produced the mutant variable-SCN (Fig. 3.3A). One-way analysis of variance (ANOVA) confirmed that there is a significant effect of strain on the SCN ($F(3, 143) = 7.23$, $p = 2e - 04$). Post hoc Tukey HSD tests showed that the SCN in *eff-1(hy21/icb4)* animals is significantly different from wild-type and *eff-1(hy21)* showing that *icb4* does not complement *hy21* ($p < 0.001$). Moreover, there is no statistically significant difference in SCN between *eff-1(hy21/icb4)* and

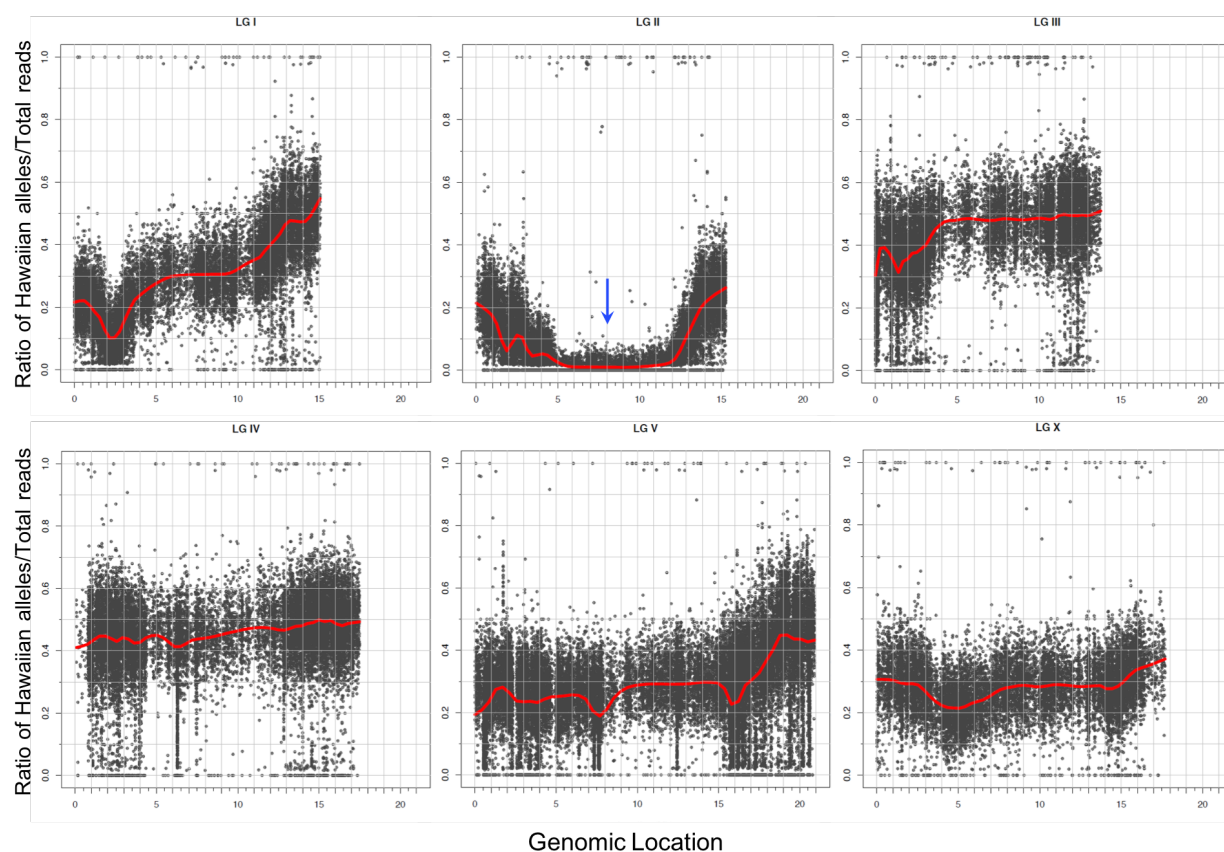


Figure 3.2: **Mutation in MBA21 maps to chromosome II (6 Mb – 10 Mb).** Allele frequency of CB4856 SNPs along the six chromosomes. Scatter plots for each *C. elegans* chromosome denoting the ratio of sequencing reads carrying Hawaiian allele to the total number of sequencing reads. Red curves represent locally weighted scatterplot smoothing (LOESS) regression lines from the allele frequencies at known SNP positions along the chromosomes with a span parameter of 0.1. Blue arrow points to the enrichment of N2 reads on chromosome II between 6 Mb – 10 Mb.

eff-1(icb4) ($p = 0.15$). Levene’s median test showed that there was a significant increase in the SCN variability between wild-type and *eff-1* mutants ($p < 0.03$) suggesting that lack of cell fusion increases variability in SCN. Interestingly, there was a significant difference in the average and variance of SCN between *eff-1(hy21)* and *eff-1(hy21/icb4)* suggesting that *icb4* is likely to be a stronger allele than the hypomorphic *hy21* allele ($p < 0.001$).

To understand the effect of severity of the loss of function of *eff-1* on SCN variability, I counted SCN in different alleles of *eff-1* (Fig. 3.3B). Two lines of evidence suggest that *icb4* is a strong loss-of-function allele of *eff-1*. First, *eff-1(icb4)* animals are dumpy like strong loss-of-function *eff-1* mutants (e.g. *ku433*, *zz10* and *zz8* alleles) in contrast to weaker *eff-1* mutants (e.g. *zz7*, *oj55* and *zz1* alleles), which have wild-type body length (Shinn-Thomas et al., 2016). Second, there is no statistically significant difference in SCN or SCN variability in animals carrying *icb4* allele compared to *ok1021* allele in which 7 out of the 8 exons of the *eff-1* gene have been deleted, and is considered functionally null (Sapir et al., 2007; Oren-Suissa, Hall, et al., 2010) ($p > 0.64$). Animals carrying a strong loss-of-function allele *ok1021* have a significant difference in SCN and SCN variability compared to wild-type ($p < 0.001$), and animals carrying a weak loss-of-function allele *oj55* (S441L) do not have a significant difference in SCN or in SCN variability compared to wild-type ($p > 0.18$). Animals carrying hypomorphic allele *hy21* (P183L) do not have a significant difference in SCN ($p = 0.77$) but significant difference in SCN variability ($p = 4e - 04$) compared to wild-type. Taken together, these results suggest that there is an increase in SCN variability based on the severity of the loss of cell fusion.

3.2.4 Phenotypic characterisation of *eff-1(icb4)* mutant

Although *eff-1* was previously studied in the context of cell fusion, it was unexpected to see changes in the seam cell number. To understand the phenotypic consequences of loss of function of *eff-1*, we characterised the *eff-1(icb4)* mutant qualitatively by scanning electron microscopy (SEM) and quantitatively by fluorescence microscopy. *eff-1* mutants display the morphological abnormalities such as a deformed tail spike and aberrant alae due to loss of cell fusion (Mohler et al., 2002). 100% of *eff-1(icb4)* animals display a bulbous tail (Fig. 3.4B) compared to the tapered tail in wild-type (Fig. 3.4A) from L1 larval stage. The alae appear discontinuous, in doublets and bifurcated in places (Fig. 3.4D) compared to wild-type alae, which are continuous

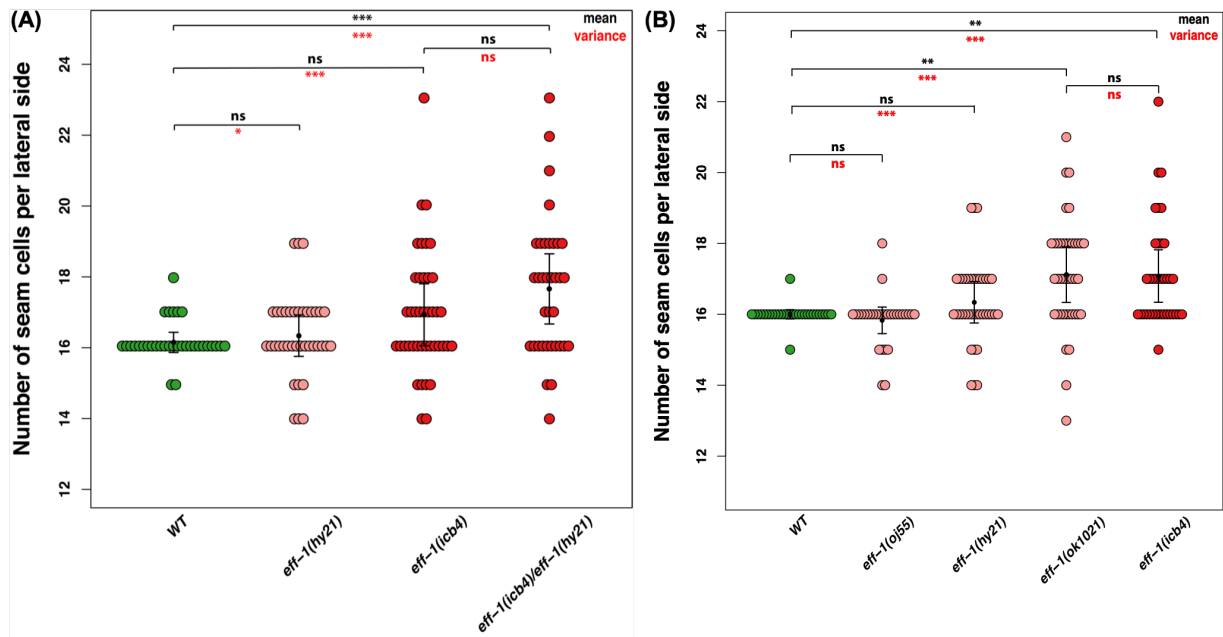


Figure 3.3: **The *icb4* mutation represents a strong loss of function of *eff-1*.** (A) *icb4* fails to complement mutant *hy21* allele of *eff-1*. $33 \leq n \leq 41$ per strain. One-way ANOVA showed that the effect of strain on SCN was significant ($F(3, 143) = 7.23$, $p = 2e - 04$). (B) The variability in SCN is increased in severe loss-of-function *eff-1* mutants. $30 \leq n \leq 40$ per strain. One-way ANOVA showed that there was a significant effect of strain on SCN ($F(4, 170) = 8.67$, $p = 2.17e - 06$). In both A and B, error bars indicate average SCN \pm 95% confidence intervals. Black stars show statistically significant changes in the average SCN by post hoc Tukey's HSD test, and red stars depict changes in variance with a Levene's median test. *** corresponds to $p < 1 \times 10^{-4}$.

and in triplets (Fig. 3.4C, Hall and Altun (2008)). The seam cells terminally differentiate, fuse at the end of L4 stage and secrete alae (C. Kenyon, 1986). The seam cells are misaligned and bifurcate like alae (Fig. 3.4F) compared to a straight line of seam cells in wild-type (Fig. 3.4E). The anterior seam cell daughters do not move out of the seam line and do not lose their *SCMp::GFP* immediately following division as in wild-type (white arrows in Fig. 3.4F). Seam cells in *eff-1* mutant adults do not have the characteristic eye-shaped morphology of wild-type seam cells. 90% (36/40) of *eff-1(icb4)* adults have breaks in the seam line compared to wild-type (Fig. 3.4H). 22% (9/41) of the mutant animals have clusters of seam cells near H1 in the head compared to 0% in wild-type. 38% (25/66) of the mutant animals have more than 3 seam cells in the anterior region (defined here as the region of the worm from the mouth to the end of pharyngeal bulb) in contrast to 3 seam cells in all wild-type animals. 27% (18/66) of the animals have additional clusters of seam cells in the mid-body compared to 0% in the wild-type.

3.2.5 Quantitative analysis of cell shape and cell division axis

Seam cell contact has been thought to be an important cue for the seam cells to stop elongating and for asymmetric cell division in the case of V5. V5 requires contact with its neighbours to produce postdeirid neuroblast (PDE) (Austin and C. Kenyon, 1994). Seam cell shape changes dynamically after every cell division to reconnect with their neighbours. *eff-1(icb4)* mutants are born with a fragmented hypodermis at L1 due to failure of fusion in embryogenesis (Mohler et al., 2002). To explore if the fragmented hypodermis affected seam cell shape, I characterised seam cells at the end of L1 asymmetric division using *ajm-1p::mCherry* to mark apical junctions in *eff-1(icb4)* animals with confocal microscopy (Fig. 3.5A and B). As reported by Mohler et al. (2002), I found that *eff-1(icb4)* indeed had a fragmented hypodermis unlike the syncytial hypodermis in wild-type animals. Interestingly, the seam cells in *eff-1(icb4)* did not have the characteristic rectangular shape of wild-type animals.

Cell shape was affected in *eff-1(icb4)* animals compared to wild-type animals (Fig. 3.5C and D). Cell shape is dynamic in seam cell divisions; they round up before cell division, and elongate to connect with neighbours after cell division. To quantify cell shape in *eff-1(icb4)*, we measured descriptive cell shape parameters such as area, perimeter, minor and major axes. I performed

CHAPTER 3. IDENTIFICATION OF A NOVEL ROLE FOR FUSOGEN *EFF-1* IN THE ROBUSTNESS OF SEAM CELL NUMBER

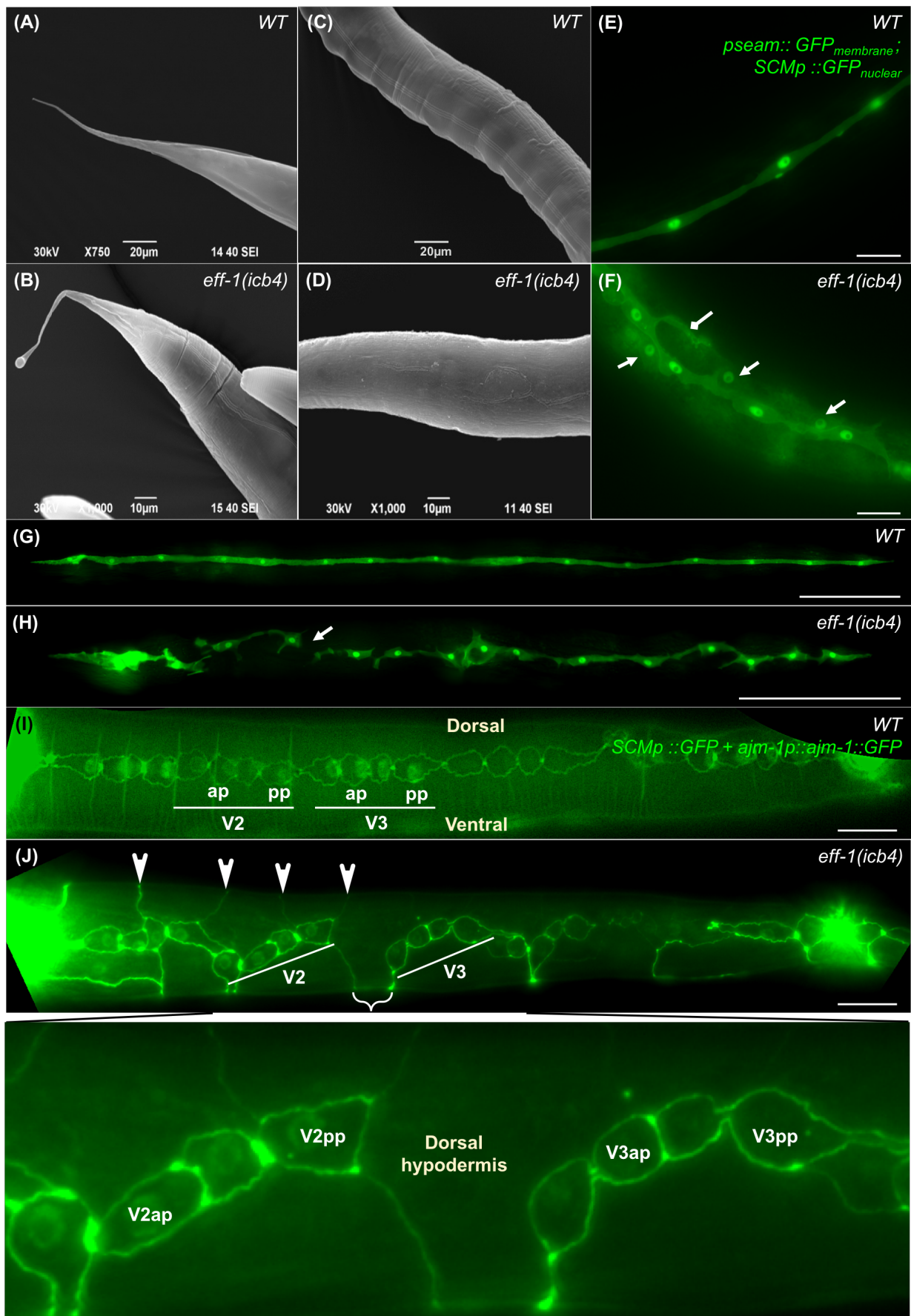


Figure 3.4: *eff-1(icb4)* animals display tail, alae and seam cell defects. Continued on next page.

CHAPTER 3. IDENTIFICATION OF A NOVEL ROLE FOR FUSOGEN *EFF-1* IN THE ROBUSTNESS OF SEAM CELL NUMBER

Figure 3.4: ***eff-1(icb4)* animals display tail, alae and seam cell defects.** (A,B) Representative SEM images of tail spike in young adult wild-type and *eff-1(icb4)* animals, respectively. Mutant animals have bulbous tail instead of tapered tail spike in wild-type animals. (C,D) Representative SEM images of alae in young adult wild-type and *eff-1(icb4)* animals, respectively. Mutant animals have cuticle defects and defective alae. (E, G, I) and (F, H, J) Representative images of seam cells in young adult wild-type and *eff-1(icb4)* animals, respectively. (E,F,G,H) Seam cells are visualised by *pseam::GFP* and *SCMp::GFP* markers. (F) Square arrow in points to a bifurcation of a seam cell in *eff-1(icb4)*. Triangular arrow points to *hyp7* seam daughters that are adjacent to the seam line. (H) Seam cells in *eff-1(icb4)* are misaligned to each other and move away from lateral position in contrast to wild-type animals. Triangular arrow points to a break in the seam line in *eff-1(icb4)*. (I,J) Seam cells are visualised by *ajm-1p::ajm-1::GFP* and *SCMp::GFP* markers. (J) The arrow heads point to intact apical junctions marked by *AJM-1::GFP* in *eff-1(icb4)* animals. The inset highlights the migration of dorsal hypodermis between the quadruplet of V2 and V3 seam cells. The bracketed area shows the loss of contact between V2 and V3 quadruplet in *eff-1(icb4)*. Scale bars in E,F,I and J are 20 μm . Scale bars in G and H are 100 μm . Animals in G, H, I, J are positioned from anterior to posterior (left to right). Images in G and H were acquired by an undergraduate student Fu Xiang Quah.

principal component analysis (PCA) on these shape parameters on individual (H0, H1, H2, V5 and T) or pooled (V1 – V4 and V6) seam cells based on their similarity in cell division pattern. Only the first two components displayed eigen values greater than 1 and scree tests suggested that the first two components were meaningful. The first two components were enough to account for > 90 percent of the total variance in the case of each cell. The corresponding amount of variance explained by each principal component is presented in appendix (Appendix B.2). I found that most of the seam cells (H1, H2, V1 – V4 and V6) in *eff-1(icb4)* are different in shape compared to wild-type (Fig. 3.5E). The seam cells in *eff-1(icb4)* are less elongated on the anteroposterior axis but extended on the dorsoventral axis in contrast to wild-type. Seam cells in *eff-1(icb4)* have a variable seam cell shape compared to wild-type as seen as scattering of the points in Fig. 3.5E (H1, H2, V1 – V4, V5, V6, T). There are no differences in shape in the most anterior seam cell (H0), which incidentally does not divide.

Elongated cell shape has been previously shown to be an important determinant for axis of cell division (Wildwater et al., 2011). To investigate if the aberrant seam cell shape in *eff-1(icb4)* animals affected cell division axis, we measured the angle between the seam cell daughters after L2 symmetric (proliferative) and L2 asymmetric cell division as shown in Fig. 3.6A,B,E and F. Seam cells (H1, V1 – V4 and V6) with the exception of H2 undergo proliferative division in L2. The magnitude of angle of cell division is lesser than 11° in wild-type as expected by the linear arrangement of seam cells (Fig. 3.6D). There is a significant difference in the angle of cell division (H1, H2, V1 – V4 and V6) in L2 symmetric division between *eff-1(icb4)*

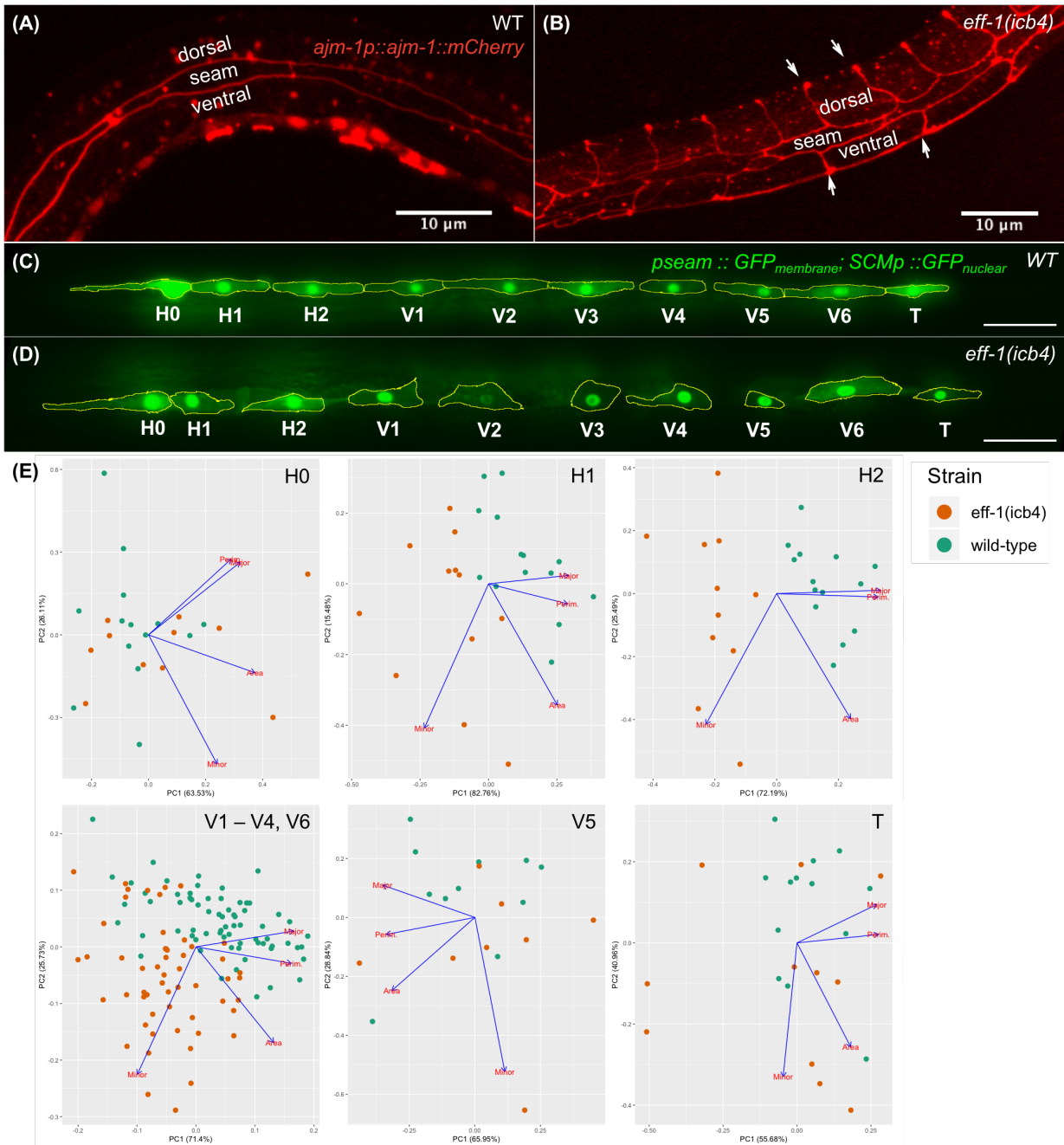


Figure 3.5: **Seam cell shape is affected in *eff-1(icb4)* mutants.** (A, C) and (B, D) Representative images of epidermis of late L1 larvae at the end of first asymmetric seam cell division in wild-type and *eff-1(icb4)* animals, respectively. (A,B) Apical junctions are marked with *ajm-1p::ajm-1::mCherry*. (A) Wild-type animals have fused dorsal and ventral hypodermis. Individual seam cells are in lateral position, rectangular in shape and have apical junctions between them. (B) *eff-1(icb4)* animals have unfused dorsal and ventral hypodermis. Dorsal and Ventral hypodermis in *eff-1* animals have apical junctions (white arrows) between them due to failure of cell fusion. Seam cells in *eff-1(icb4)* animals do not have the characteristic rectangular shape of wild-type. (C,D) Seam cell membranes are visualised by *pseam::GFP* marker and highlighted in yellow as detected by segmentation pipeline in Fiji. Note that the seam cells are misshapen in *eff-1(icb4)* compared to wild-type. Scale bars are 20 μm . (E) Principal component analysis of descriptive cell shape parameters in L1 wild-type and *eff-1(icb4)* animals. In this biplot, the arrows pointing in direction of the variables used in PCA. Individual cells are plotted with respect to first and second principal components. Together they accounted for more than > 90 of the total variance. Continued on the next page

Figure 3.5: **Seam cell shape is affected in *eff-1(icb4)* mutants.** Green and orange dots correspond to seam cells in wild-type and *eff-1(icb4)*, respectively. Note that the seam cells (H1, H2, V1 – V4 and V6) are less elongated in *eff-1(icb4)* compared to wild-type. Fluorescent seam cell images in C and D and raw cell shape were acquired by an undergraduate student Fu Xiang Quah.

and wild-type animals ($p < 0.05$). These two cells of V1 – V4 and V6 undergo asymmetric cell division in L2 and form pairs of four cells. We measured two angles of division (aa-ap and pa-pp) and (ap-pa) as shown in Fig. 3.6G, which is related to the cell division angle from previous division (a-p from Fig. 3.6C). There is a significant difference in the angle of cell division (V1 – V4 and V6) in L2 asymmetric division between *eff-1(icb4)* and wild-type animals ($p < 0.001$). Also, there is a significant difference in the angle between ap-pa (formerly, a-p angle in Fig. 3.6C) ($p = 5.59e - 15$). Interestingly, there is a significant difference between the angles between a-p compared to ap-pa ($p = 3.26e - 08$) in mutant animals in contrast to wild-type ($p = 0.66$) suggesting an increase in the misalignment between seam cells from L2 symmetric and asymmetric division.

3.2.6 Developmental basis for the seam-cell-number variability in *eff-1(icb4)*

To understand the two-sided errors of SCN, we performed long-term time-lapse imaging of postembryonic cell divisions to understand the developmental basis of the SCN variability in *eff-1(icb4)* in microchambers with 20 min resolution (Gritti et al., 2016). 18 seam lineages were imaged and lineaged manually as described in section 2.2.4 (Fig. 3.7B,C,D). We analysed 18 lineages (Fig. C.1 and Fig. C.2.) and found there were errors that changed SCN (Fig. 3.7D) and errors that did not change SCN (Fig. 3.7E). Surprisingly, most of the seam lineages were wild-type based on the error frequencies.

Most frequent cell division errors were symmetrisation toward seam cell fate causing an increase in the SCN. There was symmetrisation of asymmetric cell divisions, i.e., gain of seam cell fate based on the ability to divide subsequently in H1 and H2 in L1 stage resulting in clusters of seam cells in the anterior part of the animal (Fig. 3.7B,C and Fig. 3.1C). The symmetrisation errors were highest in the L4 stage compared to other stages. There were other rare seam cell losses in L1, L2 and L3 stages followed by gain of seam cell fate in the subsequent divisions in L4 suggestive of a compensatory mechanism. For instance, in lineage shown in Fig.

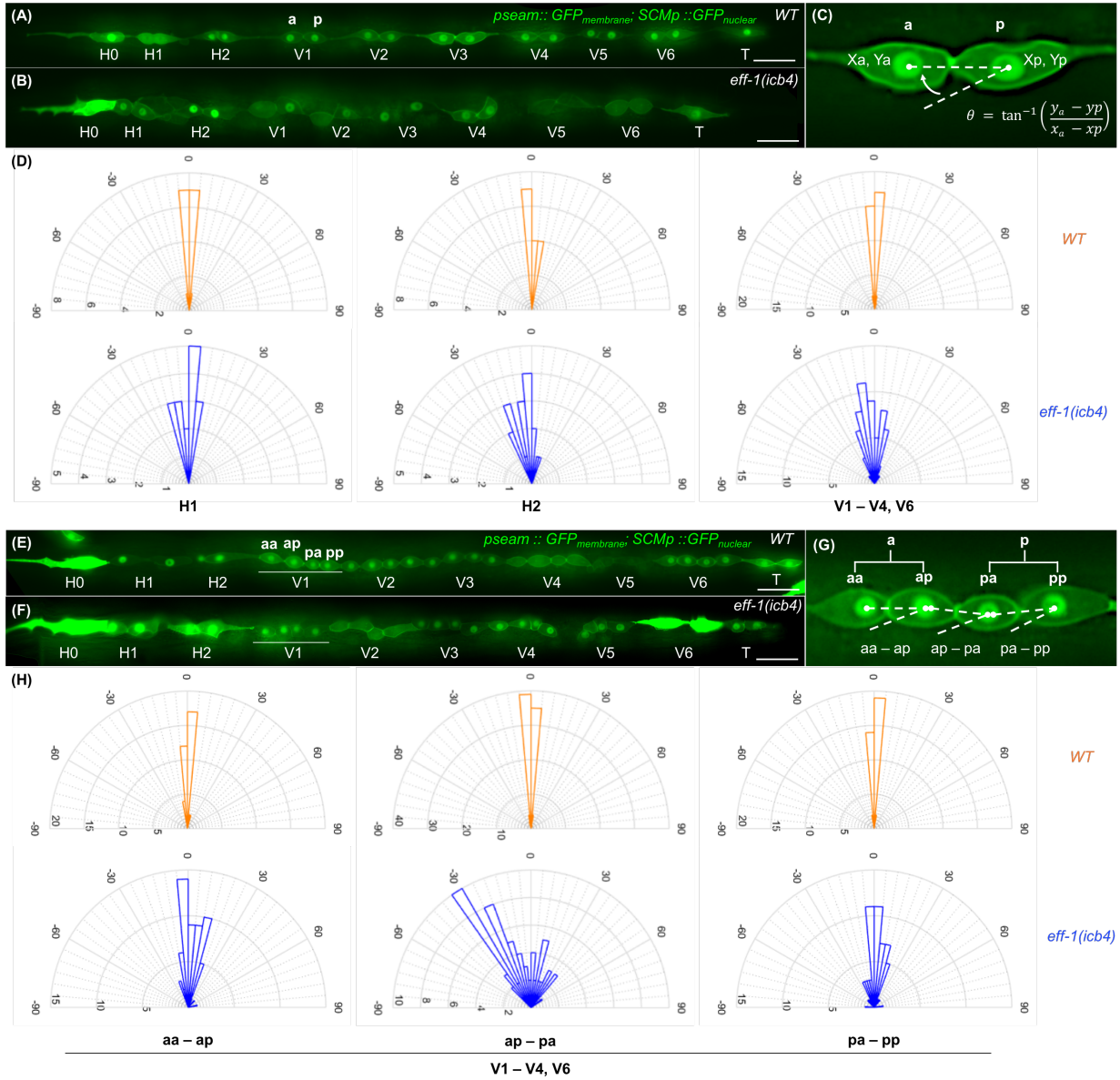


Figure 3.6: Seam cells become progressively misaligned between L2 symmetric and L2 asymmetric cell division in *eff-1(icb4)*. (A,B) Representative images of seam cells in L2 symmetric in wild-type and *eff-1(icb4)* animals, respectively. Seam cells undergo proliferative seam cell division and seam cell daughters (a and p) appear in pairs. (C) The angle between pairs of cells (a and p) is calculated as shown in the figure. (D) Rose plots showing angles between pairs of cells (a-p) in *eff-1(icb4)* and wild-type, respectively. One-way ANOVA showed a significant difference in the angle between pairs of cells in *eff-1(icb4)* (ANOVA tables are presented in appendix (Appendix B.1). $10 \leq n \leq 21$ per cell per strain. (E,F) Representative images of seam cells in L2 asymmetric cell division in wild-type and *eff-1(icb4)* animals, respectively. (G) Angle between pairs of cell daughters of a and p (e.g., aa and ap) as shown in the figure is calculated. aa-ap and pa-pp are angle of cell division. Angle between ap-pa corresponds to angle between a-p in the previous stage. (H) Rose plots showing angles between pairs of cells (aa-ap, ap-pa and pa-pp) in *eff-1(icb4)* and wild-type, respectively. One-way ANOVA showed a significant difference in the angle between pairs of cells in *eff-1(icb4)* (ANOVA tables are presented in appendix (Appendix B.1). $38 \leq n \leq 61$ per cell pair per strain. Seam cells are visualised by *pseam::GFP* and *SCMp::GFP* markers. Scale bars in A, B, E and F are 20 μ m. Fluorescent seam cell images, raw angles data and graphs were generated by an undergraduate student Fu Xiang Quah.

3.7B, V1 underwent symmetric division in L2 following the loss of V2 lineage in L1. In one case, there were a loss of seam cell fate in H1 and H2 lineages in L4 and L3 stages, which could cause a decrease in SCN in Fig. 3.7D. The lineaging errors suggest that there are low frequency random errors that occur in all lineages except H0 in different larval stages to produce variable SCN.

Seam cells fate is not affected in most of the cell divisions in *eff-1(icb4)* animals as shown in Fig. 3.7, however, there was a proportion of animals with increased SCN. In order to test if this increase of SCN in *eff-1(icb4)* animals is dependent on GATA transcription factor *egl-18*, a repressor of *eff-1*, I counted SCN in double mutant *eff-1(icb4); egl-18(ga97)* (Fig. 3.8). In *egl-18* mutants, seam cell fate is lost and they fuse to the hypodermis resulting in decreased SCN (Koh and Rothman, 2001; Gorrepati, K. W. Thompson, and Eisenmann, 2013). There was a significant decrease in SCN in double mutant *eff-1(icb4); egl-18(ga97)* compared to wild-type and individual mutants ($p < 0.04$) suggesting that the seam cell increase in *eff-1(icb4)* requires *egl-18*.

3.2.7 Quantification of *eff-1* mRNA transcripts in seam cells with single molecule fluorescence in situ hybridisation imaging

It is thought that *eff-1* expression is tightly regulated to ensure wild-type development (Shemer, Suissa, et al., 2004). *eff-1* expression occurs in bursts in the anterior seam cell daughters (Vaa, Vpa) but not posterior seam cell daughters (Vap, Vpp) (Katsanos et al., 2017). I used single molecule mRNA in situ hybridisation imaging (smFISH) to quantify *eff-1* mRNA transcripts to discover dynamics of *eff-1* at the level of the seam tissue in L2 asymmetric cell division (Fig. 3.6E). As reported previously by Katsanos et al. (2017), *eff-1* was expressed in bursts in anterior seam cell daughters (Vaa, Vpa) of the pairs of V cells (Fig. 3.9A). Interestingly, I found a new pattern of *eff-1* expression in the seam cells. There was significantly higher *eff-1* expression in Vaa compared to Vpa for V2 – V6 ($p < 0.01$, 3.9B). However, V1 did not follow this pattern, there was no statistically significant difference in the *eff-1* expression between V1aa and V1pa ($p = 0.35$).

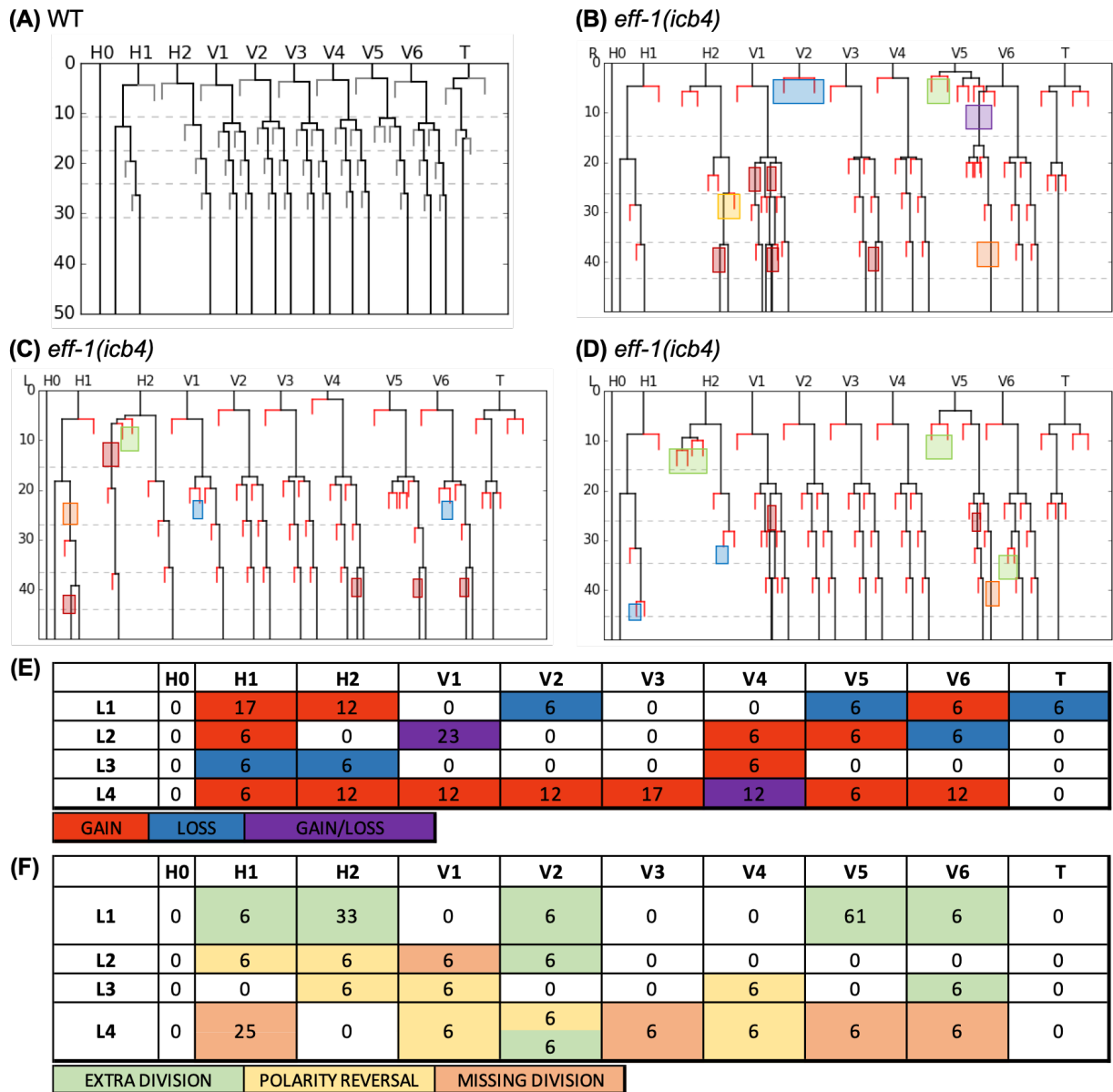


Figure 3.7: **Developmental basis of SCN variability in *eff-1(icb4)*.** (A),(B,C,D) Representative seam cell lineages of wild-type and *eff-1(icb4)* animals, respectively. The coloured boxes in (B,C,D) highlight the different errors coded as shown in (E) and (F). Note the loss of V2 lineage in L1 and gain of V1 lineage in L2. This animal has a final SCN of 19 despite the complete loss of V2 lineage. (E) Summary of percentage of lineage errors which cause increase or decrease SCN. Notice the higher percentage of symmetrisation errors (coloured red) in L4 stage, and in H1, H2 seam cells in L1 stage. (F) Summary of percentage of lineage errors which do not increase or decrease SCN. Most of these errors are rare (1 in 18 lineages) but are atypical compared to wild-type. Notice the extra divisions in H2 and V5 in L1 stage and missed asymmetrical divisions in L4. The errors per seam lineage were grouped based on the developmental stage (L1 – L4). $n = 18$ lineages. The seam lineages in A,B,C and D were generated by Ritobrata Ghosh, a research assistant in the lab

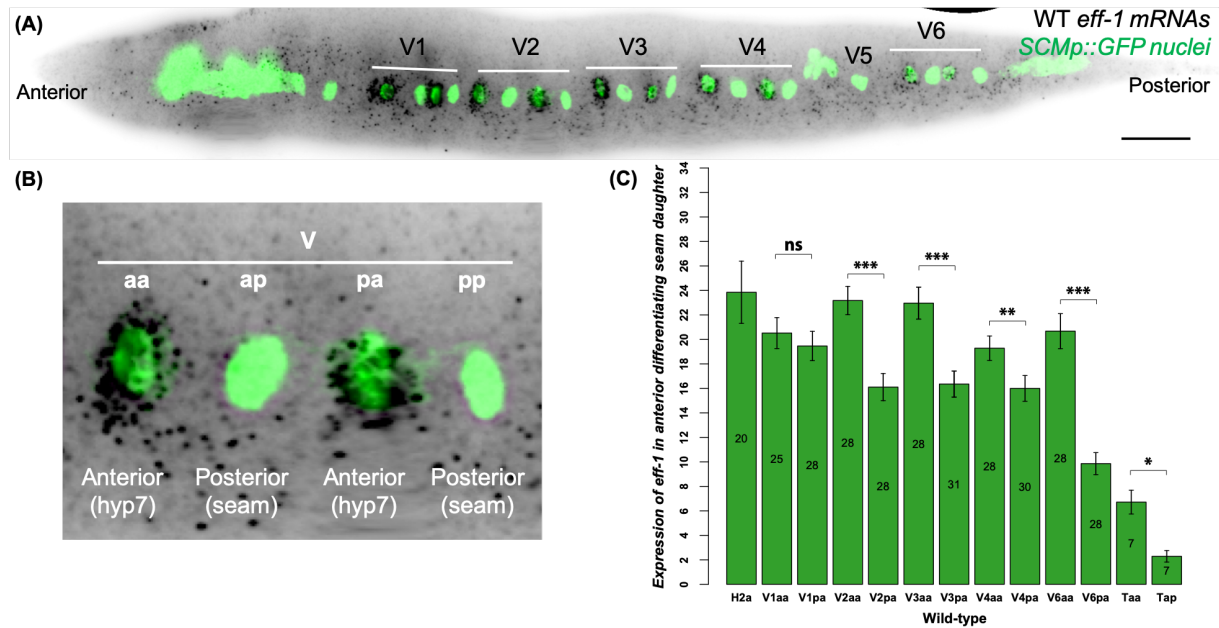


Figure 3.9: *eff-1* is expressed in anterior seam cell daughters. (A). Schematic of single molecule fluorescence in situ hybridisation. 48 probes labelled with fluorophore Cy5 were used to detect *eff-1* mRNAs. (B). A diagram of L2 asymmetric seam cell division. A pair of cells (Va and Vp) per V cell undergo asymmetric cell division in L2 stage. The relevant four-cell stage in which mRNAs were measured is highlighted by a green panel. (C). Representative smFISH image showing *eff-1* expression in anterior seam cell daughters. Seam cells are labelled in green due to *SCMp::GFP* expression and black spots correspond to mRNAs in wild-type V cells after the L2 asymmetric division. Scale bar is 20 μm . (D). Quantification of *eff-1* mRNAs in anterior seam cell daughters after the L2 asymmetric division. Error bars indicate 95% confidence intervals. Number of cells analysed is printed inside the bars. $20 \leq n \leq 31$. Black stars shows statistically significant differences in *eff-1* expression with pairwise t-test. *** corresponds to $p < 1 \times 10^{-4}$.

3.2.8 *eff-1* is not required for the differentiation of anterior seam daughters

To investigate the developmental state of anterior daughters in *eff-1(icb4)*, I used two approaches. First, a new single-copy *dpy-7* reporter using the full *dpy-7* promoter sequence was made. *dpy-7* is expressed primarily in anterior fated daughters and at a low level in seam-fated daughters (Fig. 3.10A and C). I found that anterior daughters in *eff-1(icb4)* animals expressed *dpy-7::mCherry* like wild-type animals. Interestingly, the breaks in seam line correlated with the presence of *dpy-7*-positive cells, which have not moved out of the seam line (Fig. 3.10B). I confirmed that these *dpy-7*-positive cells expressed *AJM-1::GFP* boundaries ectopically using *ajm-1::gfp + SCMP::GFP* marker (Fig. 3.10D). Second, to address the fate of these cells, I used smFISH to look at the expression of a seam cell-specific marker *nhr-73* (Miyabayashi et al., 1999; Cao et al., 2017). As expected, I found *nhr-73* to be expressed exclusively by the posterior seam cell daughters in wild-type (Fig. 3.10E). Like in wild-type, posterior seam cell daughters and not anterior expressed *nhr-73* in *eff-1(icb4)* animals suggesting that the anterior daughter cells are not seam-fated (Fig. 3.10F). Taken together, the presence of *dpy-7* hypodermal marker and absence of *nhr-73* seam marker in the anterior seam cell daughters suggests that these cells are not stuck in developmental limbo and are differentiated to the hypodermal fate.

3.2.9 Differentiation of seam cells to hypodermal cells in *eff-1(icb4)* may be dependent on *nhr-25*

Seam cells are able to differentiate in *eff-1(icb4)* animals despite the lack of fusion, which uncouples *eff-1* from the differentiation programme. Therefore, I sought to identify factors required for differentiation. *nhr-25* encodes a nuclear hormone receptor orthologous to *Drosophila* Fushi tarazu transcription factor (ftz-F1), and is known to regulate vulval and seam cell differentiation (Chen, Eastburn, and Han, 2004; Silhánková, Jindra, and Asahina, 2005). To investigate if *nhr-25* is required for the ability of anterior seam daughters to differentiate despite the lack of fusion, I performed RNAi knockdown of *nhr-25* in *eff-1(icb4)* animals. There is a significant increase in SCN in *eff-1(icb4)* animals upon *nhr-25* knockdown compared to *eff-1(icb4)* animals

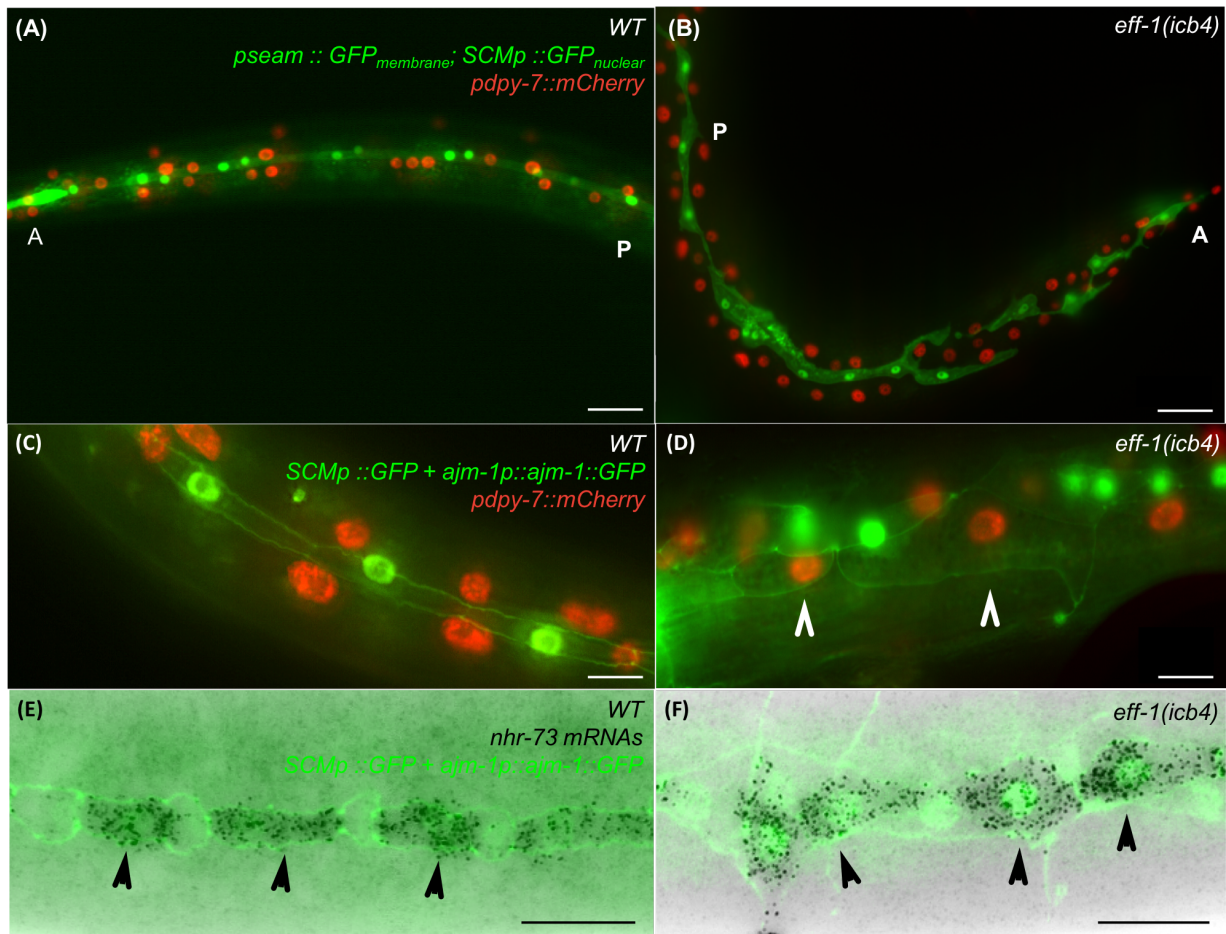


Figure 3.10: **Anterior seam daughters differentiate in the absence of EFF-1.** (A,B) Representative images of seam cells and hyp7 cells in adult wild-type and *eff-1(icb4)* animals, respectively. Seam membranes and nuclei are visualised by *pseam::GFP* and *SCMp::GFP* markers, respectively. Hyp7 cells are visualised by *pdp-7::mCherry* marker. The anterior daughter differentiates, and expresses *dpy-7* at a higher level compared to the posterior cell. Note the presence of *dpy-7*-positive hyp7 daughters in *eff-1(icb4)* animals. (C,D) Representative images of seam and hyp7 cells in an adult wild-type and *eff-1(icb4)* animals, respectively. Seam cells and hyp7 cells are visualised by *ajm-1::gfp + SCMp::GFP* and *pdp-7::mCherry* markers, respectively. Arrow heads point to cell with apical junction boundary expressing *dpy-7* marker. *eff-1(icb4)* animals have hyp7 cells with ectopic AJM-1::GFP boundaries. (E,F) Representative smFISH images of seam and hyp7 cells in adult wild-type and *eff-1(icb4)* animals, respectively. The posterior cell expresses seam specific markers like *nhr-73*. Seam cells are labelled in green due to *ajm-1::gfp + SCMp::GFP* expression and black spots correspond to *nhr-73* mRNAs in L2 asymmetric cell division. Black arrow heads point to seam cell expressed *nhr-73*. Note the lack of expression of *nhr-73* in anterior daughters in both wild-type and *eff-1(icb4)* animals. Scale bars in A and B are 50 μ m. Scale bar in C,D,E and F is 10 μ m.

pared to wild-type animals ($p_{\text{variance}} = 3e - 04$, $p_{\text{average}} = 0.64$). Secondly, these transgenic animals showed similarities to *eff-1(icb4)* such as clusters of seam cells in the head and bifurcations in the seam (Fig. 3.12B,C,D). Curiously, these animals with the strongest variable SCN were also extremely dumpy. The expression of the seam marker *SCMp::GFP* was seen in the hypodermal nuclei.

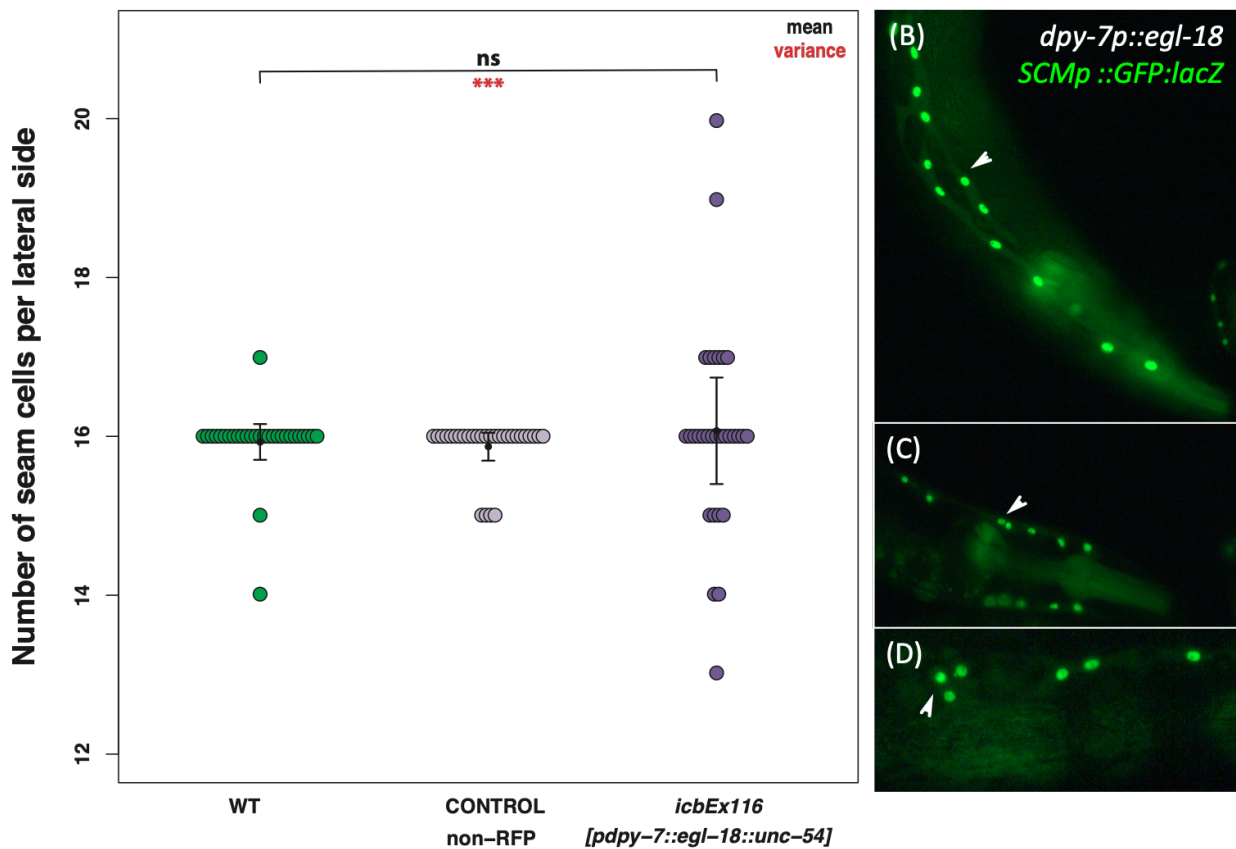


Figure 3.12: Overexpression of *egl-18* in anterior seam daughters increase SCN variability. (A) Seam cell number upon overexpression of *egl-18* in anterior seam daughters. There was a significant difference in SCN variability ($p = 3e - 04$) in animals overexpressing *egl-18* compared to wild-type animals. However, one-way ANOVA showed that there was no effect of strain on SCN ($F(2, 87) = 0.44$, $p = 0.64$). $n = 30$ per strain. Error bars indicate average SCN \pm 95% confidence intervals. Black stars show statistically significant changes in the average SCN, and red stars depict changes in variance with a Levene's median test. In both cases, *** corresponds to $p < 1 \times 10^{-4}$. (B, C, D) Transgenic animals carrying *dpy-7p::egl-18* phenocopy *eff-1* loss-of-function mutants. (B) Arrow head points to a bifurcation in the anterior part of the worm. (C) Arrow head points to cluster in head of the animal. (D) Arrow points to cluster in the mid-body of the animal.

3.3 Discussion

3.3.1 *EFF-1* is required for robust seam cell patterning

An open question about developmental robustness is whether robustness-conferring genes belong to developmental gene networks or are core cell regulators like chromatin modifiers or homeostasis factors such as molecular chaperones (Rutherford and Lindquist, 1998; Queitsch, Sangster, and Susan Lindquist, 2002; Levy and Siegal, 2008). To uncover the mechanisms underpinning developmental robustness in multicellular organisms, I use the seam cell number as a model. We recovered *eff-1(icb4)* from an unbiased forward genetic screen for variable seam cell number (VSC) mutants. VSC is a proxy for loss of developmental robustness. The mutation *icb4* represents a new strong loss-of-function allele of *eff-1*. I found that *eff-1(icb4)* affects SCN variance but not the average SCN. Therefore, I describe a new role for *eff-1* in the robustness of SCN patterning, which has only been studied as a fusogen that facilitates differentiation of hypodermal cells and morphogenesis of neurons (Mohler et al., 2002; Oren-Suissa, Hall, et al., 2010). The results in this chapter taken together indicate a new mechanism by which core components of the seam cell gene network can modulate seam cell number variance.

There is an increase in SCN variance with the severity of the mutation in *eff-1* showing that the timely cell fusion of anterior seam cell daughters promotes SCN robustness. To test this hypothesis, I suppressed cell fusion in anterior daughters by expressing *egl-18*, a GATA transcription factor known to repress *eff-1* (Cassata et al., 2005). I found that overexpression of *egl-18* causes two-sided errors of SCN. The increase in SCN can be thought of as conversion of anterior cells to seam cell fate by the overexpression of *egl-18*. However, there is a decrease in the SCN as well, which can happen by the rare loss of seam cell fate like in *eff-1(icb4)* animals. Overexpression of *egl-18* in anterior cells also phenocopies seam cell phenotypic characteristics of *eff-1* mutant such as bifurcations and clusters of seam cells in the anterior region and the mid-body of the animals. My prediction is that a delay in fusion should lead to SCN variability without a change in the mean.

Yang et al. (2017) showed that there is a positive feedback loop between spectraplakins/VAB-10A, *EFF-1* and F-actin that leads to accumulation of *EFF-1* at fusion synapses. They quan-

tified cell fusion time from birth to fusion of anterior seam cell daughters to be 219 ± 42 min in wild-type animals. The time from birth to fusion of anterior seam cell daughters was significantly prolonged in *wsp-1* and *vab-10a* conditional mutants. In addition, overexpression of VAB-10 fusogen binding domain also delays fusion. Therefore, I could investigate if this delay in fusion is enough to cause SCN variability seen in *eff-1* mutants. These experiments could also elucidate the consequences of lack of contact between seam cells for seam cell patterning.

I show with smFISH that *eff-1* is expressed asymmetrically in the anterior seam daughters. Consistently, Yang et al. (2017) found that *eff-1* was enriched at the fusion sites 155 ± 20 min after anterior daughter (Vaa/Vpa) formation in L2 asymmetric cell division using endogenously tagged *EFF-1::GFP* knockin. Interestingly, I found that there was a significant difference in the *eff-1* transcripts between Vaa and Vpa for V2 – V4 and V6 that has not been characterised. This may suggest that the anterior pair (Vaa-Vap) may be ahead in cell division compared to the posterior pair (Vpa-Vpp). In support of this observation, there seem to be instances where the anterior pair divided first compared to the posterior pair (Fig. 2b, Gritti et al. (2016)). Unfortunately, Yang et al. (2017) have not reported *EFF-1::GFP* fluorescence in individual seam cells along the length of the animal in order to compare the dynamics at the protein level. It would be interesting to see if the differences at the transcript level persist at the level of protein and if this is specific to L2 asymmetric division or to other larval divisions as well.

3.3.2 *eff-1* is not required for differentiation of anterior seam cell daughters

There are more patterning errors that increase SCN compared to errors that decrease SCN. One potential reason for this is that *eff-1* may be required for differentiation of anterior seam cells daughters. However, if *eff-1* was essential for differentiation, we would expect an increase of SCN by an addition 119 anterior cells that did not differentiate. We do not observe this even in deletion mutation of *eff-1*. The increase in SCN in *eff-1(icb4)* is suppressed by loss of function of *egl-18* suggesting that the increase in SCN depends on seam cell proliferative potential and not just lack of differentiation. The complete lack of postembryonic hyp7 cells could be lethal in *eff-1* mutants. In support of this, despite the severe epidermal defects in *eff-1* mutants, they are viable suggesting that anterior seam cell daughters differentiate

to hypodermal (*hyp7*) fate correctly. Brabin, Appleford, and Woollard (2011) claimed that *eff-1* is required for differentiation of anterior cells as they retained apical junctions and did not express *dpy-7p::yfp*, a hypodermal marker in *eff-1* mutant. This may be expected because apical junctions are considered important for maintenance of seam cell identity and *eff-1* causes breakdown of the apical junctions, facilitating differentiation in the anterior daughter following an asymmetric cell division (Brabin and Woollard, 2012). Contrary to this pattern, I found that anterior cells that retained apical junctions expressed *dpy-7p::mCherry*. There could be multiple reasons for this discrepancy, one obvious reason is the transgene *arIs99* used only 216 bp of the *dpy-7* promoter and not the full 317 bp promoter like I used. Perhaps, the smaller *dpy-7* promoter is missing cis-regulatory sites that are essential for the full expression pattern of *dpy-7* in hypodermis.

Anterior cells expressing *AJM-1::GFP* ectopically in *eff-1* mutant are considered to be undifferentiated and in developmental limbo as they do not express *SCMp::GFP* nor *dpy-7p::yfp* (Brabin, Appleford, and Woollard, 2011; Brabin and Woollard, 2012). They concluded from this observation that the differentiation signal is received only after apical junction boundary breakdown by EFF-1, and the hypodermal fate is adopted. Contrarily, I found that anterior seam daughters in *eff-1* mutant did not express highly seam-specific marker *nhr-73*. Unlike *SCMp::GFP*, which is expressed at first by both daughters and then fades in the anterior daughter, *nhr-73* is specifically expressed in seam cells (≈ 22 transcripts per million in seam cells compared to 0 per million in non-seam hypodermis Cao et al. (2017)). The presence of hypodermal marker *dpy-7* and absence of seam marker *nhr-73* in anterior seam cell daughters in *eff-1* mutant suggests that they have an intrinsic differentiation programme that is not dependent on extrinsic clues from *hyp7*. Interestingly, *SCMp::GFP* fades slower in anterior seam daughters in *eff-1* mutant animals suggesting that differentiation may be delayed. Taken together, fusion programme may act redundantly with the differentiation programme for the timely differentiation of anterior seam cell daughters. *nhr-25*, a nuclear hormone receptor may be required for differentiation of anterior cells in *eff-1* as knockdown of *nhr-25* increases SCN drastically.

3.3.3 Loss of *eff-1* may affect seam cell patterning indirectly by disrupting seam cell contacts

eff-1 is a surprising candidate in a screen for modulators of variance because *eff-1* mutants are viable and cell patterning has been reported to be normal despite the complete failure of fusion (Mohler et al., 2002). However, this discrepancy may be because postembryonic divisions at the level of animal have not been investigated in *eff-1* mutants. Contrary to the literature (Huang et al., 2009), we found that there are many errors in seam cell divisions in *eff-1* mutant. While most of them do not change SCN, they are highly atypical of wild-type animals. The most common errors that lead to increase in SCN were symmetrisation of H1 and H2 in the first postembryonic asymmetric division in L1. These errors seem to translate to one or two extra cells in the anterior part of the animal. The H1 asymmetric division is unusual in that the posterior cell differentiates to *hyp7* and the anterior cell inherits the seam cell fate in contrast to all other asymmetric seam cell divisions (Sulston and Horvitz, 1977). H1 and H2 seam cells produce two daughters between them in L1 stage that differentiate and fuse to *hyp7*. These cells do not move away and remain between H1 and H2. These cells sometimes gain proliferative potential and lead to increase in SCN.

eff-1 animals show significant cell shape changes due to unfused hypodermis in L1 and unfused anterior seam cell daughters that stay in the seam line. In the wild-type condition, seam cells are rectangular in shape. Gaps are left in the seam line after every asymmetric division as the anterior daughter differentiates and fuses with the *hyp7* syncytium. The seam cells elongate on both sides to contact their neighbours to form uninterrupted seam tissue through secreted semaphorin (*mab-20*) and plexin receptor (*plx-2*) (P. J. Roy et al., 2000; Fujii et al., 2002). In *eff-1* mutants, *mab-20* may cue the seam cells to extend in different directions to reach their neighbours and cause them to become misshaped.

The first L1 asymmetrical division followed by the proliferative cell division is responsible for misalignment seen in *eff-1* mutants. Wild-type L1 larvae hatch with a fused dorsal hypodermis, 10 seam cells and a fragmented ventral hypodermis consists of twelve P cells organised in two rows of six in the midbody (Sulston and Horvitz, 1977). The L1 stage in *C. elegans* is characterised by a number of cellular migration events that reorganise the ventral hypodermis

(shown in Fig. 1.5). The first postembryonic division in L1 stage is asymmetric and follows a distinct division pattern. The divisions of V2 – V6 cells contribute hypodermal cells exclusively to the ventral side. The anterior seam cell daughters fuse to hyp7 to cover the ventral surface of the animal (Podbilewicz and J. G. White, 1994; Austin and C. Kenyon, 1994). However, in *eff-1* animals, all cells remain unfused but migrate in a wild-type pattern causing cells to become misaligned. We show that this misalignment increases between L2 symmetrical and asymmetrical division (3.6).

Wild-type seam cell divisions are synchronised and they follow a stereotyped division pattern. This is due to the local cues transmitted from their neighbours through their contacts. Seam cell contact has been shown to be an important cue for asymmetric cell division of V5 to produce a postdeirid neuroblast in L2. Symmetrisation of V5 division has been reported upon the ablation of cells anterior or posterior to V5 (Waring, L. Wrischnik, and C. Kenyon, 1992; Austin and C. Kenyon, 1994). However, the effect of seam cell contact is not characterised in the context of other seam cells. The anterior daughters do not move out of the seam line in *eff-1* mutant due to the lack of fusion. The seam cells are misshapen and less elongated on the lateral surface in *eff-1* mutants. The seam cell daughters are thought to not reconnect properly with their neighbouring seam cells and have breaks in the seam line as a result (Brabin, Appleford, and Woollard, 2011). Wildwater et al. (2011) show that the elongated cell shape is an important cue for seam cell division orientation. They report that cell shape changes combined with inhibition of Wnt signalling resulted in seam cell patterning errors. I postulate that loss of seam cell contacts caused by lack of fusion may underlie seam cell patterning errors in *eff-1*. The signals that are transmitted between the seam cells through contacts are not known and remain to be investigated.

Chapter 4

Role of Natural Genetic Variation in Seam Cell Development

4.1 Introduction

Postembryonic seam cell development has been proposed as a simplified model for elucidating molecular mechanisms that underlie symmetrical proliferative and asymmetric maintenance of stem cell-like division patterns (Joshi et al., 2010; Brabin and Woollard, 2012). Seam cell development has been studied in the reference laboratory strain N2 (isolated initially from Bristol, UK), which has several adaptations suited to living in a petri dish with a continuous supply of food. N2 carries many laboratory-derived alleles for genes such as *npr-1* and *nath-10*, which influence a large number of phenotypes (Andersen, Bloom, et al., 2014; Sterken et al., 2015). For example, the N2 allele of *nath-10* affects many life-history traits like age at maturity, brood size, and egg-laying speed through an increase in the production of sperm. The *n2* allele was identified due to its effect on the vulval cell-fate specification. The *n2* allele compared to the ancestral *haw6805* allele was found to lower the effect of mutations in the epidermal growth factor receptor (EGFR) gene (*let-23*) (Duveau and Félix, 2012). Therefore, it remains to be investigated if any of these laboratory-derived alleles in N2 affect seam cell development and if N2 background is representative of the *C. elegans* species.

The highly genetically and developmentally tractable nature of *C. elegans* is ideal for elucidating genetic pathways and molecular mechanisms underlying development. However, mutations in genes affecting SCN have been discovered and studied only in the N2 background. It is known that mutations in different genetic backgrounds can vary in penetrance or expressivity (Fig. 1.1). Genetic background specific effects have been described in many model organisms like worms, flies, mice and even humans (Milloz et al., 2008; Dworkin, Kennerly, et al., 2009; Chandler, 2010; Abitbol et al., 2005; Cutting, 2010). Studying the effect of a mutation in only one background may even lead to an incomplete picture or the wrong conclusion. For example, *pha-1* has been considered to be a pharyngeal developmental gene based on its mutant phenotype. However, much to their surprise, researchers discovered that it was an antidote expressed to a maternal-effect toxin *sup-35* (Ben-David, Burga, and Kruglyak, 2017). Therefore, understanding this genetic background dependence of mutations is very profound and of great importance for the prognosis of human diseases (Chow, 2016; Gasch, Payseur, and Pool, 2016).

Mechanisms of robustness in development suppress the phenotypic effects of genetic varia-

tion in normal wild-type conditions (Félix and Wagner, 2008). For example, the heat shock protein *Hsp90* acts as a capacitor for phenotypic variation by buffering genetic variation (Queitsch, Sangster, and Susan Lindquist, 2002). This type of genetic variation that is hidden is called cryptic genetic variation (CGV), and it can be detected empirically through environmental (Genotype-by-Environment interactions) and genetic perturbations (Genotype-by-Genotype interactions) (Gibson and Dworkin, 2004). *C. elegans* has been sampled around the globe and has significant genetic diversity that can be used to study its effect on SCN. Moreover, their whole genome sequences are available from *C. elegans* Natural Diversity Resource (CeNDR) (Cook et al., 2017) to facilitate quantitative genetics approaches such as genome wide association study (GWAS) and quantitative trait loci (QTL) mapping. Combining the ability to control environment for *C. elegans* development with the ability to precisely engineer specific mutations in genes with CRISPR-Cas9 genome editing facilitates the study of genotype-by-environment ($G \times E$) and genotype-by-genotype ($G \times G$) interactions.

Seam cell lineage and seam cell number (SCN, 16 cells) at the end of postembryonic development in N2 background are mostly invariant (Sulston and Horvitz, 1977; Mestek Boukhibar and Barkoulas, 2016; Katsanos et al., 2017). It is unclear if there are genetic background specific effects on SCN in wild-type *C. elegans* and on the penetrance/expressivity of mutations affecting SCN. Therefore, we sought to investigate the effect of natural genetic variation on SCN. In order to reveal genetic background specific effects on SCN, we introduced different environmental (temperature) and genetic perturbations (mutations) in specific seam cell regulatory genes in genetically divergent wild-type backgrounds. We also investigate the phenotypic consequences of the laboratory derived allele (*n2*) versus wild-type allele (*haw6805*) of a highly pleiotropic gene *nath-10* on SCN. The work presented in this chapter is aimed at understanding the effect of natural genetic variation on SCN.

4.2 Results

4.2.1 Seam cell number is robust to standing genetic variation

Seam cell lineages have been reported to be invariant and seam cell number (SCN, 16 per lateral side of the animal) is a robust phenotype at 20°C (Sulston and Horvitz, 1977; Mestek

4.2.2 There is a $G \times E$ interaction between natural genetic variation and temperature that affects seam cell number

Despite the absence of phenotypic variation associated with genetic variation in SCN, hidden phenotypic variation could be revealed upon environmental perturbation. For example, wild-type *C. elegans* vulval development is a robust phenotype, where the spatial pattern of vulval cell fates is invariant to stochastic noise, genetic variation and environmental perturbations (Félix and Wagner, 2008; Braendle and Félix, 2008). However, different types of vulval defects were observed in different genetic backgrounds upon environmental perturbation (Braendle and Félix, 2008). Previous work from the lab showed that the robustness of SCN breaks down at 25 °C (Mestek Boukhibar and Barkoulas, 2016). In order to test if SCN changes upon environmental perturbation, I allowed *C. elegans* wild isolates to develop postembryonically at 25 °C rather than the typical 20 °C in developmental studies. Average SCN in the wild isolates increased from 16.07 – 16.21 to 16.38 – 17.44 when temperature at which development occurs was shifted from 20 °C to 25 °C (Fig. 4.2A). A two-way ANOVA was conducted to compare the main effects of temperature and genetic variation on SCN. Type of genetic variation included six levels (N2, JU2007, JU2519, JU775, CB4856, XZ1516) and type of temperature included two levels (20 °C and 25 °C). The two main effects significantly affected SCN ($F(5, 1436) = 21.073$, $p < 2.2 \times 10^{-16}$; $F(1, 1436) = 238.65$, $p < 2.2 \times 10^{-16}$). There was a significant interaction between the type of genetic variation and temperature on SCN at the 0.05 significance level ($F(5, 1436) = 30.74$, $p < 2.2 \times 10^{-16}$). Therefore, the effect of temperature on SCN was dependent on the type of genetic variation, and was suppressed in CB4856. The effect of temperature on SCN was highest in XZ1516 (Fig. 4.2B).

The position of seam cells relative to each other in the epidermis is constant after L2 symmetrical division due to the invariant seam cell division pattern in *C. elegans*. After the L4 division, there are 16 seam cells per lateral side in *C. elegans*, eight of them (H0, H1a, H1p, H2, V1a, V1p, V2a, V2p) are anterior to the vulva, two are either side of the vulva (V3a and V3p) and six of them are posterior to the vulva (V4a, V4p, V5, V6a, V6p, T) as shown in a wild-type animal at 20 °C (Fig. 1.6). Due to the consistent positioning of seam cells, it is possible to attribute increase or decrease in SCN to specific cell lineages.

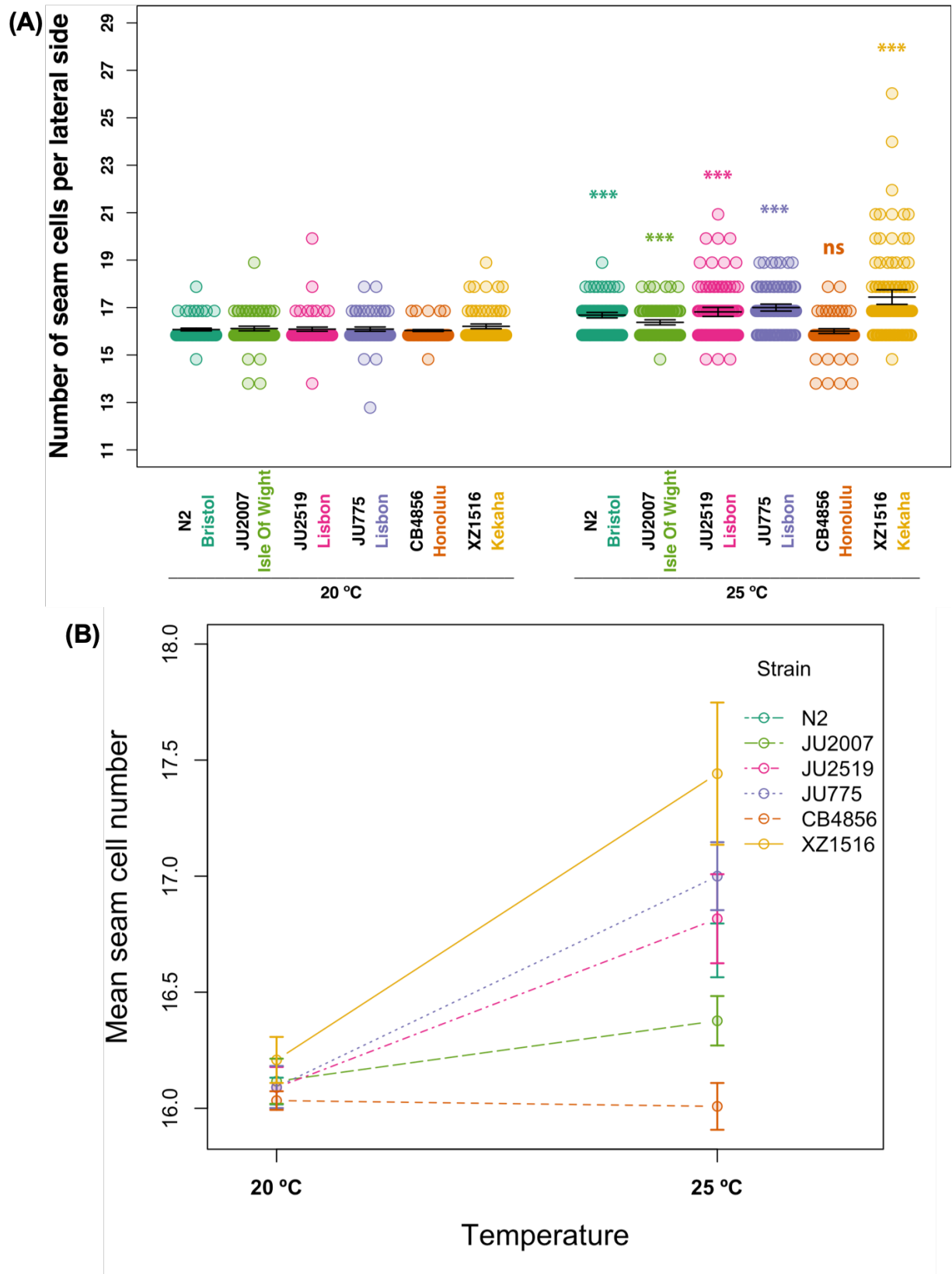


Figure 4.2: Presence of a $G \times E$ interaction between natural genetic variation and temperature in *C. elegans* that affects seam cell number. Continued on next page.

CHAPTER 4. ROLE OF NATURAL GENETIC VARIATION IN SEAM CELL DEVELOPMENT

Figure 4.2: **Presence of a $G \times E$ interaction between natural genetic variation and temperature in *C. elegans* that affects seam cell number.** (A) Error bars indicate average SCN \pm 95% confidence intervals. $120 \leq n \leq 123$ per strain. average SCN increased from 16.07 – 16.21 to 16.38 – 17.44 when temperature at which development occurs was shifted from 20 °C to 25 °C. A two-way ANOVA showed that there was a significant interaction between the type of genetic variation and temperature, $F(5, 1436) = 30.74$, $p < 2.2 \times 10^{-16}$. The main effect for genetic variation yielded a F ratio of $F(5, 1436) = 21.073$, $p < 2.2 \times 10^{-16}$. The main effect for temperature yielded a F ratio of $F(1, 1436) = 238.65$, $p < 2.2 \times 10^{-16}$. Since there was a significant interaction between the main effects of strain and temperature, SCN at different developmental temperatures was compared within each wild isolate. There was a significant increase in SCN in all wild isolates ($p < 4 \times 10^{-4}$) except CB4856 ($p = 0.65$) grown at 25 °C compared to 20 °C. *** $p < 1 \times 10^{-4}$ correspond to significant differences by one-way ANOVA. (B) Interaction plot of genetic variation and temperature on seam cell number in wild isolates of *C. elegans*. The animals were grown at 20 °C and 25 °C. Error bars indicate average SCN \pm 95% confidence intervals. $120 \leq n \leq 123$ per strain. The effect of temperature on SCN was dependent on the type of genetic variation and was completely suppressed in CB4856. The effect of temperature on SCN was highest in XZ1516.

Previous work in the laboratory led to the discovery that the increase in SCN came from specific cells (V1a, V2a, V5 and V6). In order to find out if this pattern was representative of other wild isolates, I traced the increase in seam cell number at 25 °C to specific seam cells (Fig. 4.3). Increase in SCN could not be attributed in each case due to equidistance of the extra cell to its two neighbours. In such instances, the extra cell was designated to an unknown category. I found that indeed there was an increase in the percentage of animals with extra cells in the following lineages (V1a, V2a, V5 and V6) in N2. The percentage of animals with extra an V6a was highest (17.5% – 53.33%) followed by V5 (5.74% – 12.5%), V1a (1.64% – 13.33%) and V2a (1.64% – 11.67%) compared to animals with extra cells in other seam cell lineages. However, this pattern of extra cells was suppressed in CB4856 at 25 °C. A binomial test indicated that the proportion of animals with an extra V6a cell of 0.05 was lower than the expected 0.38, $p < 2.2 \times 10^{-16}$ (two-sided). There were small differences in the percentage of animals with extra seam cells in specific lineages between the wild isolates. For example, XZ1516 had a higher percentage of animals with extra cells in seam lineages (H1a, H1p, H2, V2p and T), and JU2519 had a higher percentage of animals with extra cells in a different set of seam lineages (V2p, V3a, V3p, V4a).

CB4856 was the only strain amongst the wild isolates in which the effect of temperature on SCN was suppressed. Interestingly, I found that CB4856 strain developed slower than N2 when grown at 25 °C, and population growth was asynchronous. This may be because of the introgression of seam cell marker on chr. V. In order to confirm that the V6a suppression

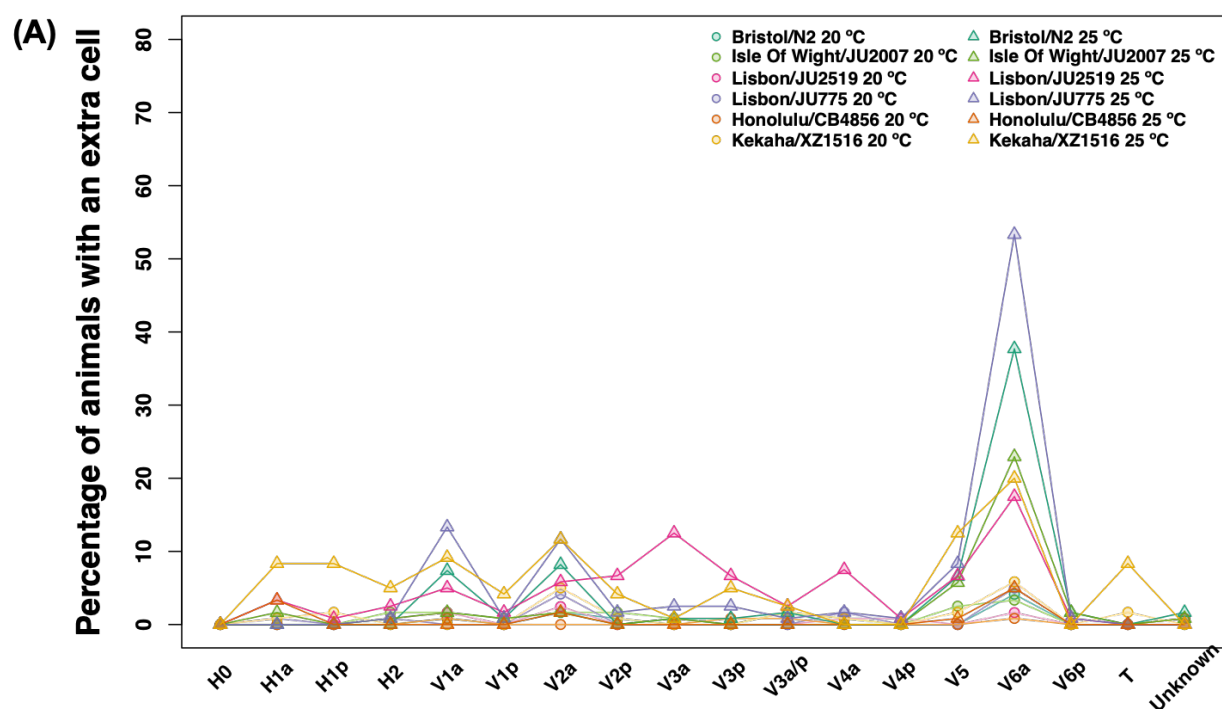


Figure 4.3: Percentage of animals with extra seam cells in wild isolates of *C. elegans* grown at 20 °C and 25 °C. The percentage of animals carrying extra seam cells in specific seam lineages was calculated. $120 \leq n \leq 123$ per strain. The percentage of animals carrying an extra V6a seam cell was the highest followed by V5 and this pattern was suppressed in CB4856 (two-sided binomial test, $p < 2.2 \times 10^{-16}$). JU2519 and XZ1516 exhibited a different pattern of extra cells in specific seam lineages compared to N2, JU2007 and JU775.

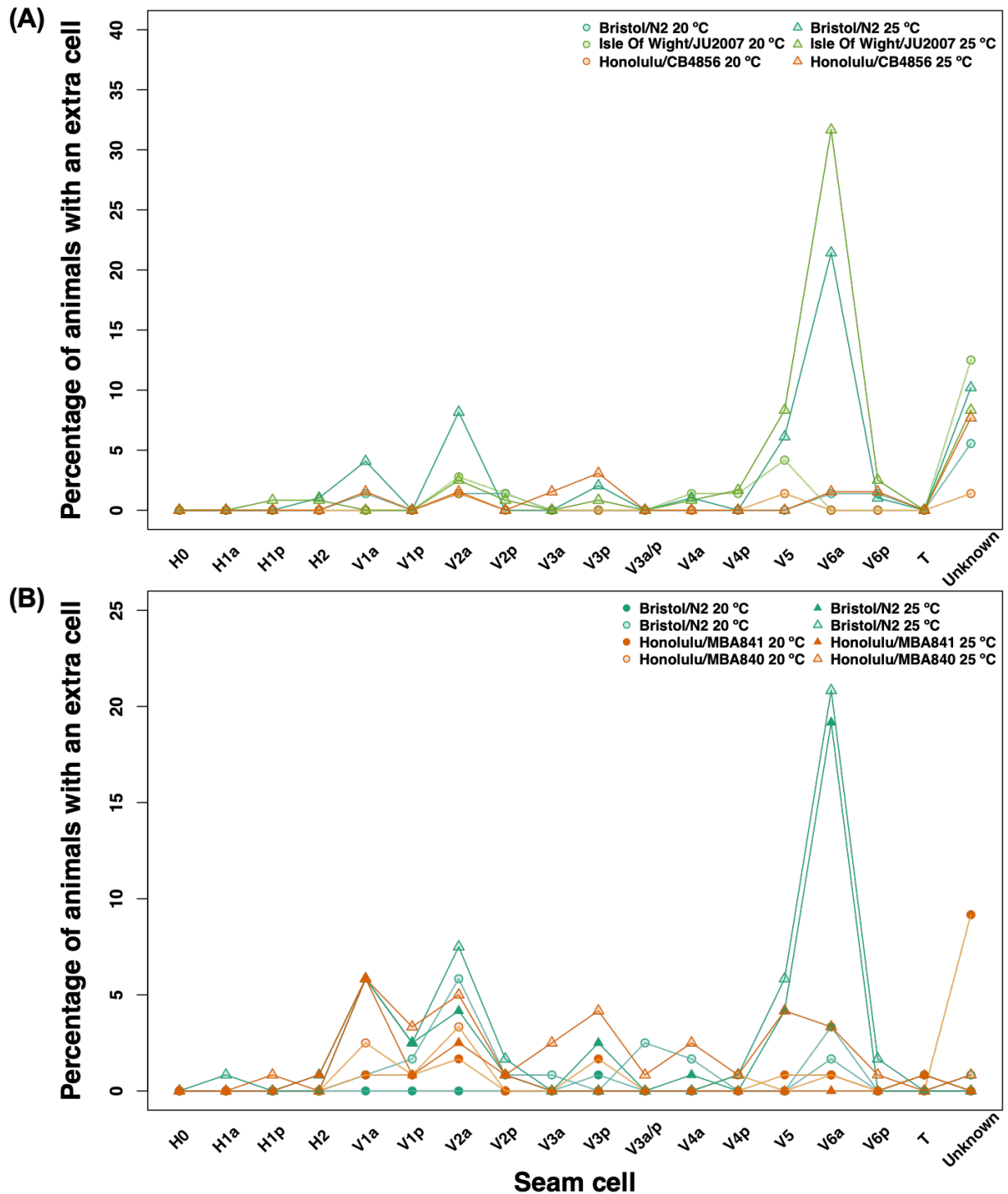


Figure 4.4: Percentage of animals with extra seam cells in a subset of wild isolates of *C. elegans* grown at 20 °C and 25 °C. (A) Extra V6a cell phenotype at 25 °C is suppressed in the CB4856 genetic background. The percentage of animals carrying extra seam cells in specific seam lineages was calculated. $65 \leq n \leq 120$ per strain. The percentage of animals carrying an extra V6a seam cell was the highest followed by V5 and this pattern was suppressed in CB4856 (two-sided binomial test, $p = 5.29 \times 10^{-6}$). (B) Suppression of extra V6a cell phenotype at 25 °C in CB4856 is not dependent on the introgressed seam marker *SCMp::GFP*. The percentage of animals carrying extra seam cells in specific seam lineages was calculated. $n = 120$ per strain. The extra V6a seam cell phenotype was suppressed in MBA840 and MBA841 (two-sided binomial test, $p = 5.388 \times 10^{-8}$ and 1.629×10^{-11}).

was not due to CB4856 animals being sick, I repeated the experiment with CB4856, N2 and JU2007 (Fig. 4.4 A). The percentage of animals with extra V6a was 21.43% – 31.67% and V5 was 6.12% – 8.33% in N2 and JU2007. The pattern of extra cells was indeed suppressed in CB4856. A binomial test indicated that the proportion of animals with an extra V6a cell of 0.02 was lower than the expected 0.21, $p = 5.29 \times 10^{-6}$ (two-sided) showing that CB4856 indeed suppressed the extra V6a cell phenotype at 25 °C.

A confounding factor was that CB4856 strain had seam cell marker *SCMp::GFP* introgressed on chromosome V, which could potentially suppress the extra V6a phenotype at 25 °C. In order to test this, two independently integrated CB4856 lines of seam cell marker *SCMp::GFP* were phenotyped at 20 °C and 25 °C (Fig. 4.4B). The percentage of animals with extra V6a was 19.17% – 20.83% and V5 was 4.17% – 5.83% in N2. MBA840 and MBA841 both suppressed the extra V6a phenotype. The data from MBA840 and MBA841 was collected in separate experiments with a N2 control on different days. A binomial test indicated that the proportion of animals with an extra V6a cell of 0.03 and 0 in MBA840 and MBA841, respectively was lower than expected 20.83 and 19.17, $p = 5.388 \times 10^{-8}$ and 1.629×10^{-11} (two-sided). This result strongly suggests that the extra V6a cell phenotype at 25 °C is not due to the introgressed marker and is specific to the CB4856 genetic background.

4.2.3 The expressivity of mutations in *lin-22* and *bro-1* did not differ between wild isolates

Another approach to reveal hidden phenotypic variation is to use genetic perturbations (Milloz et al., 2008; Duveau and Félix, 2012). Seam cell divisions are known to be affected by mutations in *bro-1*, the CBF β homologue (binding partner of Runx) and *lin-22*, a *Hes*-related bHLH transcription factor (L. A. Wrischnik and C. J. Kenyon, 1997; Kagoshima et al., 2007). SCN is reduced in *C. elegans* carrying mutations in *bro-1* gene due to loss of proliferative seam cell division in L2 stage and V and T seam lineages (Kagoshima et al., 2007). Variability in SCN is increased in animals carrying mutations in *lin-22* due to loss and gain of symmetric cell divisions (Katsanos et al., 2017). In order to reveal hidden phenotypic variation, I used the CRISPR-Cas9 genome editing tool to create de novo precise mutations in *bro-1* and *lin-22* in wild isolates (N2, JU2007 and CB4856) using the same CRISPR single guide RNA (sgRNA).

A previous attempt to mutate *bro-1* by CRISPR-Cas9 genome editing was successful and *icb44* allele, a 8 bp deletion in N2 was produced. I attempted to produce the same allele in other backgrounds using the same sgRNA, however, I produced a new allele *icb45*, which is a 9 bp in-frame deletion allele in JU2007 and CB4856 backgrounds but not in the N2 background. Another attempt at mutating *bro-1* yielded two different alleles *icb46* (an indel in N2 background) and *icb47* (a 9 bp in-frame deletion in JU2007). I counted SCN in these mutants grown at 20 °C (Fig. 4.5A).

As expected in *bro-1* mutants, the average SCN decreased from 16 in wild-type *C. elegans* to 11.92 – 13.13. One-way ANOVA showed that SCN was significantly affected by the strain ($F(5, 325) = 23.85$, $p < 2.2 \times 10^{-16}$). Post hoc Tukey HSD tests showed that SCN in *bro-1* mutant strains is significantly different from wild-type ($p < 2.2 \times 10^{-16}$). Interestingly, there was a significant difference in SCN between MBA243 and two strains (MBA173, MBA272) ($p = 0.0038$ and 0.0021). The average SCN in MBA243 ($\mu = 11.92$) carrying an in-frame deletion was lower than the average SCN of MBA173 ($\mu = 13.08$), which carries a deletion that causes a pre-mature stop codon. There was no statistically significant difference in SCN between the strains MBA274 (CB4856) and MBA243 (JU2007) carrying the same deletion allele *icb45* in *bro-1* ($p = 0.41$). Since there was no difference in the expressivity of *icb45* in two different genetic backgrounds, I decided to target other genes affecting seam cell number like *rnt-1*, *nhr-25* and *egl-18* with CRISPR. However, my attempts at gene editing were not successful.

Previous studies had shown that *lin-22* mutations increase the number of PDE due to homeotic transformation of V1 – V4 to V5 (L. A. Wrischnik and C. J. Kenyon, 1997; Katsanos et al., 2017). I used this phenotype to find animals with more than one *dat-1p::GFP* neuron per lateral side to screen for successful gene-editing events in *lin-22*. I recovered three new mutant alleles affecting the first exon of *lin-22*, *icb49* (5 bp deletion), *icb50* (1 bp deletion) and *icb52* (9 bp in-frame deletion). *icb49* was the only allele that was recovered in both N2 and CB4856 background.

SCN was counted in order to compare expressivity of *lin-22* mutations (Fig. 4.5B). The average SCN increased slightly from the expected 16 to 16.94 – 18.32 in the *lin-22* mutants.

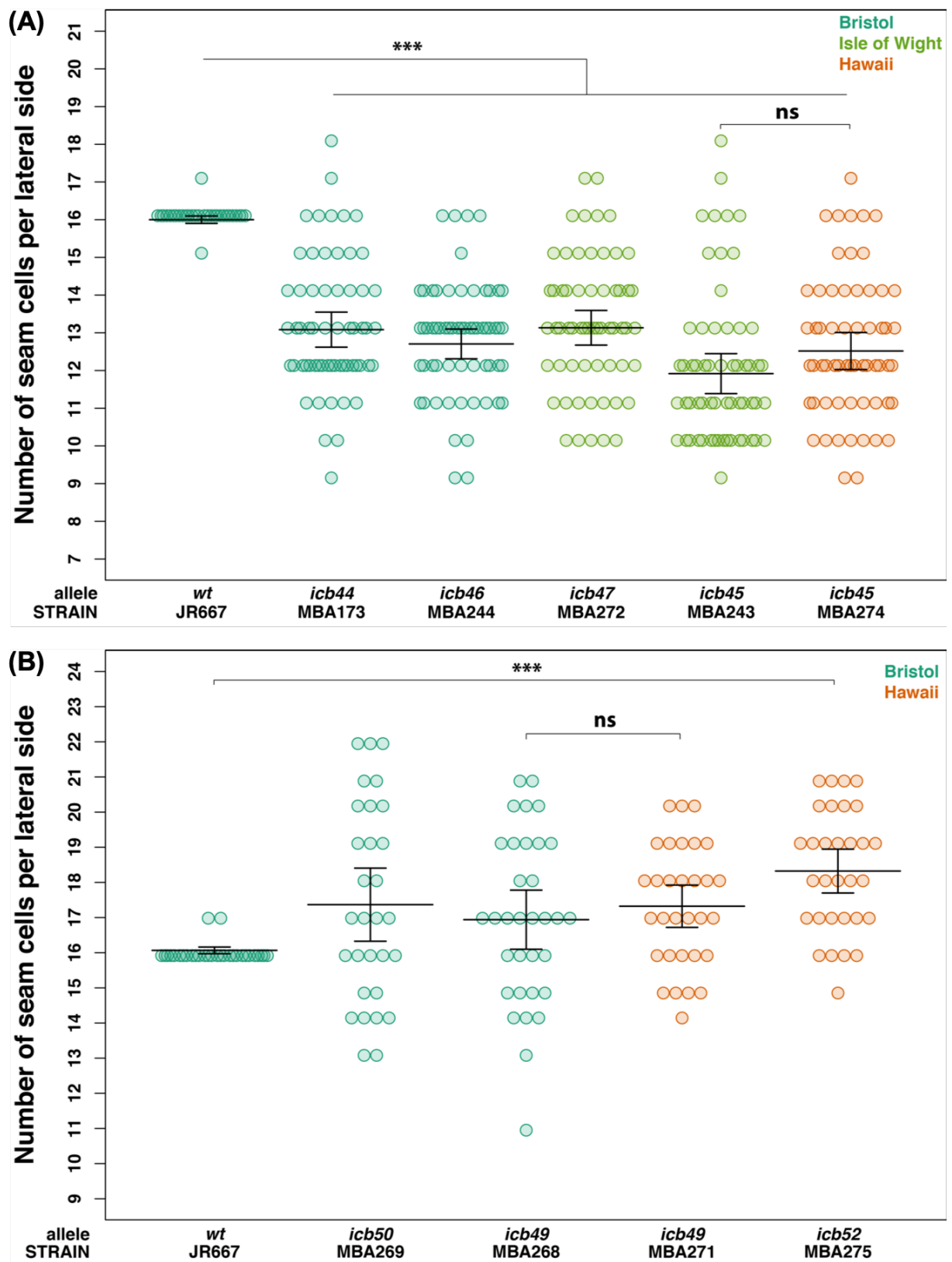


Figure 4.5: No difference in the expressivity of *bro-1* or *lin-22* mutations between wild isolates. Continued on next page.

Figure 4.5: **No difference in the expressivity of *bro-1* or *lin-22* mutations between wild isolates.** SCN in CRISPR mutants in different genetic backgrounds. (A) One-way ANOVA showed that SCN was effected by the strain ($F(5, 325) = 23.851, p < 2.2 \times 10^{-16}$). Post hoc Tukey HSD test showed that SCN in *bro-1* mutant strains is significantly different from wild-type ($p < 2.2 \times 10^{-16}$). However, there was no statistically significant difference in SCN between the strains MBA274 (CB4856) and MBA243 (JU2007) carrying the same deletion allele *icb45* in *bro-1* was not significantly different from each other ($p = 0.41$). WT ($n = 30$), MBA173 ($n = 61$), MBA244, MBA272, MBA243 and MBA274 ($n = 60$). (B) One-way ANOVA showed that SCN was affected by strain ($F(5, 183) = 23.953, p = 5 \times 10^{-4}$). Post hoc Tukey HSD test showed that SCN in MBA275 (CB4856) carrying *lin-22(icb52)* was significantly different from wild-type ($p = 1 \times 10^{-4}$). There were no statistically significant differences in SCN between the two *lin-22* mutants in N2 background ($p = 0.91$), and the two *lin-22* mutants in CB4856 background ($p = 0.27$). There was no significant difference in SCN between the strains MBA268 (N2) and MBA271 (CB4856) carrying the same putative null allele *icb49* in *lin-22* ($p = 0.94$). $30 \leq n \leq 33$ per strain. The data pertaining to *icb49* and *icb50* was published in Katsanos et al. (2017). In both A and B, error bars indicate average SCN \pm 95 % confidence intervals. *** $p < 1 \times 10^{-4}$ correspond to significant differences by post hoc Tukey HSD test.

One-way ANOVA showed that SCN was affected by strain ($F(5, 183) = 23.953, p = 5 \times 10^{-4}$). Post hoc Tukey HSD test showed that SCN in MBA275 (CB4856) carrying *lin-22(icb52)* was significantly different from wild-type ($p = 1 \times 10^{-4}$). There were no statistically significant differences in SCN between the two *lin-22* mutants in N2 background ($p = 0.91$), and the two *lin-22* mutants in CB4856 background ($p = 0.27$). There was no statistically significant difference in SCN between the strains MBA268 (N2) and MBA271 (CB4856) carrying the same putative null allele *icb49* of *lin-22* ($p = 0.94$) suggesting that there are no differences in the expressivity of *lin-22* mutations in different genetic backgrounds.

4.2.4 The expressivity of *eff-1(icb4)* did not differ between wild isolates

eff-1 is a fusogen required for the hyp7 fated seam daughters to fuse to the hyp7 syncytium (Mohler et al., 2002). Mutations in *eff-1* result in failure of all cell fusion events during embryonic and postembryonic development leading to a fragmented hypodermis and disorganised seam. In Chapter 3, I showed that mutations in *eff-1* increase SCN by causing disorganisation of the seam cells (section 3.2.2). CRISPR-Cas9 genome editing approach did not work to introduce the mutations in *eff-1* in different genetic backgrounds. Another approach to studying the expressivity of mutations in different backgrounds is to introgress them from one background to another. A putative null mutation *eff-1(icb4)* was introgressed into JU2007 from N2 background and SCN was counted. The average SCN increased slightly from the expected

16 to 18.15 – 18.32 in the *eff-1(icb4)* mutants (Fig. 4.6A). One-way ANOVA showed that SCN was significantly affected by strain ($F(2, 116) = 12.82$, $p = 9.36 \times 10^{-6}$). Post hoc Tukey HSD test showed that SCN in *eff-1(icb4)* mutants was significantly different from wild-type ($p = 3.44 \times 10^{-5}$ and 1×10^{-4}). However, there was no statistically significant difference in SCN between the strains MBA79 (N2) and MBA368 (JU2007) carrying the same putative null allele *icb4* in *eff-1* ($p = 0.93$) suggesting that there are no differences in the expressivity of *eff-1* mutations in different genetic backgrounds.

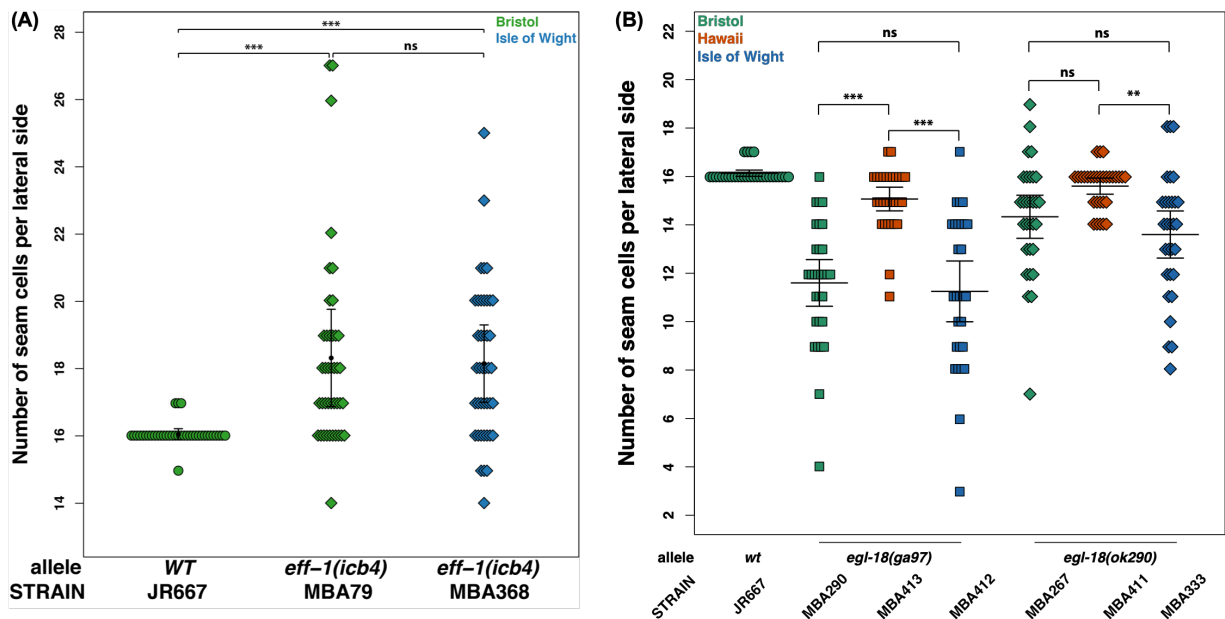


Figure 4.6: **Expressivity of *eff-1(icb4)* and *egl-18(ga97)* mutation in SCN between wild isolates.** SCN in CRISPR mutants in different genetic backgrounds. (A) No difference in the expressivity of *eff-1(icb4)* mutation in wild isolates. One-way ANOVA showed that SCN was affected by strain ($F(2, 116) = 12.82$, $p = 9.36 \times 10^{-6}$). Post hoc Tukey HSD test showed that SCN in *eff-1(icb4)* mutants was significantly different from wild-type ($p = 3.44 \times 10^{-5}$ and 1×10^{-4}). However, There was no statistically significant difference in SCN between the strains MBA79 (N2) and MBA368 (CB4856) carrying the same putative null allele *icb4* in *eff-1* ($p = 0.93$). $37 \leq n \leq 41$ per strain. (B) Significant difference in the expressivity of *egl-18* alleles in SCN between wild isolates. One-way ANOVA showed that SCN was affected by strain carrying *ga97* allele ($F(2, 85) = 21.25$, $p = 3.29 \times 10^{-8}$) or *ok290* allele ($F(2, 87) = 6.94$, $p = 0.0016$). Post hoc Tukey HSD showed that there were no statistically significant difference in SCN between N2 and JU2007 carrying *ga97* ($p = 0.85$) or *ok290* ($p = 0.37$). However, there was a significant difference in SCN between CB4856 and N2 ($p = 1.8 \times 10^{-6}$) or JU2007 ($p = 3 \times 10^{-7}$) carrying *ga97* mutation. There was no statistically significant difference in SCN between CB4856 and N2 carrying the weaker loss-of-function *ok290* mutation ($p = 0.06$), but there was a significant difference in SCN between CB4856 and JU2007 carrying *ok290* mutation ($p = 0.0012$). $37 \leq n \leq 41$ per strain. In both A and B, error bars indicate average SCN \pm 95% confidence intervals. *** $p < 1 \times 10^{-4}$ correspond to significant differences by post hoc Tukey HSD test.

4.2.5 The expressivity of mutations in GATA transcription factor *egl-18* was lower in CB4856

egl-18 is a GATA transcription factor required for seam cell fate and inhibiting fusion of seam cells to the hyp7 syncytium (Koh and Rothman, 2001). *egl-18* is a target of the Wnt/ β -catenin asymmetric (W β A) pathway and specifies seam cell fate during larval asymmetric seam cell divisions (Gorrepati, K. W. Thompson, and Eisenmann, 2013). *egl-18* mutations have been previously reported to decrease SCN (Koh and Rothman, 2001; Gorrepati and Eisenmann, 2015).

Since CRISPR-Cas9 genome editing approach did not work to introduce mutations in *egl-18* in different genetic backgrounds, I introgressed two different alleles, a null allele *ga97* and an in-frame 698 bp deletion allele *ok290* into CB4856 and JU2007 from N2. SCN was counted for strains carrying the *ga97* and *ok290* alleles by blinding the strain name. The average SCN decreased from 16.13 in wild-type to 11.25 – 15.60 in the *egl-18* mutants (Fig. 4.6B). One-way ANOVA showed that SCN was affected by strain carrying *ga97* ($F(2, 85) = 21.25$, $p = 3.29 \times 10^{-8}$) or *ok290* allele ($F(2, 87) = 6.94$, $p = 0.0016$). Post hoc Tukey HSD showed that there were no statistically significant difference in SCN between N2 and JU2007 carrying *ga97* ($p = 0.85$) or *ok290* ($p = 0.37$). However, there was a significant difference in SCN between CB4856 and N2 ($p = 1.8 \times 10^{-6}$) or JU2007 ($p = 3 \times 10^{-7}$) carrying *ga97* mutation. There was no statistically significant difference in SCN between CB4856 and N2 carrying the weaker loss-of-function *ok290* mutation ($p = 0.06$), but there was a significant difference in SCN between CB4856 and JU2007 carrying *ok290* mutation ($p = 0.0012$).

To test if the difference in expressivity of *ga97* between CB4856, N2 and JU2007 was recapitulated by knockdown of *egl-18* with RNAi, I subjected the strains to postembryonic RNAi knockdown of *egl-18* in wild-type animals (Fig. 4.7A). SCN was significantly different between JU2007 and the two strains N2 and CB4856 upon *egl-18* knockdown ($p_{JU2007 \text{ vs. } N2} = 0.0054$ and $p_{JU2007 \text{ vs. } CB4856} = 0.023$). N2 and CB4856 did not have statistically different SCN upon *egl-18* knockdown ($p = 0.87$) in contrast to them carrying *egl-18(ga97)* mutation. Thus, the differential expressivity of loss of *egl-18* seems to be observed only in N2 and CB4856 carrying the *ga97* mutation. In other words, there is a $G \times G$ interaction suggesting that there is cryptic genetic variation affecting SCN.

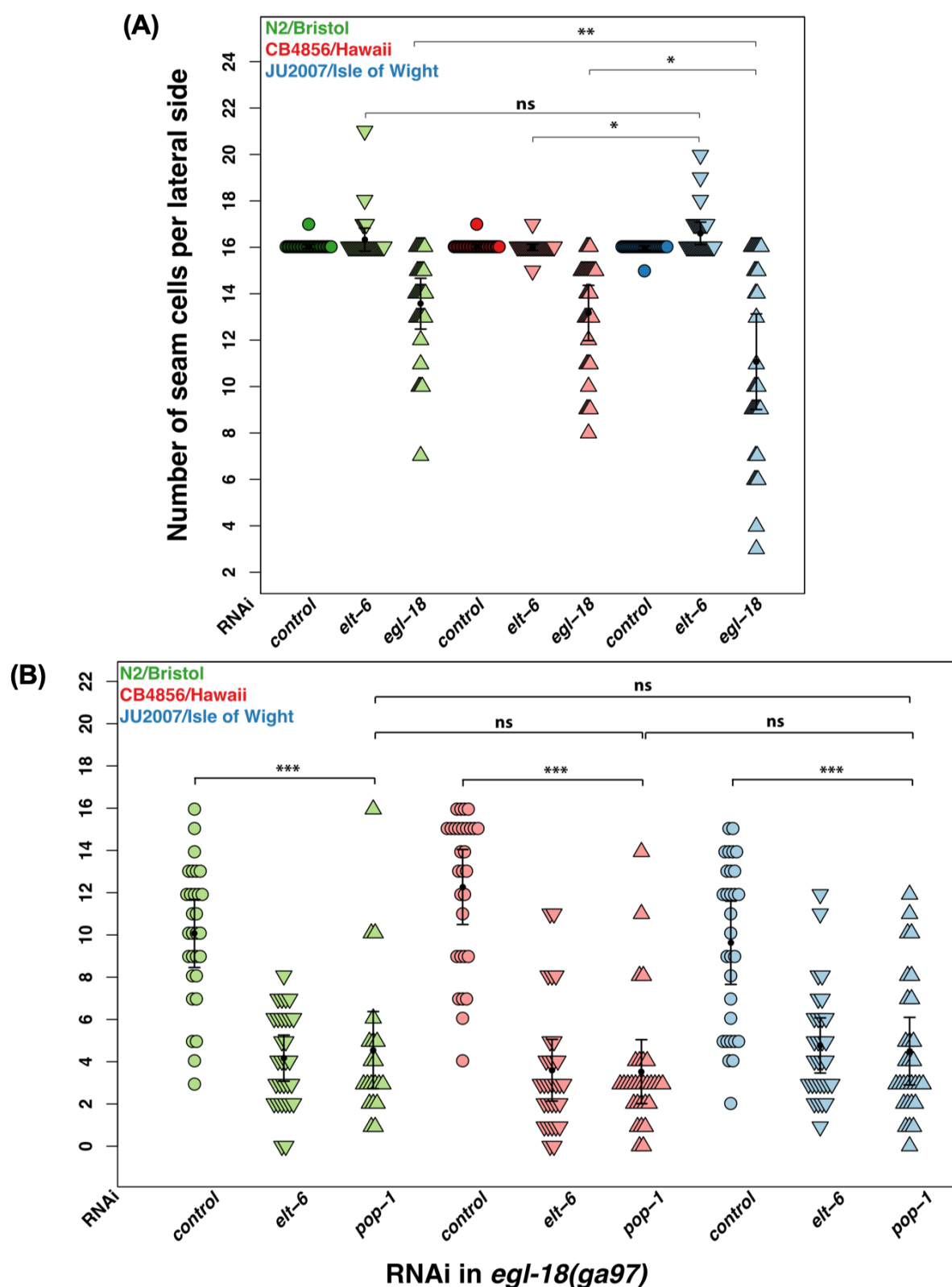


Figure 4.7: Differential expressivity of *egl-18(ga97)* is not recapitulated by *egl-18* RNAi. Continued on text page.

Figure 4.7: **Differential expressivity of *egl-18(ga97)* is not recapitulated by *egl-18* RNAi.** (A) SCN in animals upon *egl-18* RNAi. One-way ANOVA showed that SCN was affected by strain for *egl-18* RNAi ($F(2, 85) = 5.92$, $p = 0.0039$). Post hoc Tukey HSD showed that SCN was significantly different between JU2007 and the two strains N2 and CB4856 upon *egl-18* knockdown (0.0054 and 0.023). There was no statistically significant difference in SCN between N2 and CB4856 upon *egl-18* knockdown ($p = 0.87$). (B) Knockdown of *elt-6* or *pop-1* by RNAi in *egl-18(ga97)* animals abrogates the differential expressivity of SCN between wild isolates. One-way ANOVA showed that SCN was not affected by strain upon *elt-6* RNAi ($F(2, 85) = 0.07$, $p = 0.93$) or *pop-1* RNAi ($F(2, 85) = 0.54$, $p = 0.59$). In both A and B, error bars indicate average SCN \pm 95% confidence intervals. $n = 30$ per strain. *** $p < 1 \times 10^{-4}$ correspond to significant differences by post hoc Tukey HSD test.

4.2.6 The differential expressivity of *egl-18(ga97)* between wild isolates requires *elt-6* and *pop-1*

The number of seam cells in *egl-18(ga97)* is not zero and this is because of its paralogue *elt-6* functioning redundantly with *egl-18* to promote seam cell fate. Loss of *elt-6* alone does not cause postembryonic defects. However, loss of function of *elt-6* in a *egl-18* loss-of-function background during postembryonic development leads to a severe reduction in seam cells compared to loss of function of *egl-18* alone. In order to test if the difference in SCN between the wild isolates carrying *egl-18(ga97)* required *elt-6* function, I used RNAi to knockdown *elt-6* in wild isolates carrying *egl-18(ga97)* mutation and counted SCN (Fig. 4.7B). One-way ANOVA showed that SCN was not affected by strain upon *elt-6* RNAi ($F(2, 85) = 0.07$, $p = 0.93$), suggesting that differential expressivity of *egl-18(ga97)* requires *elt-6*.

Since *egl-18* and *elt-6* are downstream of Wnt/ β -catenin asymmetric ($W\beta A$), we wanted to determine if the difference in SCN between the wild isolates carrying *egl-18(ga97)* was dependent on the TCF homolog *pop-1*, which activates the targets of Wnt signalling pathway. To address this, I subjected animals to *pop-1* RNAi knockdown (Fig. 4.7B). One-way ANOVA showed that SCN was not affected by strain upon *elt-6* RNAi ($F(2, 85) = 0.54$, $p = 0.59$), suggesting that the differential expressivity of *egl-18(ga97)* requires *pop-1*.

4.2.7 The differential expressivity of *egl-18(ga97)* is postembryonic

egl-18 is required for seam cell fate in embryonic development (Koh and Rothman, 2001) and again postembryonically for larval seam cell divisions (Gorrepati, K. W. Thompson, and Eisenmann, 2013). We decided to narrow our focus on the genetic basis for the difference in *ga97* allele expressivity between N2 and CB4856 background. To find if the developmental basis for the dif-

CHAPTER 4. ROLE OF NATURAL GENETIC VARIATION IN SEAM CELL DEVELOPMENT

ference in expressivity of *egl-18* between N2 and CB4856 is embryonic or postembryonic, I scored SCN after L1 asymmetric division and at L4 stage after all seam cell divisions were completed (Fig. 4.8A). One-way ANOVA on L1 SCN data showed that there was no statistically significant difference in SCN between N2 and CB4856 after L1 division ($F(1, 122) = 0.95, p = 0.33$). However, one-way ANOVA on L4 SCN data showed that there was a significant difference in SCN between N2 and CB4856 after L4 division ($F(1, 125) = 44.91, p = 6.31 \times 10^{-10}$) suggesting that the difference in the expressivity of *egl-18(ga97)* between N2 and CB4856 is postembryonic.

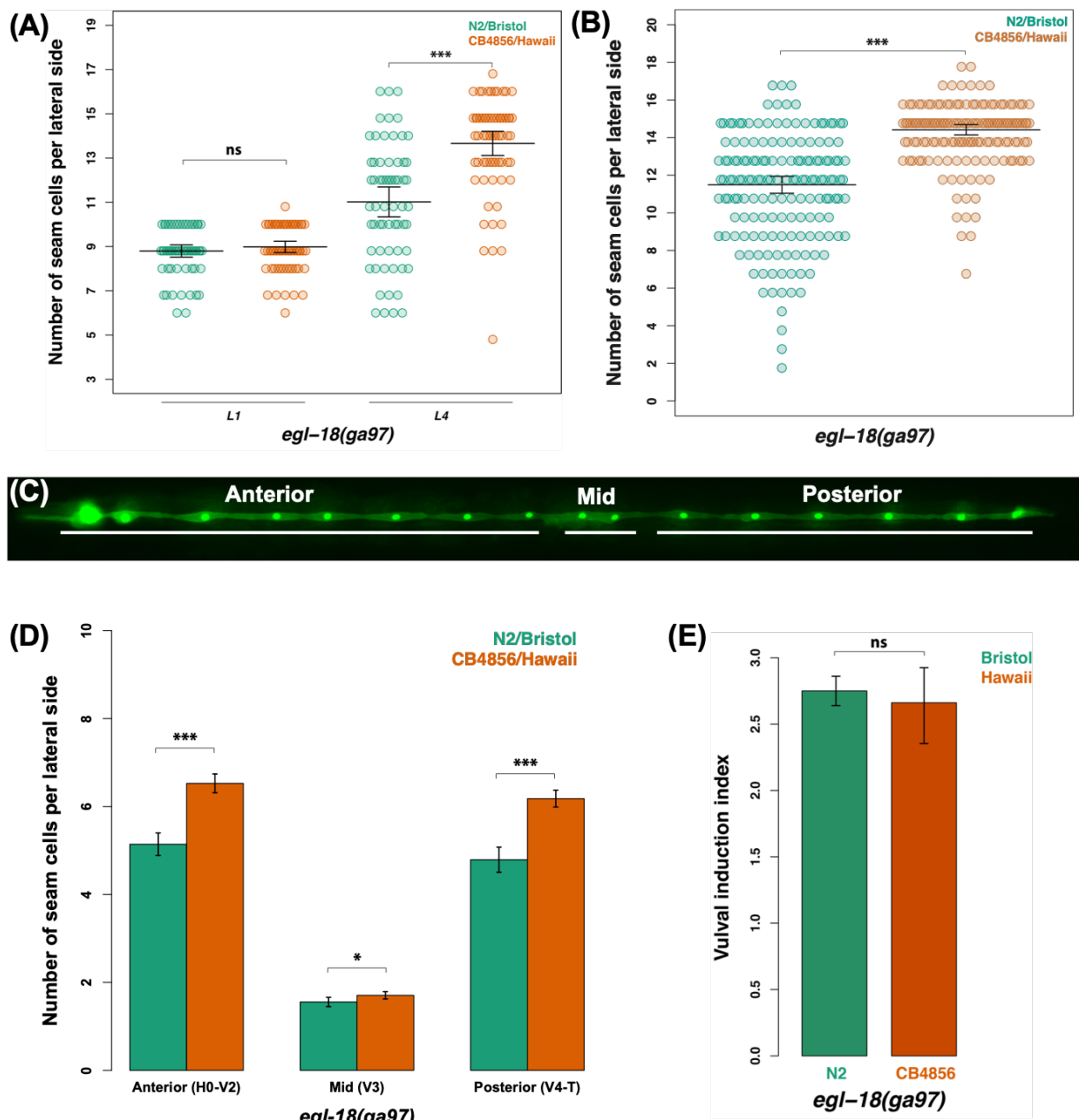


Figure 4.8: The differential expressivity of *egl-18(ga97)* mutation in N2 and CB4856 is postembryonic and specific to seam cell number. Continued on next page.

CHAPTER 4. ROLE OF NATURAL GENETIC VARIATION IN SEAM CELL DEVELOPMENT

Figure 4.8: **The differential expressivity of *egl-18(ga97)* mutation in N2 and CB4856 is postembryonic and specific to seam cell number.** (A) The differential expressivity of *egl-18(ga97)* mutation in N2 and CB4856 is postembryonic and specific to seam cell number. There are 10 and 16 SCN in wild-type L1 and L4 stage larvae, respectively. One-way ANOVA on L1 SCN data showed that there was no statistically significant difference in SCN between N2 and CB4856 after L1 division ($F(1, 122) = 0.95$, $p = 0.33$). However, one-way ANOVA on SCN in L4 animals showed that there was a significant difference in SCN between N2 and CB4856 after L4 division ($F(1, 125) = 44.91$, $p = 6.31 \times 10^{-10}$). SCN each strain was scored on two different trials and the counts were plotted together. Data for L1 and L4 were acquired on separate days but were plotted together for visualisation. $60 \leq n \leq 64$ per strain. (B) Seam cell number in N2 and CB4856 carrying *egl-18(ga97)* mutation in large sample size. One-way ANOVA showed that there is a significant difference in SCN between the two strains ($F(1, 310) = 115.02$, $p < 2.2 \times 10^{-16}$). $150 \leq n \leq 162$ per strain. (C) Fluorescent image of an L4 stage larvae showing the expression of *wIs51[SCMp::GFP]* positive seam cells. The seam cells have been counted in panel B in three categories - anterior (H0, H1a, H1p, V1a, V1p, V2a, V2p), mid (V3a, V3p) and posterior (V4a, V4p, V5, V6a, V6p, T) as shown in the image. (D) Seam cells in B are differentiated in three categories (anterior, mid, posterior) along the anteroposterior axis of *C. elegans* as shown in C. One-way ANOVA yields that there are differences in SCN in the three categories - anterior, mid and posterior ($F(1, 310) = 66.41, 4.87, \text{ and } 61.78$, $p < 0.03$). (E) A Student's t-test showed that there was no difference in the vulval induction index between the two strains N2 and CB4856 carrying *egl-18(ga97)* mutation. $31 \leq n \leq 36$ per strain. In A, B, D and E, error bars indicate 95% confidence intervals. *** $p < 1 \times 10^{-4}$ correspond to significant differences.

Loss of *egl-18(ga97)* leads to loss of seam cell fate and reduction in SCN in adult hermaphrodite (Gorrepati and Eisenmann, 2015). However, it is not clear which seam cells are lost upon the loss of *egl-18*. As mentioned in section 4.2.2 the positioning of seam cells is stereotypical. Using this observation, I scored SCN by placing them into three different categories - anterior (H0, H1a, H1p, V1a, V1p, V2a, V2p), mid (V3a, V3p) or posterior (V4a, V4p, V5, V6a, V6p, T) as shown in Fig. 4.8C. In wild-type *C. elegans*, there are 8, 2 and 6 seam cells in the anterior, mid and posterior category. The average SCN in the three categories were 5.14, 1.56 and 4.79 in *egl-18(ga97)* in N2 background. The average SCN in the three categories were 6.53, 1.71 and 6.18 in *egl-18(ga97)* in CB4856 (Fig. 4.8D). There seems to be a greater loss of seam cells in the anterior part of the animals in *egl-18(ga97)* mutants suggesting that these cells are more sensitive to the loss of *egl-18*. One-way ANOVA showed that SCN in all three categories was significantly different between N2 and CB4856 ($F(1, 310) = 66.41, 4.87 \text{ and } 61.78$, $p < 0.03$). Interestingly, there seems to be a greater loss of seam cells in *egl-18(ga97)* in N2 background compared to CB4856 in all three categories. As expected there is a significant difference in the total SCN in *egl-18(ga97)* between the two isolates with a one-way ANOVA ($F(1, 310) = 115.02$, $p < 2.2 \times 10^{-16}$) (Fig. 4.8B.)

4.2.8 The differential expressivity of *egl-18(ga97)* phenotype is specific to the seam tissue

N2 animals carrying *egl-18(ga97)* appeared severely egg laying defective and as a result moved less than CB4856 animals carrying *egl-18(ga97)* (data not shown). *egl-18* is also required for ensuring vulval cell fate and inhibiting fusion of the vulval precursor cells with hyp7 syncytium (Eisenmann and S. K. Kim, 2000; Koh, Peyrot, et al., 2002). *egl-18* mutations cause vulval abnormalities like protruding vulva (pvl) and egg-laying defective (egl) phenotypes (Trent, Tsuing, and Horvitz, 1983; Eisenmann and S. K. Kim, 2000). Since *egl-18* is involved in vulval development, I was interested to see if there was a difference in the vulval induction index between N2 and CB4856 carrying the *ga97* mutation. The vulva induction index decreased from 3 in wild-type to 2.66 – 2.75 in N2 and CB4856 strains carrying *ga97* mutation (Fig. 4.8E). However, there was no statistically significant difference in the vulval induction index between N2 and CB4856 strains carrying *ga97* mutation (Student’s t-test, $p = 0.56$) suggesting that the difference in expressivity of *ga97* was specific to SCN.

4.2.9 The differential expressivity of *egl-18(ga97)* seam phenotype is not dependent on *haw6805* polymorphism in *nath-10*

nath-10 is an essential RNA cytidine acetyltransferase gene and an ortholog of human NAT10 (N-acetyl transferase). A study by Duveau and Félix (2012) found that a non-synonymous polymorphism *haw6805* in the *nath-10* gene affects the expressivity of mutation in the EGF-receptor-family tyrosine kinase gene *let-23* in different *C. elegans* wild-type backgrounds. The polymorphism was found to affect the expressivity of other mutations in the EGF/Ras pathway. A gain of function mutant of Ras, *let-60* was found to be affected in a similar way. Previous work in the lab showed that *nath-10* RNAi increases SCN variability. Since *egl-18* also acts in the vulval fate specification I wanted to test if the differential expressivity of *egl-18(ga97)* between N2 (Bristol) and CB4856 (Hawaii) was also *nath-10(haw6805)* dependent. To this end, I introduced the *haw6805* polymorphism present in CB4856 into N2 background carrying *egl-18(ga97)* and counted SCN in this strain (Fig. 4.9A). One-way ANOVA showed that there was a significant effect of strain on SCN ($F(4, 145) = 77.19$, $p < 2.2 \times 10^{-16}$). Post hoc Tukey HSD

showed that there was a significant difference between *nath-10(haw6805); egl-18(ga97)* in N2 and CB4856 ($p < 2.2 \times 10^{-16}$). I found that there was no difference between *nath-10(haw6805); egl-18(ga97)* and *nath-10(n2); egl-18(ga97)* ($p = 0.98$). Therefore, *haw6805* polymorphism in *nath-10* in CB4856 was not responsible for the differential expressivity of *egl-18(ga97)* in N2 and CB4856.

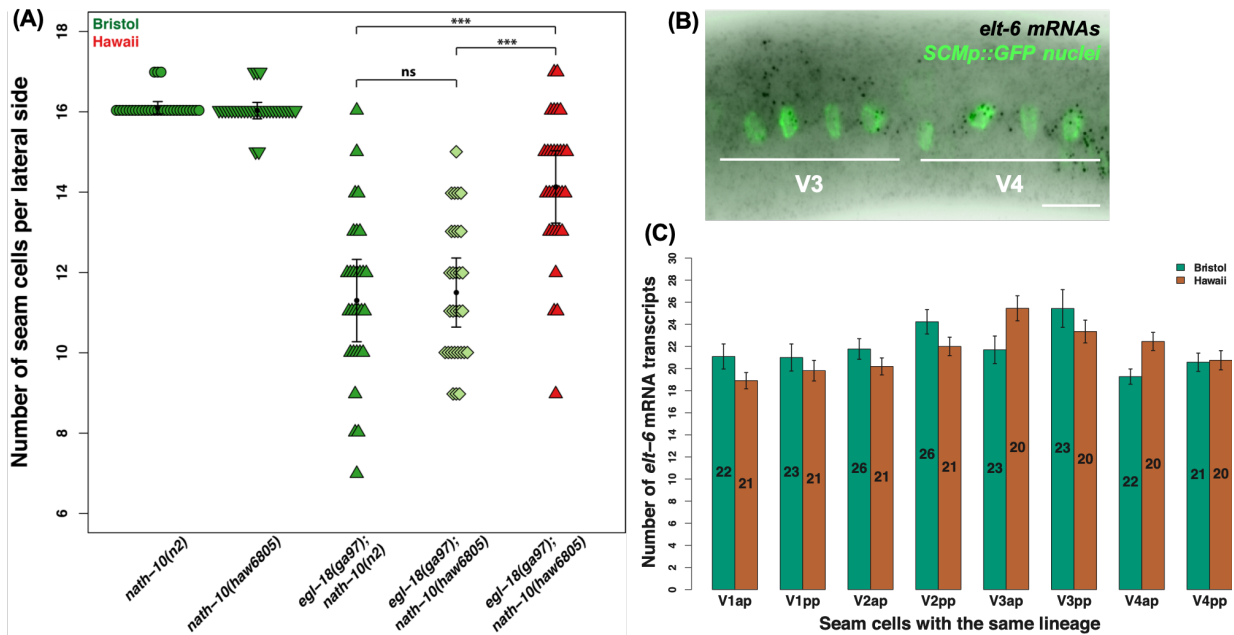


Figure 4.9: **The differential expressivity of *egl-18(ga97)* seam phenotype is not due to *nath(haw6805)* or *elt-6* expression differences.** (A) The differential expressivity of *egl-18(ga97)* seam phenotype is not dependent on *haw6805* polymorphism in *nath-10*. One-way ANOVA showed that there was a significant affect of strain on SCN ($F(4, 145) = 77.19$, $p < 2.2 \times 10^{-16}$). Post hoc Tukey HSD showed revealed that there was a significant difference between *nath-10(haw6805); egl-18(ga97)* in N2 (Bristol) and CB4856 (Hawaii) ($p < 2.2 \times 10^{-16}$). There was no difference between *nath-10(haw6805); egl-18(ga97)* and *nath-10(n2); egl-18(ga97)* ($p = 0.98$). $n = 30$ per strain. (B) Representative smFISH image showing *elt-6* expression in posterior seam cell daughters. Seam cells are labelled in green due to *SCMp::GFP* expression and black spots correspond to mRNAs in wild-type V cells after the L2 asymmetric division. Scale bar is $20 \mu\text{m}$. (C) No differences in *elt-6* expression between N2 and CB4856 strains carrying the *egl-18(ga97)* mutation. n per cell are presented inside the bars. One-way ANOVA showed that there were no statistically significant differences in the mRNA counts between N2 and CB4856 *egl-18(ga97)* mutants ($F(1, 348) = 0.13$, $p = 0.72$). Error bars indicate average SCN $\pm 95\%$ confidence intervals.

4.2.10 The differential expressivity of *egl-18(ga97)* phenotype is not dependent on increase in expression of its paralogue *elt-6*

One mechanism of robustness is redundancy in the form of gene duplicates in a genetic pathway, wherein one gene duplicate compensates for the loss of the other (Masel and Siegal, 2009; Félix

and Barkoulas, 2015). Therefore, it is conceivable that *elt-6* in CB4856 compensates for the loss of *egl-18*. I wanted to test if the underlying basis for the differential expressivity of *egl-18(ga97)* is due to difference in expression of its paralogue *elt-6*. If *elt-6* compensates for the loss of *egl-18*, there should be a difference in *elt-6* expression between N2 and CB4856 *egl-18(ga97)* mutants. To test this, I performed single molecule fluorescent in situ hybridisation (smFISH) to measure *elt-6* expression in *egl-18(ga97)* mutants in N2 and CB4856 backgrounds (Fig. 4.9B). I found with one-way ANOVA that there were no statistically significant differences in the mRNA counts between N2 and CB4856 *egl-18(ga97)* mutants ($F(1, 348) = 0.13$, $p = 0.72$).

4.3 Discussion

4.3.1 Effect of natural genetic variation on seam cell number

We have shown by introgressing seam cell marker *wIs51[SCMp::GFP + unc-119(+)]* into *C. elegans* wild isolates that seam cell number is robust to natural genetic variation. However, there seems to be a $G \times E$ interaction between natural genetic variation and temperature that affects seam cell number. Increase of developmental temperature from 20°C to 25°C increases SCN in all isolates except in CB4856. The extent of increase in SCN was dependent on the wild-isolate. XZ1516 was most responsive to temperature in terms of SCN showing an increase in SCN even at 20°C which is standard *C. elegans* development temperature. Its preferred developmental temperature is 15°C – 18°C. The effect of temperature on SCN was suppressed in CB4856. CB4856 and XZ1516 are from two different islands on Hawaii (Honolulu and Kekaha, respectively), and these isolates represent genetically divergent strains from N2 phylogenetically (Fig. 1.3 and Fig. 1.3 Cook et al. (2017), <http://www.elegansvariation.org/>). It is intriguing that temperature has opposing effects on SCN on these two isolates from Hawaii. However, this could be due to high genetic diversity amongst the isolates sampled from various locations on the Pacific Rim (Andersen, Gerke, et al., 2012; Cook et al., 2017). Since, the effect of temperature on SCN was variable in different isolates, there may be natural variation affected SCN. 330 wild isolates are available from CeNDR (<http://www.elegansvariation.org/>) that could be phenotyped for SCN and the underlying causative genetic variation could be mapped with GWAS.

Previous data in the lab showed that the increase in SCN upon increase in developmental temperature from 20 °C to 25 °C was driven by increase in the number of specific seam lineages, V6a, V5, V1a and V2a. All isolates show higher relative frequencies in these cells except CB4856. XZ1516 and JU2519 are also notable in that they show increase in number of a few other cells like H1a, H1p and T in the case of XZ1516 and V2p, V3a, V3p and V4a in the case of JU2519. The increase in SCN at 25 °C was mostly driven by an extra cell in the V6a lineage. The most notable exception is CB4856 which completely suppresses this phenotype. I confirmed that this phenotype is not driven by the introgressed marker *wIs51[SCMp::GFP]* by using independently integrating seam cell marker into strains MBA840 and MBA841 (CB4856 background). MBA840 and MBA841 carry transgenes *icbIs16[arf-3::GFP::unc-54]* and *icbIs18[arf-3::GFP::unc-54]*, respectively. Therefore, there is natural variation in CB4856 suppressing the V6a phenotype at 25 °C that can be mapped using quantitative trait locus (QTL) mapping. Additionally, there could be more natural variation in *wild isolates* not tested in this study affecting the frequency of extra cell in the V6a lineage. Other wild isolates could be phenotyped for this trait and causative variation could be mapped using GWAS.

4.3.2 No differences in the expressivity of mutations in *lin-22*, *bro-1* and *eff-1*

In order to test phenotypic differences in the expressivity of different mutant alleles, I employed two approaches. First approach was to produce de novo mutations in genes affecting SCN. Second was to introgress existing mutations in genes affecting SCN from N2 into wild isolates. The CRISPR-Cas9 approach had to be efficient and precise in order to produce the same mutation in multiple wild isolates. The co-CRISPR approach, while great for screening worms with successful genome editing events, was still dependent on the sgRNA for the gene of interest being efficient (Arribere et al., 2014). The CRISPR-Cas9 approach was not successful in 2015, because the chosen sgRNA against many genes *egl-18*, *rnt-1*, *nhr-25* and *eff-1* may not be efficient.

The CRISPR-Cas9 approach was successful in the editing of CBF β homologue (binding partner of Runx) *bro-1* and *Hes*-related bHLH transcription factor *lin-22* genes. The sgRNA for *lin-22* had a GGNGG motif (Farboud and Meyer, 2015), which made it efficient, produced

multiple mutations in *lin-22* in very few attempts. In fact, *icb49* was recovered independently from two different injections in the CB4856 (henceforth referred to as Hawaii) background. *icb49* mutation was produced in N2 and CB4856 but not in the JU2007 isolate. However, there was no difference in SCN between N2 and CB4856 carrying *lin-22(icb49)*. I recovered four different (*icb44*, *icb45*, *icb46* and *icb47*) but not same alleles in *bro-1* after multiple attempts. Only *icb45*, a 9 bp deletion allele was produced in CB4856 and JU2007 but not in N2. However, there was no difference in SCN between CB4856 and JU2007 carrying *bro-1(icb45)*. There was no differential expressivity of the mutations in the strains that were tested. However, these experiments were not exhaustive and there is scope for improvement in the CRISPR-Cas9 protocol to produce the same mutation in more number of wild isolates.

There are many algorithms which predict the likelihood of successful genome editing with a specific sgRNA like <http://genome.sfu.ca/crispr/search.html> specifically for *C. elegans*, <http://crispr-era.stanford.edu/contact.jsp> and <https://www.benchling.com/crispr/>. There is on-going research on getting more precise genome editing events in *C. elegans*. One study inserted the highly efficient sgRNA targeting *dpy-10* into the gene of interest and a second round of CRISPR targeted the gene of interest using the same sgRNA *dpy-10* with an repair template to scarlessly edit the genome. (El Mouridi et al., 2017). CRISPR-Cas9 ribonucleoprotein complexes and use of single stranded DNA oligos with short homology arms (30 bp – 60 bp) were found to increase the editing efficiency, and without any cloning involved in the protocol (Paix, Folkmann, and Seydoux, 2017). A high-throughput method on getting putative null mutations using a universal 43-nucleotide-long knock-in cassette, which has stop codons in all three reading frames has been described (Wang et al., 2018). I used Cas9 plasmid *Peef-1A.1::Cas9-SV40-NLS::tbb-2 3'UTR* from Friedland et al. (2013), which was shown to be less efficient and have low fidelity for precise knock-ins compared to pDD162 (*Peef-1A.1::Cas9* + sgRNA) (Dickinson et al., 2013). Also, *Peef-1A.1::Cas9-SV40-NLS::tbb-2 3'UTR* caused insertion-deletion (InDels) mutations instead of precise knock-ins in contrast to pDD162. The two Cas9 constructs had the same promoter, the same Cas9 sequence and the same 3' UTR, however, the C-terminal tag attached to the Cas9 was found to play a critical role in determining the editing efficiency and the fidelity of Cas9. Also, a flexible linker between NLS and Cas9

was also important for editing accuracy of Cas9 (Zhao et al., 2016). If the CRISPR strategy were to be repeated successfully, it would have to be with CRISPR-Cas9 ribonucleoprotein complexes and with short single stranded DNA oligo repair template containing a universal 43-nucleotide-long knock-in cassette.

Introgression of known mutations into different wild isolates is a time consuming process involving backcrossing at least 10× that takes about 5 weeks. However, it is a time tested method for discovering natural genetic variation (Gibson and Helden, 1997). We introgressed a nematode specific fusogen *eff-1(icb4)* from N2 into JU2007. However, there was no difference in SCN between N2 and JU2007 carrying *eff-1(icb4)*.

4.3.3 There are differences in the expressivity of mutations in *egl-18* between wild isolates

Introgression of a null mutation *egl-18(ga97)* and hypomorphic mutation *egl-18(ok290)* from N2 into JU2007 and CB4856 showed that there was a difference in SCN between CB4856 and N2/JU2007. This difference in expressivity between the isolates was not recapitulated by RNAi knockdown of *egl-18*. Phenotypic differences between RNAi knockdown and null mutants have been observed in several model organisms. One study in *Danio rerio* found that knockdown of an endothelial extracellular-matrix (ECM) gene *egf17*, unlike mutations in the gene, produce severe vascular defects. Based on their observations, they proposed the activation of a compensatory response in the gene network to buffer against null mutations but not RNAi knockdown (Rossi et al., 2015). Recently, such a mechanism termed genetic compensation response (GCR) was discovered in *Danio rerio*. GCR's activation was dependent on the simultaneous presence of mRNA bearing premature termination codon (PTC) and the nucleotide sequence of the transgene mRNA, which is homologous to compensatory genes. Members of the nonsense-mediated mRNA decay (NMD) pathway and COMPASS complex was involved in GCR (Ma et al., 2019). Interestingly, *egl-18* has a paralogue *elt-6* which functions redundantly to promote seam cell fate (Koh and Rothman, 2001; Gorrepati, K. W. Thompson, and Eisenmann, 2013). The difference in SCN between the wild isolates abrogated upon knockdown of *elt-6* suggesting that the difference in the expressivity of *egl-18(ga97)* requires function of *elt-6*. There are no statistically significant differences in SCN between wild

isolates in wild-type animals upon knockdown of *elt-6*. The difference in SCN between wild isolates carrying *egl-18(ga97)* is cancelled upon knockdown of *TCF/pop-1* suggesting that the difference in the expressivity of *egl-18(ga97)* requires function of *pop-1*. *pop-1* is the end point mediator of Wnt signalling pathway. Therefore, differences in the response of Wnt signalling pathway to loss of *egl-18* may underlie differential expressivity of *egl-18(ga97)*.

egl-18 function is required in seam cells during embryonic and postembryonic development (Koh and Rothman, 2001; Gorrepati, K. W. Thompson, and Eisenmann, 2013). The difference in the expressivity of SCN between N2 and CB4856 was present in postembryonic (L2 – L4) divisions and not after embryonic and L1 division. There was higher loss of anterior seam cells (H0, H1a, H1p, V1a, V1p, V2a, V2p) in *egl-18(ga97)* compared to mid (V3a, V3p) or posterior (V4a, V4p, V5, V6a, V6p, T) seam cells. However, there were greater losses of seam cells in all three categories in N2 carrying *egl-18(ga97)* in contrast to CB4856. *egl-18* is involved in vulval cell fate specification and inhibiting fusion of the vulval primordium (Eisenmann and S. K. Kim, 2000; Koh, Peyrot, et al., 2002). However, there were no differences in the expressivity of *egl-18(ga97)* in the vulva. Consistent with this finding, a polymorphism *haw6805*, which was found to affect the expressivity of mutations in (EGFR) gene (*let-23*) by Dubeau and Félix (2012) did not rescue the phenotype of N2 to that of CB4856 carrying *egl-18(ga97)*. Therefore, the differential expressivity of *egl-18* mutations are specific to the seam tissue.

4.3.4 Genetic basis of cryptic genetic variation affecting SCN in wild isolates carrying *egl-18(ga97)*

One mechanism of robustness is redundancy in the form of gene duplicates in a genetic pathway (Woollard, 2005; Tischler et al., 2006; Masel and Siegal, 2009; Félix and Barkoulas, 2015). Therefore, it is conceivable that *elt-6* in CB4856 compensates for the loss of *egl-18*. However, this compensation could not be due to cis-regulatory variation or sequence specific differences in *elt-6* between N2 and CB4856 because *elt-6* from N2 was introgressed into CB4856 along with *egl-18(ga97)* and *elt-6* is directly downstream of *egl-18*. The putative polyadenylation site of *egl-18*, and the trans-splice site *elt-6* are separated by only ≈ 130 bp (Koh and Rothman, 2001). However, there could be trans-regulatory variation in CB4856 affecting the expression of *elt-6* in order to compensate for the loss of *egl-18*. An alternate mechanism is the genetic compensation

response (GCR) discovered in *Danio rerio* that could compensate for loss of function of a gene by producing gene expression changes in another related homologous gene. To test this hypothesis, I quantified mRNA transcripts of *elt-6* in *egl-18(ga97)* animals in N2 and CB4856 backgrounds. I found no statistically significant differences between the *elt-6* mRNA transcripts between the two isolates carrying *egl-18(ga97)*. This result suggests that while the increase in gene expression of *elt-6* is not responsible for the differential expressivity of *egl-18(ga97)* between N2 and CB4856, it is required based on the RNAi knockdown result. Therefore, the underlying causative genetic variation for the differential expressivity of *egl-18(ga97)* mutation needs to be investigated using quantitative genetics methods like QTL mapping.

Chapter 5

Mapping Genetic Variation Underlying Differences in Expressivity of the *egl-18(ga97)* mutation Between N2 And CB4856

5.1 Introduction

Seam cell lineages and seam cell number (SCN) in *C. elegans* development are mostly invariant (Sulston and Horvitz, 1977), which is indicative of a tightly regulated developmental process. The presence of developmental buffering mechanisms facilitate the accumulation of conditionally neutral genetic variation called cryptic genetic variation (CGV) (Gibson and Dworkin, 2004; Félix and Wagner, 2008). CGV is a type of genetic variation that does not affect phenotype in wild-type genetic context and the typical environmental conditions encountered by an organism. However, CGV may modify the phenotype upon introduction of novel mutations or environmental change. CGV represents a hidden pool of variation that can facilitate adaptation in the face of mutations or environmental change. Therefore, CGV has broad implications for adaptive potential as genetic variation is the ultimate substrate on which evolution acts.

In the previous chapter, we discovered cryptic genetic variation affecting SCN in *C. elegans* wild-isolates upon perturbation involving mutations in *egl-18*, a GATA transcription factor that is a target of the Wnt/ β -catenin asymmetry (W β A) pathway, and is required for specifying seam cell fate. Mutations in *egl-18* cause loss of seam cell fate — where both seam cell daughters after division adopt hyp7 fate — leading to a decrease in SCN (Gorrepati, K. W. Thompson, and Eisenmann, 2013). We found that the expressivity of mutations in GATA transcription factor *egl-18* varied between wild-isolates (Fig. 4.6B). Specifically, there were two more seam cells on average in CB4856 compared to N2 or JU2007 in the *egl-18(ga97)* (mutant) background. N2 is a laboratory-adapted strain originally isolated from Bristol, UK; JU2007 and CB4856 are wild-isolates that were isolated from Isle of Wight, UK and Honolulu, Hawaii, respectively. We aimed to discover the genetic basis for the difference in the expressivity of *egl-18(ga97)* allele specifically between N2 and CB4856. We picked N2 (not JU2007) and CB4856 because N2 is used as a reference strain and genomic differences between N2 and CB4856 were known already (Wicks et al., 2001; O. A. Thompson et al., 2015).

SCN is variable in *egl-18(ga97)* mutant and can be studied as a quantitative trait. Quantitative genetics represent a powerful method to dissect the genetic basis of such traits. Previous studies successfully employed a quantitative genetics approach to discover CGV that underlies a difference in the expressivity of mutations in epidermal growth factor receptor (EGFR) gene

let-23 in vulval development between *C. elegans* wild isolates (Duveau and Félix, 2012). The authors produced recombinant inbred lines (RILs) by crossing the parents followed by SNP-marker based genotyping to discover quantitative trait loci (QTLs) associated with high/low vulval induction index phenotype. Using SNP-marker based genotyping of several recombinants, they narrowed a QTL on chr. I to a 183 kb region which harboured a non-synonymous polymorphism *haw6805* in RNA cytidine acetyltransferase (*nath-10*) gene. Recently, the bulk segregant analysis combined with whole-genome sequencing (BSA-WGS) was used to discover QTLs that are associated with temperature-sensitive mortal germline phenotype, which is a multi-generational phenotype in *C. elegans* wild-isolate (Frézal et al., 2018). Briefly, BSA is an approach where DNA from F2 animals/plants from a cross sharing the same phenotype but not genotype is pooled. The rationale for bulking the samples is that two groups (bulks) with opposing phenotypic characteristics will be genetically similar in all genomic regions except for the causative genomic region for the phenotype (Michelmore, Paran, and Kesseli, 1991; Wicks et al., 2001). Therefore, QTLs can be detected by genotyping only two bulks and not every single RIL.

We followed a quantitative genetics approach to investigate the genetic basis of differential expressivity of *egl-18(ga97)* between N2 and CB4856. To this end, we built RILs and BSA-WGS to discover QTLs affecting SCN in *egl-18(ga97)*. In order to find the genomic interval of QTLs, we depooled the bulked samples using genetic markers based on deletions in CB4856 genome. We validated the QTLs discovered by producing near isogenic lines (NILs) and narrowed down the genomic interval of causative QTLs using classical genetics. We prioritised candidate genes in QTLs that may modulate SCN in *egl-18(ga97)* through an RNAi screen. Further, we investigated Wnt/ β -catenin asymmetry ($W\beta A$) pathway between N2 and CB4856, which is upstream of *egl-18* in specifying seam cell fate. The differential expressivity of *egl-18(ga97)* between N2 and CB4856 reveals CGV for the first time in the seam cell network. The work presented in the chapter is aimed at discovering the genetic nature and understanding the genetic architecture of CGV affecting SCN.

5.2 Results

5.2.1 Construction and phenotyping of the recombinant inbred lines between Bristol and Hawaii strains carrying *egl-18(ga97)*

To discover the molecular nature of the variation underlying the differential expressivity of *egl-18(ga97)* mutation between Bristol and Hawaii, we used QTL mapping. To this end, we produced 116 RILs by crossing the parents carrying the *egl-18(ga97)* mutation (section 2.5.3, and Fig. 5.1A). Phenotypic distribution was continuous (average SCN = 9.96 – 14.61), and most RILs had average SCN that was intermediate between the two parents (average SCN = 11.33 – 13.72). We observed transgressive segregation of SCN in RILs — some RILs had SCN higher and lower than parents — suggesting that there are multiple genetic loci underlying the differential expressivity of *egl-18(ga97)* phenotype. We followed a BSA-WGS approach for mapping the genetic loci (Michelmore, Paran, and Kesseli, 1991; Wicks et al., 2001; Frézal et al., 2018). To this end, extremes of the phenotypic distribution that resembled one of the two parents were pooled together in two groups for whole genome sequencing. 22 and 24 RILs were combined for the low-bulk (low average SCN like N2) and high-bulk (high average SCN like CB4856) pool, respectively. In addition, we selected 10 RILs, which showed consistently lower SCN or higher SCN compared to N2 or CB4856, respectively in the phenotypic distribution and pooled them into two groups (stringent low-bulk and high-bulk) to increase the likelihood of finding a QTL.

5.2.2 Bulk segregant analysis of RILs between N2 and CB4856 carrying *egl-18(ga97)* identifies four QTLs

To map the genetic variation underlying the differential expressivity, we genotyped the four bulked samples (low-bulk and high-bulk) representing pools in two conditions (stringent and less stringent) using whole genome sequencing. We analysed next generation sequencing data using CloudMap pipeline to get the single nucleotide polymorphism (SNP) frequency across the genome relative to the parental genomes as described in section 2.4 (Minevich et al., 2012). According to the null hypothesis, there should be no statistically significant differences in the

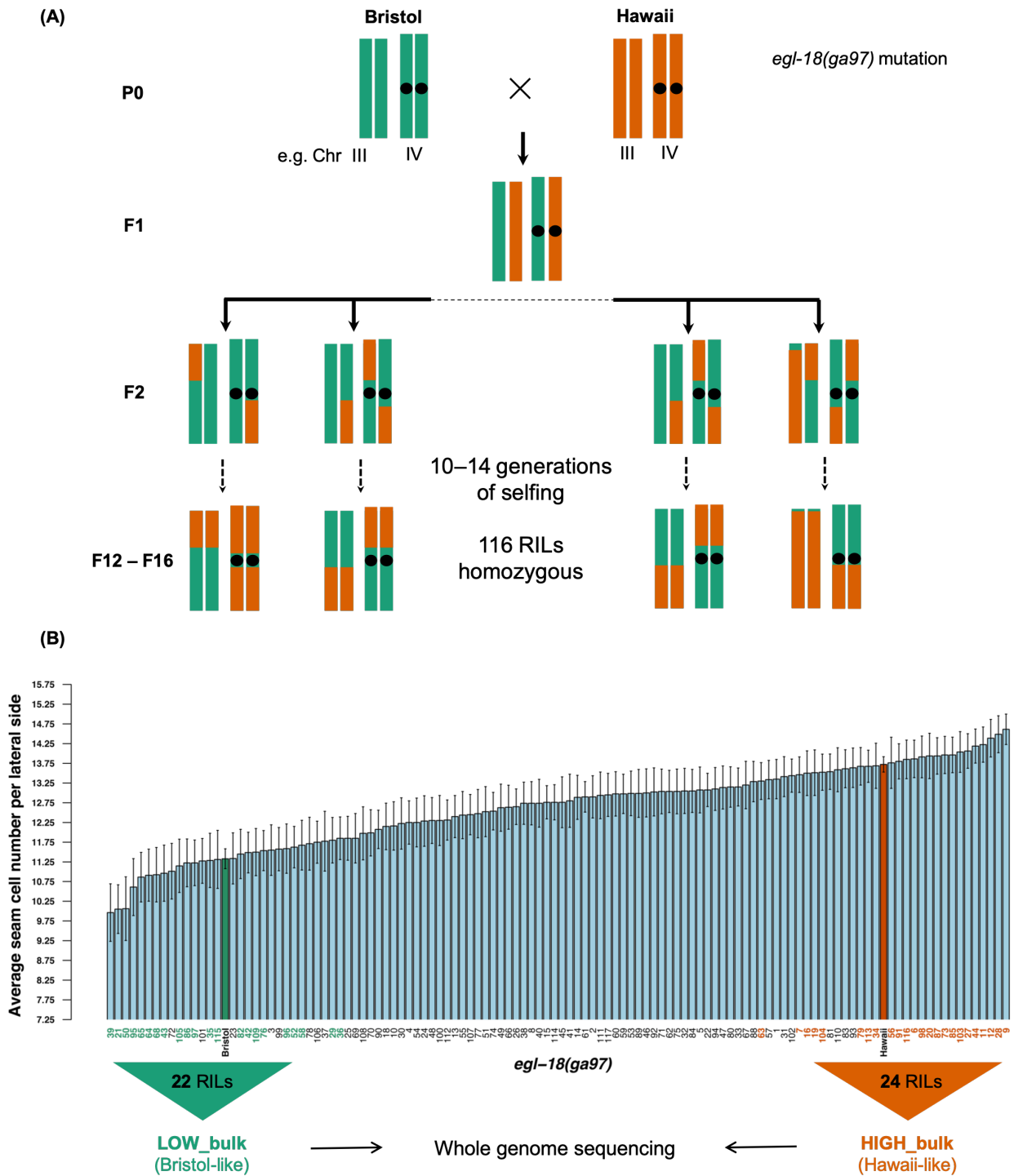


Figure 5.1: **Generation and phenotyping of recombinant inbred lines.** (A) Simplified cross scheme for generating RILs. Two chromosomes are presented. Chr. V not presented here carries the *SCMp::GFP* and *dat-1::gfp* marker. Due to asynchronous growth caused by the egg-laying defective phenotype of *egl-18* mutant, RILs went through different generations (10 – 14) of selfing (B.3). (B) Seam cell number for 116 RILs averaged from two independent replicates ranked from lowest to highest. Parental strains are shown in green and orange. The two extremes (low-bulk and high-bulk) of the phenotypic distribution were pooled for whole genome sequencing. RILs (RIL number highlighted in green and orange) which showed consistent SCN between the two replicates were selected in the two respective pools. $77 \leq n \leq 80$ per RIL. $n = 520$ for parental strains. Error bars indicate average SCN \pm 95% confidence intervals.

SNP frequencies between low-bulk and high-bulk samples in both stringent and less stringent conditions. We calculated log-odds ratio in the less stringent condition to assess the significance of deviation in SNP frequencies at $\alpha = 0.05$ (Fig. 5.2B,D,F). We observed significant deviation in SNP frequencies between low-bulk and high-bulk samples on chromosomes II (3 Mb – 13 Mb), III (3 Mb – 8 Mb) and X (3 Mb – 5 Mb) (Fig. 5.2A,C,E), indicating the presence of three QTLs. We observed a similar pattern of deviation in SNP frequencies in the stringent condition (Fig. C.3). Due to the cross scheme (section 2.5.3) used in creating the RILs, both groups had a large portion of chr. V from N2; therefore, we could not test the effect of natural variation in a large region of this chromosome. In addition, there was a significant deviation in SNP frequencies on the right arm of chr. V (18 Mb – 20 Mb) — the high-bulk contained N2 SNPs in this region compared to the low-bulk, which contained CB4856 SNPs — suggesting the presence of an antagonistic QTL on the right arm of chr. V. Reassuringly, RILs in both groups had 1.5 Mb – 2 Mb of chr. IV from N2, which represents the introgression of *egl-18(ga97)* mutation. In addition, the RILs had 2 Mb – 3.5 Mb of chr. I from N2 where the *zeel-1/peel-1* selfish element is located, which represents a known incompatibility between the two isolates (Seidel, Rockman, and Kruglyak, 2008; Seidel, Ailion, et al., 2011). In conclusion, we identified four QTLs on chromosomes II, III, V and X that may modify the *egl-18(ga97)* phenotype. Based on the log-odds ratio, QTL on chr. II, chr. III and chr. V are major QTLs followed by QTL on chr. X.

5.2.3 Genetic marker design and validation for QTL fine mapping

It was important to depool the bulked samples to know the genetic composition of RILs and to produce NILs to study the contribution of individual QTLs to SCN. For the purpose of depooling, primers were designed around genetic markers (indels, SNPs and snip-SNPs) in CB4856 compared to reference N2 genome. There are 327050 SNVs (one polymorphism per 307 bp or 3.26 per kb) and 79529 indels between the reference Bristol (N2) and the polymorphic Hawaiian (CB4856) strains (O. A. Thompson et al., 2015). SNPs rely on amplification and purification of DNA fragment containing SNP followed by Sanger sequencing for detection (listed in Appendix A.4). 3457 SNPs known as snip-SNPs modify restriction enzyme recognition sites, and can be detected as restriction fragment length polymorphisms (RFLPs) (Wicks et al., 2001). Snip-

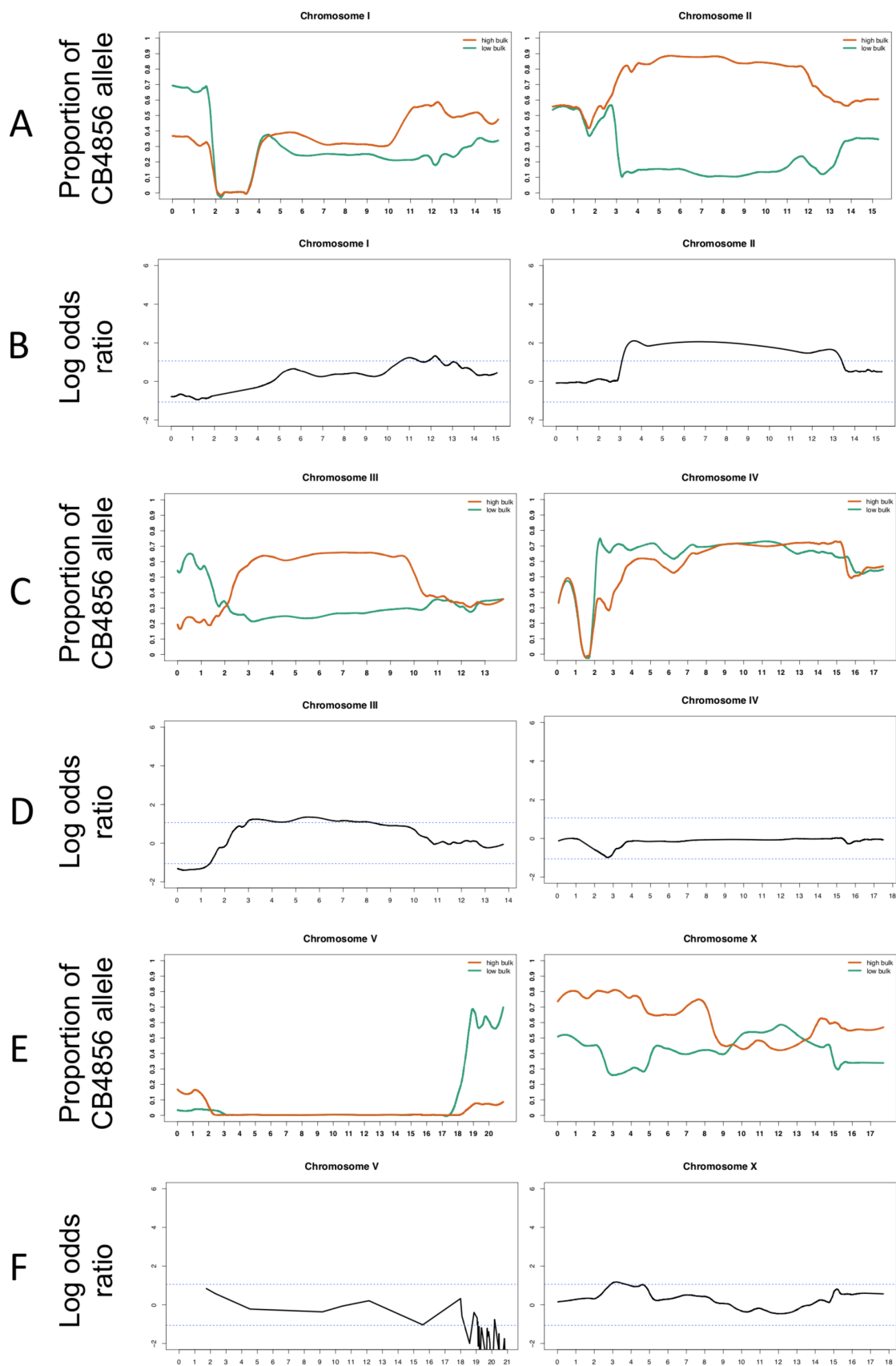


Figure 5.2: Bulk segregant analysis of recombinant inbred lines. Continued on next page.

CHAPTER 5. MAPPING GENETIC VARIATION UNDERLYING DIFFERENCES IN EXPRESSIVITY OF THE *EGL-18(GA97)* MUTATION BETWEEN N2 AND CB4856

Figure 5.2: **Bulk segregant analysis of recombinant inbred lines.** (A,C,E) Allele frequencies of CB4856 SNPs along six chromosomes. SNP frequencies in low-bulk and high-bulk are depicted by green and orange fitted curves, respectively. The curves represent locally weighted scatterplot smoothing (LOESS) regression lines from the allele frequencies at known SNP positions along the chromosomes with a span parameter of 0.1. x-axis and y-axis correspond to chromosomal position in Mb and proportion of CB4856 SNPs in the sequencing reads. 22 and 24 RILs were pooled in low-bulk and high-bulk groups, respectively. (B,D,F) Log-odds ratio of SNP frequencies along the six chromosomes. Log-odds ratio was calculated at individual SNP positions as described in section 2.5.7. The blue dashed lines indicate the thresholds for statistical significance for log-odd ratios at $\alpha = 0.05$. Processing sequencing data was performed and SNP frequency data in bulked samples shown in A were generated by Dr. Michael Fasseas, a postdoctoral scholar in the lab.

SNPs rely on amplification and purification of DNA fragment containing snip-SNP followed by restriction digestion with a restriction enzyme and DNA electrophoresis (listed in Appendix A.5). SNP detection by Sanger sequencing is expensive and snip-SNP detection is laborious due to multiple steps. Indels have an advantage over SNPs and snip-SNPs in that they can be detected easily in two steps — DNA amplification followed by DNA electrophoresis — indels are detected by a size difference between the amplicons from N2 and CB4856. However, indels are not uniformly distributed along the chromosomes. Therefore, I used indels predominantly and SNPs/snip-SNPs only when there were no indels in the genomic region. I designed primers around genetic markers as described in section 2.6.2 (listed in Appendix A.6). The agarose gels validating the genetic markers based on deletions showing the differences in the size of amplicon between N2 and CB4856 are presented in Fig. 5.3. CB4856 shows a smaller amplicon size compared to N2 due to deletions ranging from 56 bp – 293 bp.

5.2.4 Depooling the high and low-bulk RILs using genetic markers

To discover the genetic boundaries of QTLs, we depooled 24 individual RILs from high-bulk and 10 RILs from low-bulk with lowest SCN using validated genetic markers in section 5.2.3 (Fig. 5.4). Based on the preliminary genetic boundaries from genotyping individual RILs, QTLs on chr. II and chr. V may be important. 87.5% (21 out of 24) high-bulk RILs compared to 10% (1 out of 10) low-bulk RILs carry CB4856 fragment on chr. II (5.30 Mb – 8.18 Mb). 92% (22 out of 24) high-bulk RILs compared to 40% (4 out of 10) low-bulk RILs carry a part or complete N2 fragment on chr. V (18.66 Mb – 20.99 Mb). In the absence of CB4856 fragment on chr. II (RIL-16 and RIL-73), N2 fragment on chr. V and CB4856 fragment on chromosomes III and X were present suggesting that the absence of QTL on chr. II may be compensated

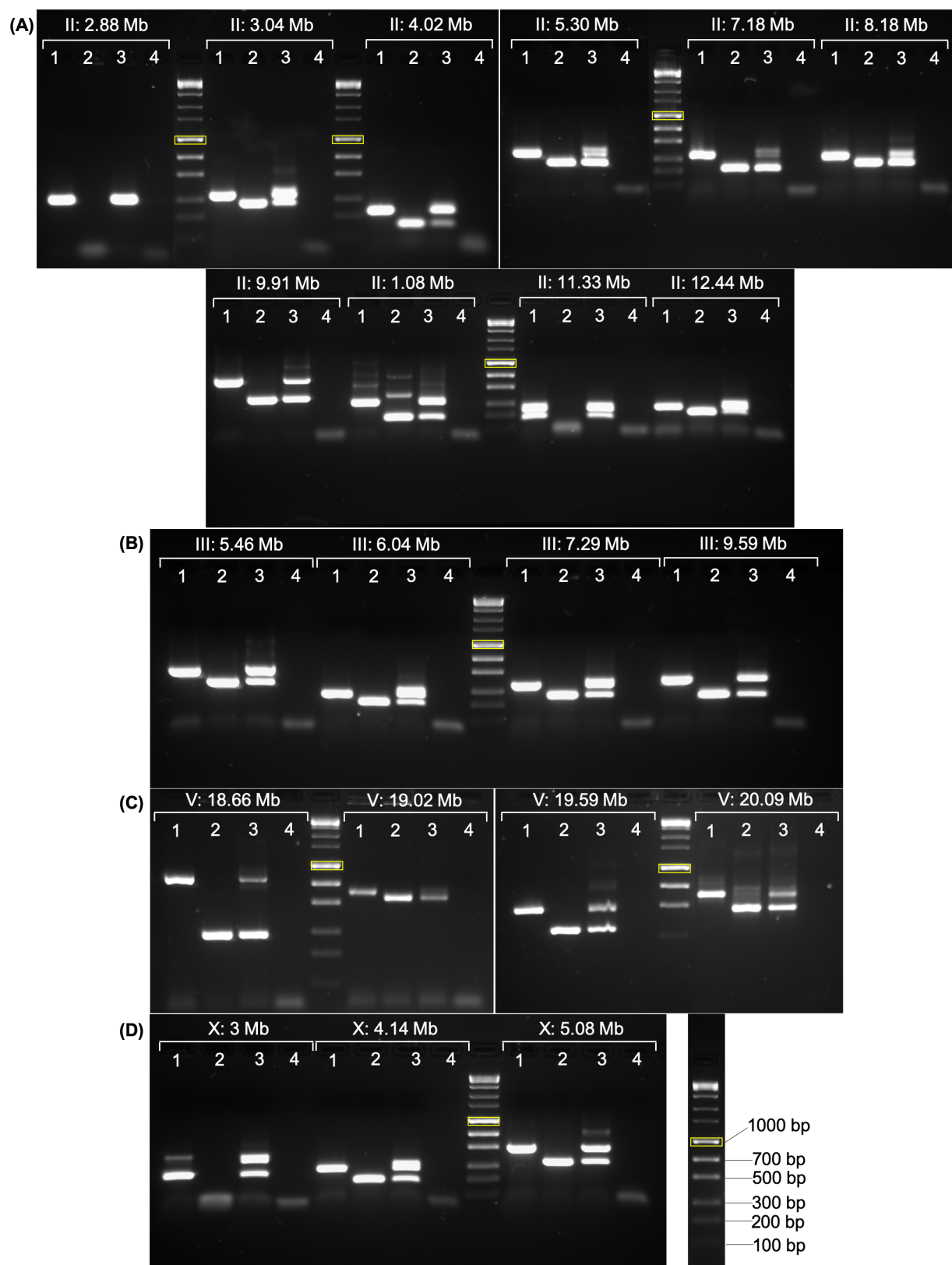


Figure 5.3: **Validation of genetic markers based on deletions in Hawaiian (CB4856) genome.** PCR products were run on 2% agarose in $1 \times$ Tris-borate-EDTA buffer for 1 h at 70 V. Samples: 1. N2, 2. CB4856, 3. N2/CB4856, 4. Negative control. (A) Ten genetic markers on chr. II. Deletions are in the range of 56 bp – 293 bp. (B) Four genetic markers on chr. III. Deletions are in the range of 60 bp – 126 bp. (C) Four genetic markers on chr. V. Deletions are in the range of 77 bp – 479 bp. (D) Three genetic markers on chr. X. Deletions are in the range of 100 bp – 173 bp.

by QTLs on chromosomes III, V and X together. Surprisingly, RIL-65 from low-bulk is the only exception, despite containing full CB4856 fragment and N2 fragment on chromosomes II and V, respectively, has low-SCN phenotype, which could suggest the presence of another QTL perhaps on chr. I. Based on the CB4856 fragments present in most high-bulk RILs and absent in most low-bulk RILs and vice versa, the interval of QTL on chr. II may be narrowed to $\approx 4.02 \text{ Mb} - \approx 8.18 \text{ Mb}$.

67% (16 out of 24) RILs in high-bulk group compared to 50% (5 out of 10) in low-bulk group carry a part or complete CB4856 fragment on chr. III (5.46 Mb – 9.59 Mb). Three high-bulk RILs (85, 16, 17) carrying the full CB4856 fragment on chr. III contained full CB4856 fragment on chr. X suggesting a positive interaction between these two fragments. Only RIL-43 (low-bulk) carried a full CB4856 fragment on chromosomes III and X. However, it contained CB4856 fragment on chr. V suggesting that CB4856 fragment chr.V antagonises the positive interaction between QTLs on chromosomes III and X. Based on the CB4856 fragments present in most high-bulk RILs and absent in most low-bulk RILs and vice versa, the interval of QTL on chr III may be narrowed to $\approx 5.46 \text{ Mb} - \approx 7.29 \text{ Mb}$.

Low-bulk RILs:RIL-64 and RIL-115 have low-SCN phenotype in spite of the presence of full CB4856 fragment on chr. X suggesting that QTL on chr. X may not act alone but may act in combination with other QTLs. In support of this, 0% (0 out of 24) of high-bulk RILs compared to 20% (2 out of 10) of low-bulk RILs carry a part or complete CB4856 fragment on chr. X (3.00 Mb – 5.08 Mb) alone.

5.2.5 The QTLs on chr II, III and X represent cryptic genetic variation

Genomes of RILs consist of shuffled parental (N2 and CB4856) genomes. To evaluate the contribution of each QTL to SCN, NILs containing individual QTLs and combination of QTLs were produced in the background of N2. Based on the narrowed QTL intervals, RIL-28 was picked for producing NILs. RIL-28 contained smaller CB4856 fragments on both chromosomes II and III; therefore, it was used to introgress individual QTLs into N2 background by backcrossing $10 \times$ to N2 as outlined in section 2.6.3. During the introgression, a smaller fragment on chr. II (4.02 Mb – 8.18 Mb) was recovered from the full fragment (2.88 Mb – 8.18 Mb) in

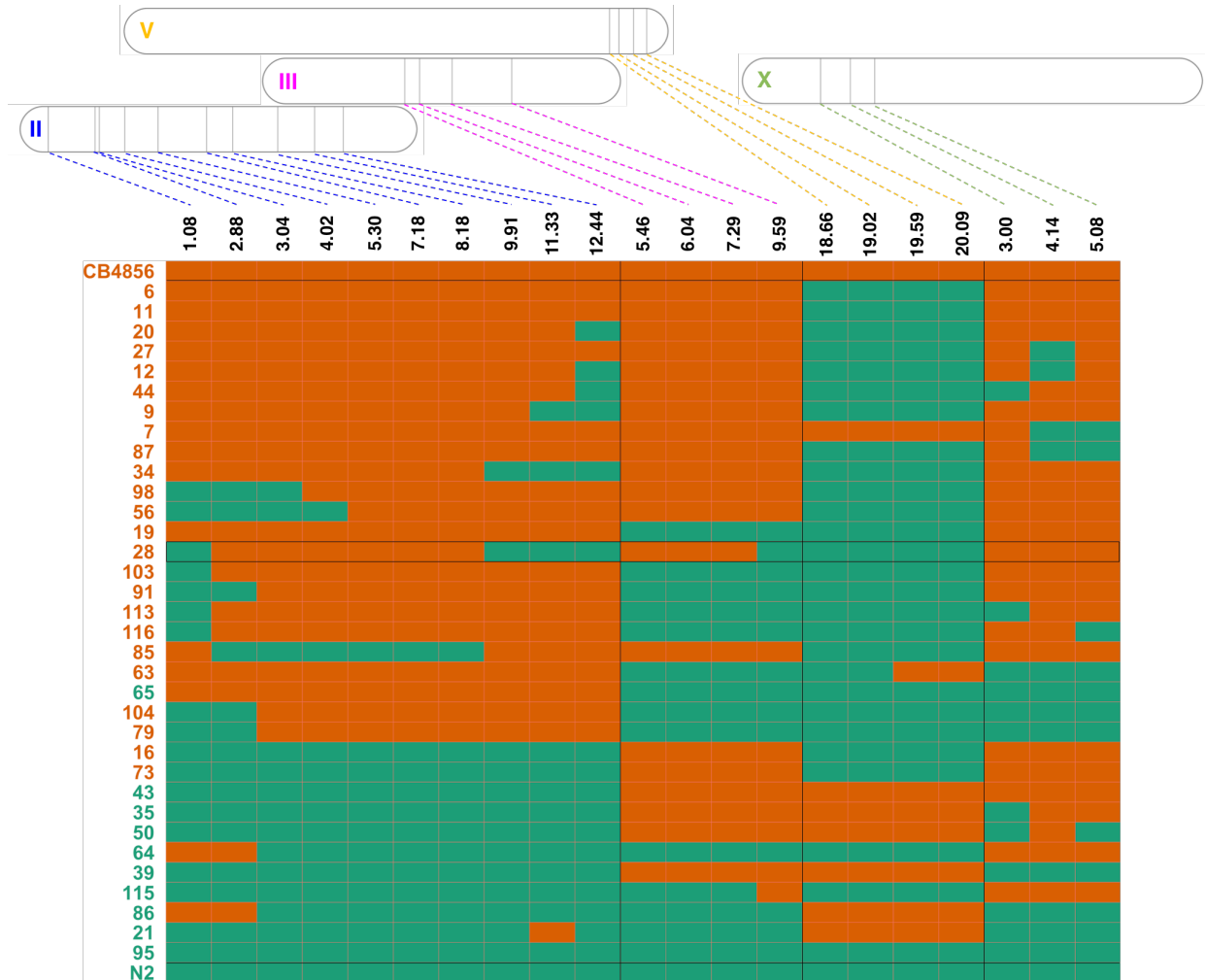


Figure 5.4: **Genotyping of recombinant inbred lines (RILs) with genetic markers based on deletions in Hawaiian (CB4856) genome.** Each row and column represents RIL line and its genetic composition at genetic markers validated in Fig. 5.3, respectively. The location of the genetic markers on the chromosomes is shown above the graph. Dotted coloured lines delineate markers on different chromosomes. Chr. II, III, V and X are blue, pink, yellow and green, respectively. Green and orange tiles represents N2 and CB4856 genomes, respectively. RILs are ranked in descending order based on number of CB4856 genomic fragments they carry. RIL numbers are colour coded based on the phenotypic similarity to their parents, N2 (green) and CB4856 (orange). Top-most and bottom-most rows delineated by black horizontal lines represent the parental genomes: CB4856 and N2, respectively. Black-box in the middle highlights genotype of RIL-28, which was used for producing near isogenic lines containing QTLs on chromosomes II, III and X. Notice smaller CB4856 fragments on chr. II and chr. III in RIL-28.

RIL-28 (Chromosomes in Fig. 5.5A). We scored SCN in wild-type NILs and found no statistically significant differences in SCN compared to control animals with no QTLs (Fig. 5.5A, $F(4, 495) = 0.87$, $p = 0.48$), indicating presence of CGV affecting SCN.

We previously found that 4.02 Mb – 8.18 Mb genomic fragment on chr. II was common to most of the high-bulk RILs and most likely contains the causative QTL. Therefore, we used this smaller genomic fragment for introducing *egl-18(ga97)* mutation. Various NIL strains containing different combinations of QTLs with *egl-18(ga97)* mutation were produced. SCN was counted in these NIL strains and we found a significant difference in SCN of NILs compared to N2 animals carrying *egl-18(ga97)* (Fig. 5.5B, $F(8, 905) = 16.55$, $p < 2.2 \times 10^{-16}$). QTLs on chromosomes II and III individually partially rescue the phenotype (increase the SCN) of the mutant in N2 ($p < 0.02$). However, QTL on chromosome X individually does not rescue the phenotype of the mutant in N2 background ($p = 0.99$). SCN in NILs containing any of the individual QTLs is significantly different to mutant in CB4856 background ($p < 0.01$) suggesting that they act in an additive manner (QTLs on chromosomes II and III) or have positive epistasis (QTLs on chromosomes III and X). SCN in NILs containing a combination of two QTLs rescues N2-mutant phenotype to that of CB4856 (double = MBA862, MBA789 and MBA790; $p_{double \text{ vs. } N2} < 2.2 \times 10^{-16}$, $p_{double \text{ vs. } CB4856} > 0.17$). Similarity NIL containing three QTLs (MBA848) rescues N2-mutant phenotype to that of CB4856 ($p_{MBA848 \text{ vs. } N2} < 2.2 \times 10^{-16}$, $p_{MBA848 \text{ vs. } CB4856} = 0.99$). The QTLs identified in the bulk segregant analysis were validated in the N2 background which contains QTL on chr. V. Furthermore, these results match the asymmetric distribution of QTLs in the high-bulk and low-bulk RILs (Fig. 5.4 and 5.1B) suggesting that a combination of two or more is required for converting the N2-mutant phenotype to CB4856-mutant phenotype. Altogether, QTLs alter SCN in mutant but not in wild-type condition showing that there is CGV affecting SCN.

5.2.6 Breaking the large genomic intervals of major QTLs residing on chromosomes II and III into smaller intervals

We discovered QTLs with a large genomic interval. In order to narrow down the genomic interval of two major QTLs on chromosomes II and III, we screened for recombinants after crossing NILs to N2 strain. Additionally, we retained recombinants that were produced by

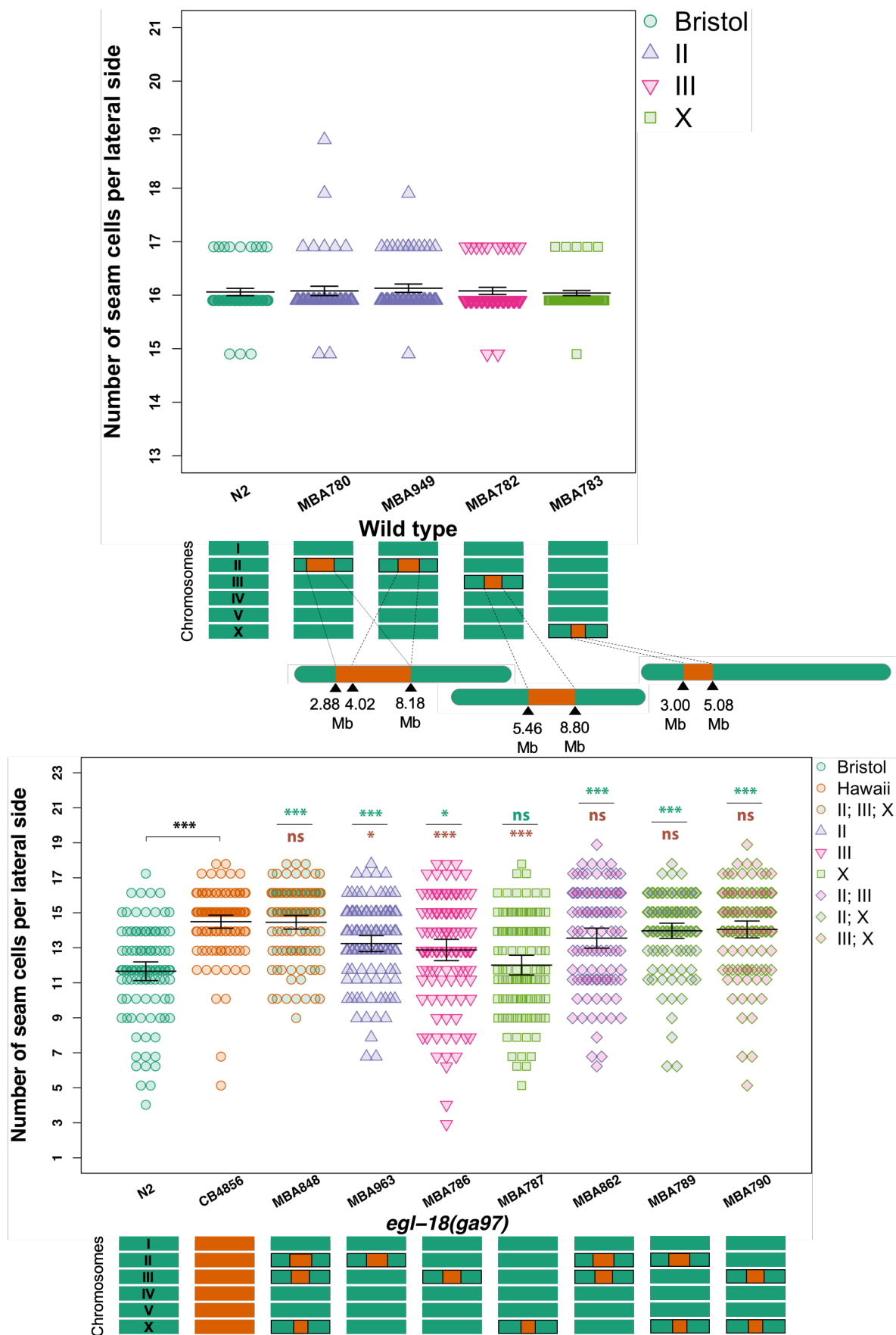


Figure 5.5: Phenotypic analysis of near isogenic lines indicates two major quantitative trait loci. Continued on next page.

CHAPTER 5. MAPPING GENETIC VARIATION UNDERLYING DIFFERENCES IN EXPRESSIVITY OF THE *EGL-18(GA97)* MUTATION BETWEEN N2 AND CB4856

Figure 5.5: **Phenotypic analysis of near isogenic lines indicates two major quantitative trait loci.** (A) Seam cell number in near isogenic lines containing individual QTLs. QTLs do not affect SCN in wild-type condition. Chromosomes below the graph depict the genotype of the strain. A smaller fragment of chr. II (4.02 Mb – 8.18 Mb) was recovered that was used in the mutant condition. One-way ANOVA shows no statistically significant differences in SCN of NILs compared to wild-type with no QTLs ($F(4, 495) = 0.87, p = 0.48$). $n = 100$ per strain. (B) Seam cell number in near isogenic lines containing individual and combination of QTLs and *egl-18(ga97)* mutation. Chromosomes below the graph depict the genotype of the strain. One-way ANOVA showed that SCN was significantly affected by the strain ($F(8, 905) = 16.55, p < 2.2 \times 10^{-16}$). QTLs on chromosomes II and III but not X converted SCN of mutant in N2 background to that of mutant in CB4856 background. Combinations of two QTLs and all three QTLs converted SCN of N2-mutant to that of CB4856-mutant. $100 \leq n \leq 105$ per strain. Error bars indicate average SCN \pm 95% confidence intervals. *** $p < 1 \times 10^{-4}$ in green and orange correspond to significant differences by post hoc Tukey HSD compared to N2 and CB4856, respectively.

recombination events during the process of producing NILs. We scored seam cell number in recombinants of chromosomes II and III (Fig. 5.6A and Fig. 5.7A, respectively). We genotyped the recombinants using additional markers based on indels in CB4856 genome (2.6.2) on chromosomes II and III (Fig. 5.6B and Fig. 5.7B, respectively). By correlating the genotype with SCN, we narrowed down the interval of the two QTLs.

We found that NILs (MBA784, MBA963, MBA944, MBA846 and MBA785) containing overlapping genomic fragments on chr. II that converted the phenotype of N2-mutant to CB4856-mutant ($p_{NIL \text{ vs. } N2} < 0.0042, p_{NIL \text{ vs. } CB4856} > 0.088$). Conversely, MBA951 containing ≈ 3.04 Mb – ≈ 5.30 Mb of chr. II did not convert N2-mutant phenotype to CB4856-mutant ($p_{MBA951 \text{ vs. } N2} = 0.99, p_{MBA951 \text{ vs. } CB4856} = 2.2e - 16$). SCN of two NILs (MBA1005 and MBA1006) containing ≈ 6.44 Mb – ≈ 6.84 Mb on chr. II was significantly higher compared to N2-mutant ($p_{NIL \text{ vs. } N2} < 0.0012$), but lower compared to CB4856-mutant ($p_{NIL \text{ vs. } CB4856} < 0.0091$). Conversely, MBA1012 that was missing this genomic region of chr. II had lower SCN compared to N2-mutant and CB4856-mutant ($p_{MBA1012 \text{ vs. } N2} = 1.3e - 04, p_{MBA1012 \text{ vs. } CB4856} = 2.2e - 16$). Additionally, genomic region between ≈ 6.44 Mb – ≈ 6.84 Mb on chr. II was common to all these NILs that had higher SCN compared to N2-mutant (Fig. 5.6B) suggesting that it may harbour the causative genetic variation for differential SCN. We found that two NILs — MBA944 and MBA846 — had mutually exclusive genomic regions with a small overlapping region around ≈ 6.48 Mb suggesting that either there are two independent QTLs on chr. II contained in ≈ 4.78 Mb – ≈ 6.44 Mb and ≈ 6.84 Mb – ≈ 7.18 Mb, respectively or the causative genetic variation is contained within the overlapping region of less than ≈ 0.14 Mb. Genes car-

rying natural variation in this genomic region are listed in appendix (Appendix B.4). We found that there are seven genes in the overlap between MBA944 and MBA846 containing natural variation. Non-overlapping regions in MBA944 and MBA846 contained 106 and 116 genes with natural variation.

We found NILs (MBA786, MBA819 and MBA945) containing overlapping genomic fragment of ≈ 7.03 Mb – ≈ 8.91 Mb of chr. III converted N2-mutant to CB4856-mutant phenotype ($p_{NIL \text{ vs. } N2} < 7.3e - 05, p_{NIL \text{ vs. } CB4856} > 0.73$). Conversely, NILs (MBA964 and MBA1056) containing ≈ 5.46 Mb – ≈ 8.26 Mb of chr. III does not convert N2-mutant to CB4856-mutant phenotype ($p_{NIL \text{ vs. } N2} < 0.12, p_{NIL \text{ vs. } CB4856} < 2.2e - 16$) suggesting that this genomic area did not harbour the causative genetic variation for differential SCN. Taken together, the genomic area present in the NILs that convert N2-mutant to CB4856-mutant phenotype but lacking in the NILs that do not convert N2-mutant to CB4856-mutant phenotype is ≈ 7.77 Mb – ≈ 8.91 Mb. Therefore, we narrowed down the QTL to a ≈ 1.14 Mb genomic fragment on chr. III. Genes carrying natural variation in this genomic region are listed in appendix (Appendix B.5). We found 59 genes in the common genomic fragment of NILs that convert N2-mutant to CB4856-mutant.

5.2.7 Identifying candidate genes in the QTLs on chromosomes II and III through an RNAi screen

In order to prioritise candidate genes to find the molecular nature of QTLs on chromosomes II and III, we performed RNAi of candidate genes in wild-type in N2 and CB4856 background. We hypothesised that the genes responsible for the differences in SCN between N2-mutant and CB4856-mutant may affect SCN in wild-type. In addition, SCN may be decreased in N2-mutant and CB4856-mutant upon RNAi knockdown of this gene as it no longer can buffer/modify the *egl-18(ga97)* phenotype. We thought it is unlikely that this gene would only function in CB4856 and not N2. Instead it may be functional in both isolates but to varying degrees. Therefore, we also performed RNAi of candidate genes in mutant (*egl-18(ga97)*) condition in N2 and NIL containing all three QTLs (MBA848). We prioritised the genes that are known to have seam cell expression according to wormbase or annotated to be involved in development or Wnt pathway.

RNAi in *C. elegans* is performed by feeding them dsRNA-expressing bacteria and down-regulating the target genes in the germline. However, CB4856 is known to harbour natural

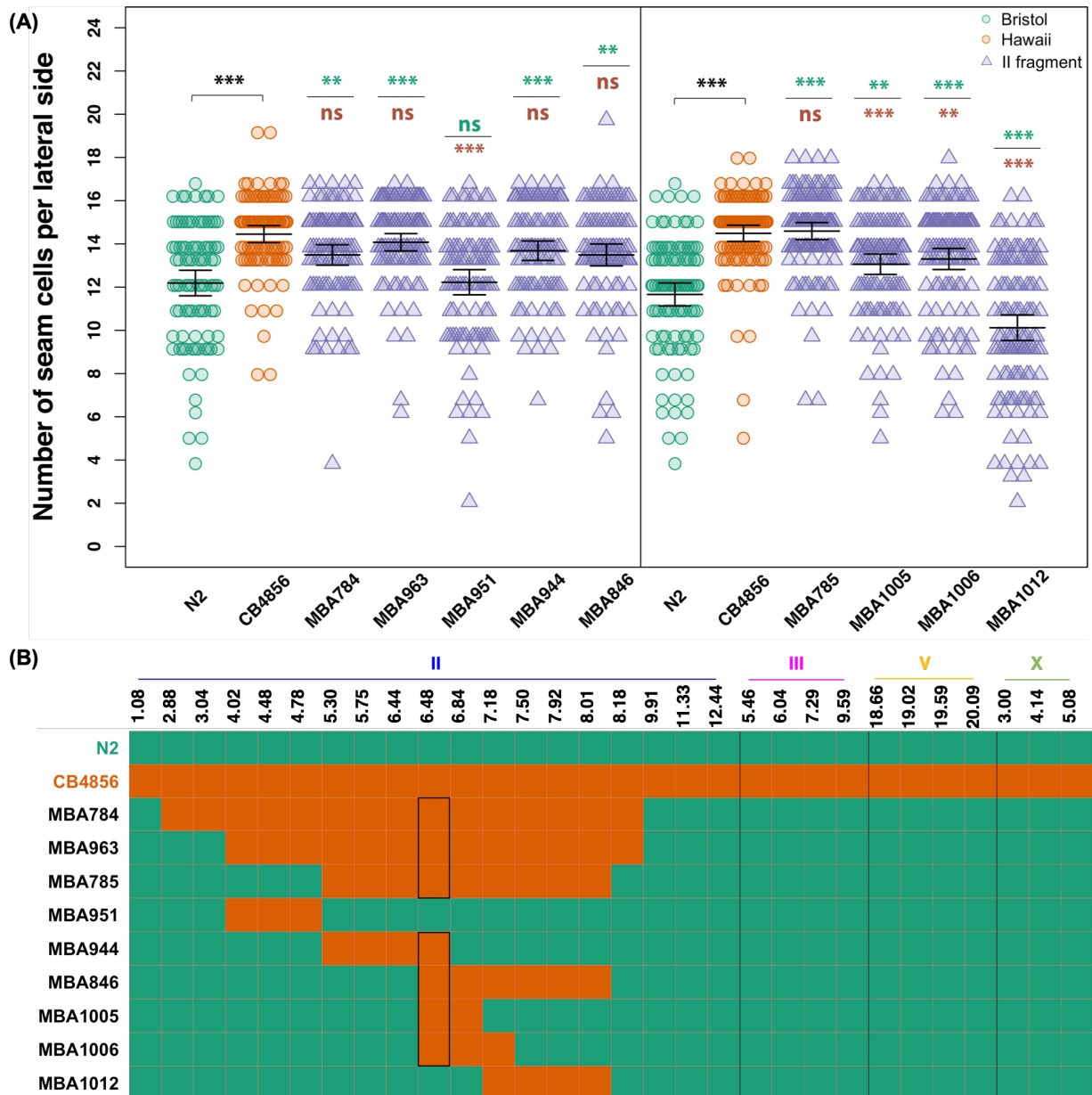


Figure 5.6: **Phenotypic and genotypic analysis of NILs carrying genomic fragments of chromosome II from CB4856.** (A) Seam cell number in near isogenic lines containing fragments of chr. II from CB4856 and the *egl-18(ga97)* mutation. Vertical black line in the graph distinguishes between two sets of experiments. In both experiments, one-way ANOVA showed that SCN was significantly affected by the strain ($F(6, 635) = 12.25$, $p < 4.5 \times 10^{-13}$; $F(5, 622) = 51.18$, $p < 2.2 \times 10^{-16}$). MBA951 and MBA1012 from the first and second experiment did not convert the phenotype of N2-mutant to CB4856-mutant. $90 \leq n \leq 93$ and $100 \leq n \leq 116$ per strain in the first and second experiment, respectively. Error bars indicate average SCN \pm 95% confidence intervals. *** $p < 1 \times 10^{-4}$ in green and orange correspond to significant differences by post hoc Tukey HSD compared to N2 and CB4856, respectively. (B) Genotyping of NILs carrying genomic fragments of chr. II from CB4856 using genetic markers based on indels in CB4856 genome. Each row and column represents RIL line and its genetic composition at genetic markers, respectively. The location of the genetic markers on the chromosomes is shown above the graph. Green and orange tiles represents N2 and CB4856 genomes, respectively. The NILs carry varying fragment size of chr. II in the N2 background. Black-box in the middle highlights genomic region around ≈ 6.48 Mb on chr. II that is common to the NILs that convert N2-mutant to CB4856-mutant phenotype and absent in the NILs that do not. Therefore, this region may contain the causative genetic variation for differential expressivity of *egl-18(ga97)* mutation.

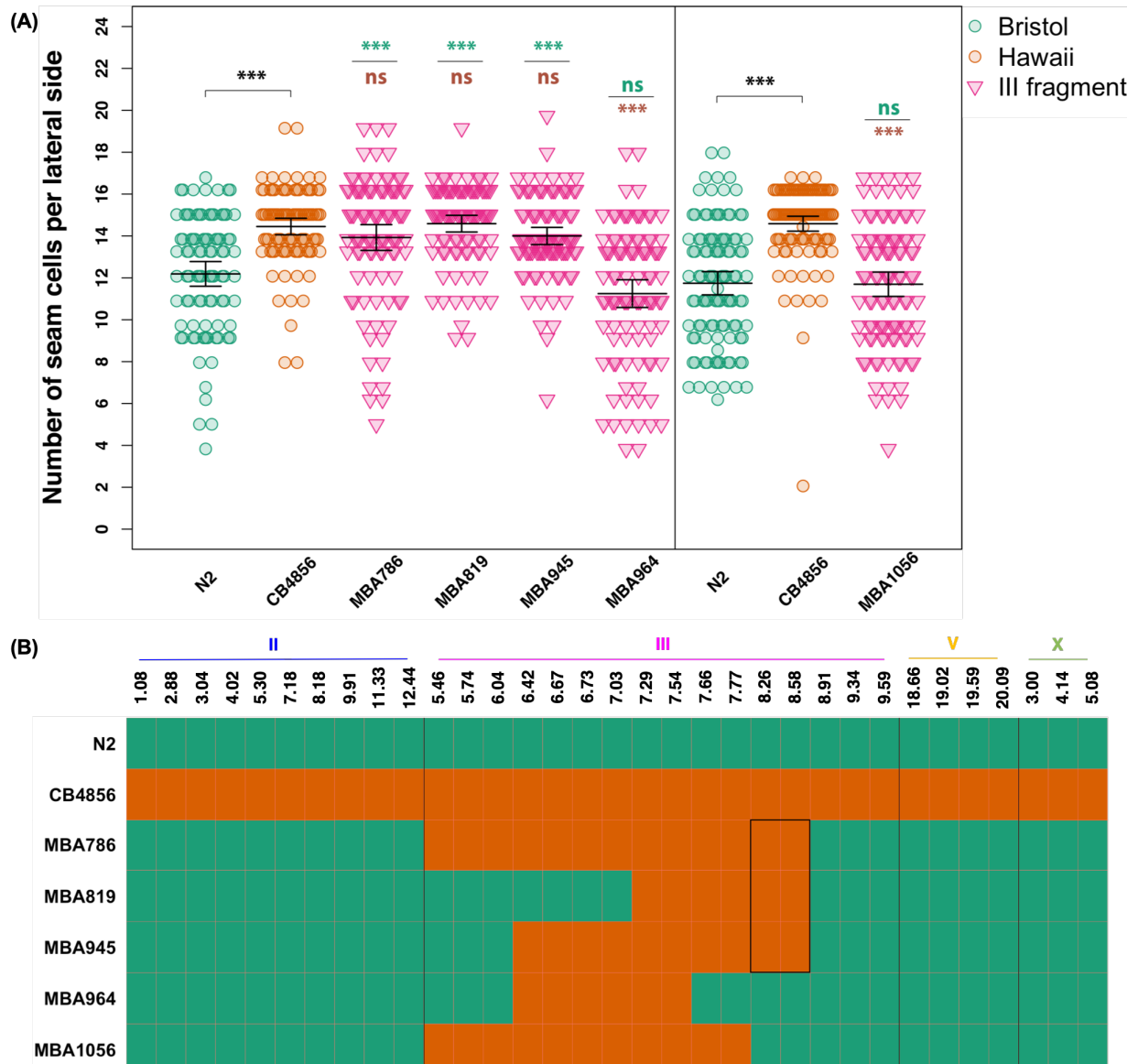


Figure 5.7: **Phenotypic and genotypic analysis of NILs carrying genomic fragments of chromosome III from CB4856.** (A) Seam cell number in near isogenic lines containing fragments of chr. III from CB4856 and the *egl-18(ga97)* mutation. Vertical black line in the graph distinguishes between two sets of experiments. In the both experiments, one-way ANOVA showed that SCN was significantly affected by the strain ($F(5, 556) = 26.58, p < 2.2 \times 10^{-16}$; $F(2, 313) = 43.31, p < 2.2 \times 10^{-16}$, respectively). MBA951 and MBA1012 from the first and second experiment did not convert the phenotype of N2-mutant to CB4856-mutant. $90 \leq n \leq 97$ and $102 \leq n \leq 109$ per strain in the first and second experiment, respectively. Error bars indicate average SCN \pm 95% confidence intervals. *** $p < 1e - 04$ in green and orange correspond to significant differences by post hoc Tukey HSD compared to N2 and CB4856, respectively. (B) Genotyping of NILs carrying genomic fragments of chr. III from CB4856 using genetic markers based on indels in CB4856 genome. Each row and column represents RIL line and its genetic composition at genetic markers, respectively. The location of the genetic markers on the chromosomes is shown above the graph. Green and orange tiles represents N2 and CB4856 genomes, respectively. The NILs carry varying fragment size of chr. III in the N2 background. Black-box in the middle highlights genomic region around ≈ 8.26 Mb $- \approx 8.58$ Mb on chr. II that is common to the NILs that convert N2-mutant to CB4856-mutant phenotype and absent in the NILs that do not. Therefore, this region may contain the causative genetic variation for differential expressivity of *egl-18(ga97)* mutation.

variation that makes it insensitive to germline RNAi (Paaby, A. G. White, et al., 2015). Loss of function due to deletion of single bp in *ppw-1* confers RNAi insensitivity in CB4856 (Tijsterman et al., 2002). Additional modifier loci have been discovered that act epistatically to *ppw-1*, and modify RNAi sensitivity (Pollard and Rockman, 2013; Paaby, A. G. White, et al., 2015). We overcame germline RNAi insensitivity of CB4856 by only performing somatic RNAi, i.e., by bleaching eggs and letting the hatched animals develop on dsRNA-expressing bacteria. In order to make sure there were no statistically significant differences in somatic RNAi between N2 and CB4856, we performed RNAi knockdown targeting GFP (Fig. 5.8). CB4856 was less sensitive to RNAi knockdown of *GFP* at a concentration of 10% V/V GFP dsRNA expressing bacteria mixed with 90% V/V empty-vector dsRNA expressing bacteria. However, no GFP-positive seam cells could be seen in N2 and CB4856 carrying *SCMp::GFP* upon RNAi knockdown of *GFP* at a concentration of 100% GFP dsRNA expressing bacteria. Therefore, to avoid differences in RNAi sensitivity between the strains we performed RNAi knockdown targeting gene-of-interest at a concentration of 100%. In the mutant condition, we counted SCN in F2 generation after bleaching P0 generation on gene-of-interest dsRNA expressing bacteria due to a technical reason — *egl-18(ga97)* animals grow asynchronously — to get enough animals at the right stage to count SCN. While germline RNAi is involved in this experiment, it does not pose a problem because MBA848 (mutant) contains N2 version of *ppw-1*, which makes it sensitive to RNAi.

In order to resolve the number of QTLs on chr. II, we counted SCN upon RNAi knockdown of 7 genes (*dgk-5*, *del-10*, *T28D9.1*, *abch-1*, *utp-20*, *wrn-1*, *C56C10.9*) with natural variation in the ≈ 0.14 Mb overlapping region around ≈ 6.48 Mb in the NILs (MBA846 and MBA944 in Fig.5.6) that converted N2-mutant phenotype to CB4856-mutant phenotype. RNAi knockdown of the 7 genes in the overlapping region did not affect SCN in wild-type or mutant animals suggesting that they are not likely to be involved in seam cell development (Fig. 5.9). We knocked down four additional genes (*dsh-2*, *egl-27*, *dnj-5* and *cutl-16*) by RNAi in both wild-type and mutant animals. We found that RNAi knockdown of *egl-27* and *dsh-2* decreased SCN in N2-mutant animals ($p < 2.2 \times 10^{-16}$). There was a decrease in SCN, albeit not significant in MBA848 (mut) animals upon knockdown of *egl-27* and *dsh-2*. The results taken together

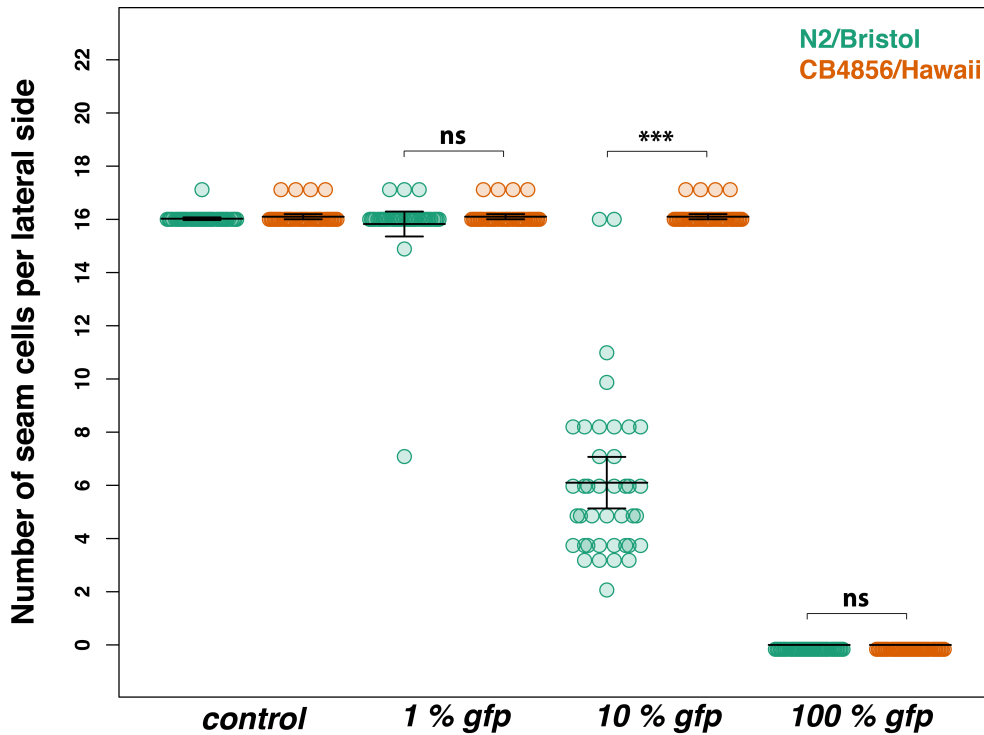


Figure 5.8: **Subtle differences in somatic RNAi between N2 and CB4856.** RNAi knock-down of *GFP* at various concentrations in N2, CB4856 and MBA840. CB4856 carries an introgressed *SCMp::GFP* from N2, while MBA840 carries an integrated *SCMp::GFP* marker. One-way ANOVA showed no statistically significant difference in SCN between the strains upon RNAi knockdown of 1% *GFP* compared to control ($F(1, 238) = 0.37$, $p = 0.55$). However, there was a significant effect of strain on SCN upon RNAi knockdown of 10% *GFP* ($F(2, 117) = 427.5$, $p < 2.2 \times 10^{-16}$). There were significant differences between SCN in N2 and two strains of CB4856 ($p < 2.2 \times 10^{-16}$). There was no difference between the two CB4856 strains (*MBA256* and *MBA840*, $p = 0.99$). $n = 40$. Error bars indicate average SCN \pm 95% confidence intervals. *** $p < 1 \times 10^{-4}$ corresponds to significant differences compared to ht115 (control) by post hoc Dunnett's multiple comparison test.

suggest the presence of two QTLs on chr. II and not a single QTL in the overlapping region. Further, *egl-27* and *dsh-2* have been shown to be involved in asymmetric seam cell divisions, therefore, they are candidate genes in the two non-overlapping genomic regions.

In order to narrow down candidates on chr. III, we counted SCN upon RNAi knockdown of 15 genes containing natural variation (*fbn-1*, *ncl-1*, *cdh-3*, *kle-2*, *clp-1*, *lin-36*, *mig-10*, *sor-1*, *hsp-110*, *ztf-30*, *sma-2*, *K02D10.1*, *zfp-1*, *lnkn-1* and *lin-9*) in the region between ≈ 7.66 Mb – ≈ 8.91 Mb in the NILs (MBA786, MBA819 and MBA945 in Fig.5.7) that converted N2-mutant phenotype to CB4856-mutant phenotype. We found that knockdown of most genes in wild-type animals did not cause a change in SCN (Fig. 5.10 and 5.11). Nevertheless, a few treatments affected SCN in both wild-type and mutant condition. Knockdown of *sor-1* increased SCN in wild-type N2 and CB4856, albeit significantly only in CB4856 ($p < 0.001$). In contrast to wild-type condition knockdown of *sor-1* caused a decrease in SCN in N2-mutant and MBA848 (mut) ($p < 0.001$). In a similar fashion, RNAi knockdown of *kle-2* significantly decreased SCN in both wild-type and mutant condition ($p < 2.21 \times 10^{-5}$). *sor-1* encodes a component of the polycomb repressive complex (PRC1) complex that is involved in development through global Hox gene repression and *kle-2*, an ortholog of human NCAPH2 (non-SMC condensin II complex subunit H2) is involved in mitotic sister chromatid segregation. *kle-2* is an important component of cell division and loss of *kle-2* may cause reduction in SCN due to seam cell division failure. Therefore, it is unlikely that natural variation in a core component of cell division underlies the differential expressivity of *egl-18(ga97)* between N2 and CB4856. *sor-1*, a polycomb group (PcG) gene that is involved in epigenetic silencing of Hox genes and for specifying cell identities along the anteroposterior axis represents a strong candidate gene.

Knockdown of some genes affected SCN only in wild-type condition — *zfp-1* is an ortholog of human MLLT10 histone lysine methyltransferase DOT1L cofactor and exhibits chromatin binding activity and is expressed in seam cells — knockdown of *zfp-1* caused a significant increase in SCN in both wild-type N2 and CB4856 ($p < 0.0012$) but did not affect SCN in mutant condition ($p > 0.33$). Therefore, *zfp-1* is an unlikely candidate. In contrast to *zfp-1*, a few genes decreased SCN only in mutant condition. Knockdown of *hsp-110* and *fbn-1* decreased SCN in both N2-mutant and MBA848 (mut) animals ($p < 0.04$). *hsp-110*, an ortholog of human

CHAPTER 5. MAPPING GENETIC VARIATION UNDERLYING DIFFERENCES IN EXPRESSIVITY OF THE *EGL-18(GA97)* MUTATION BETWEEN N2 AND CB4856

Figure 5.9: **RNAi screen to identify candidate genes in QTL on chromosome II.** (A) Seam cell number in wild-type N2 and CB4856 upon knockdown of candidate genes. There were no statistically significant differences in SCN between strains knockdown of genes on chr. II in wild-type N2 and CB4856 animals. (B) Seam cell number in mutant N2 and MBA848 (NIL) upon knockdown of candidate genes. Knockdown of most genes on chr. II did not have a significant effect on the SCN in mutant (*egl-18(ga97)*) animals in N2 and MBA848. Knockdown of two genes (*egl-27* and *dsh-2*) had a significant effect on the SCN in N2 and MBA848 ($F(4, 195) = 7.48$, $p = 1.26 \times 10^{-5}$; $F(4, 195) = 2.23$, $p = 0.03$, respectively). SCN decreased compared to control animals upon knockdown of *egl-27* and *dsh-2* in N2-mutant background ($p < 0.001$). In both A and B, Vertical black line in the graph distinguishes between different sets of experiments. One-way ANOVA was conducted separately for each strain (N2, CB4856 and MBA848) on different RNAi treatments within each experiment. $n = 40$ per strain. Error bars indicate average SCN \pm 95% confidence intervals. *** $p < 1 \times 10^{-4}$ corresponds to significant differences compared to control (ht115) RNAi by post hoc Dunnett's multiple comparison test. SCN data from RNAi in wild-type strains was generated in collaboration with Mingke Pan, a Master's student in the lab.

HSPA4 (heat shock protein family A (Hsp70) member 4) is a co-chaperone that is involved in the response to incorrectly folded protein and *fbn-1*, an ortholog of human FBN2 (fibrillin 2) is a component of apical extracellular matrix of epidermal sheath that is required for molting. *hsp-110* and *fbn-1* are candidate genes for the differential expressivity of *egl-18(ga97)* between N2 and CB4856 as they modify *egl-18(ga97)* phenotype. Knockdown of *lin-9* and *lin-36* caused a significant decrease in SCN in MBA848 (mut) background but not N2-mutant ($p < 0.02$). *lin-9*, an ortholog of human LIN9 (a tumor suppressor gene) is expressed in seam cells, and is involved in embryonic development. *lin-36* is expressed in several tissues including hyp7 and involved in digestive tract morphogenesis. *lin-9* and *lin-36* also represent candidate genes that underlie the higher SCN in CB4856-mutant compared to N2-mutant as they decrease SCN only in MBA848 (mutant) but not N2-mutant background. To summarise, RNAi knockdown of 8 of 15 genes tested had an effect on the SCN and most promising candidates are *sor-1*, *hsp-110*, *fbn-1*, *lin-9* and *lin-36*.

5.2.8 Prioritisation of candidate genes in QTLs

We sought to increase the resolution of QTL mapping to narrow down the candidate genes that may modulate the differences between N2-mutant and CB4856-mutant SCN. To this end, we genotyped NILs that converted the N2-mutant to CB4856-mutant phenotype and NILs that did not using SNP-based markers (listed in Appendix A.4) and snip-SNP-based markers (listed in Appendix A.5). This step also allowed us to confirm the natural variation present in these genes. On chr. II, we resolved the boundaries of MBA944 and MBA846, they had two mutually

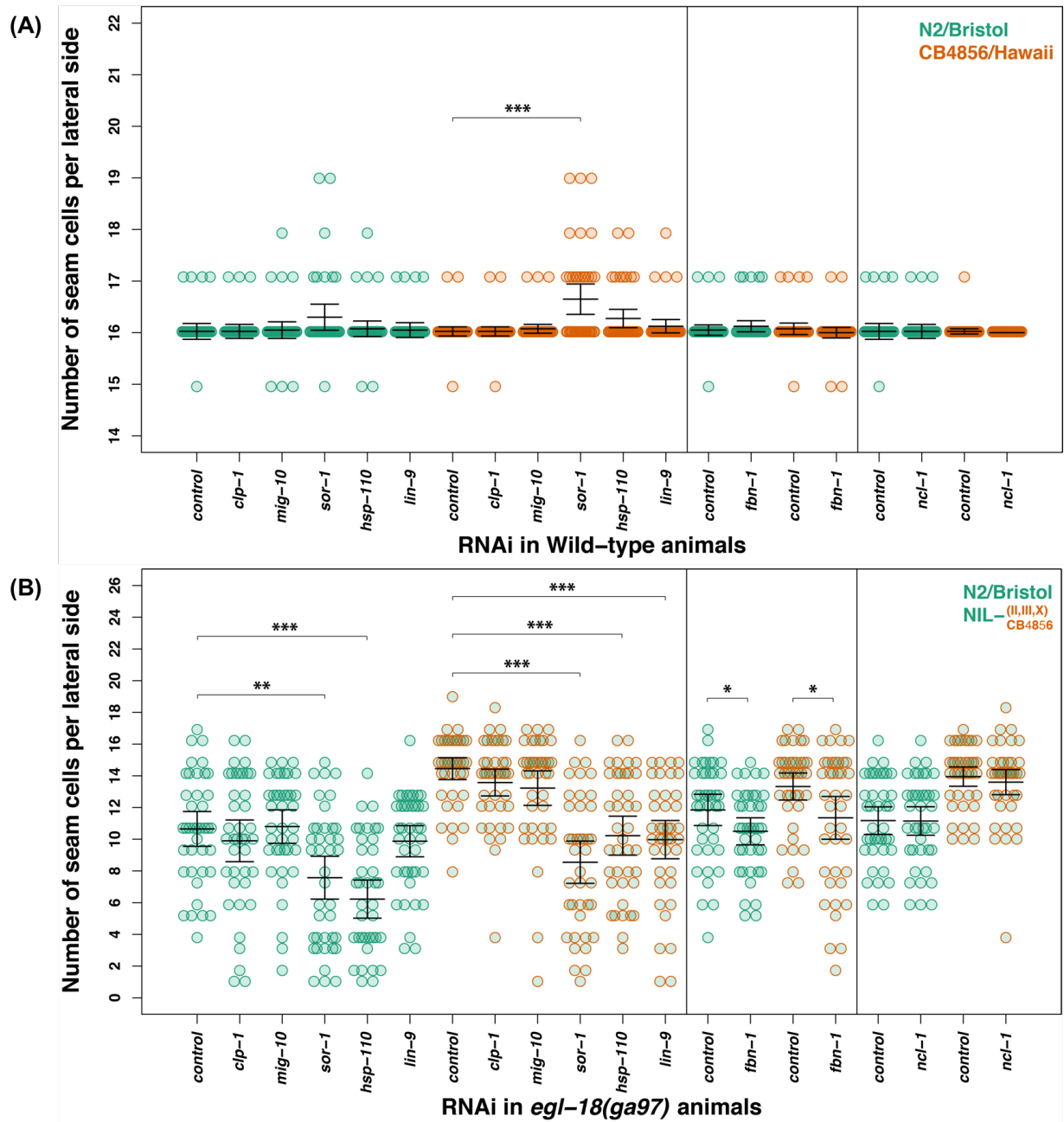


Figure 5.10: RNAi screen to identify candidate genes in QTL on chromosome III (first set). (A) Seam cell number in wild-type N2 and CB4856 upon knockdown of candidate genes. There were no statistically significant differences in SCN between strains knockdown of most genes on chr. III in wild-type N2 and CB4856 animals. Knockdown of *sor-1* had a significant effect on the SCN in CB4856 but not N2 ($F(5, 234) = 9.04$, $p = 7.17 \times 10^{-8}$; $F(5, 234) = 1.56$, $p = 0.17$, respectively). There was a increase in SCN upon the knockdown of *sor-1* compared to control in CB4856-wild-type animals ($p < 0.001$). Continued on next page.

CHAPTER 5. MAPPING GENETIC VARIATION UNDERLYING DIFFERENCES IN EXPRESSIVITY OF THE *EGL-18(GA97)* MUTATION BETWEEN N2 AND CB4856

Figure 5.10: **RNAi screen to identify candidate genes in QTL on chromosome III (first set).**

(B) Seam cell number in mutant N2 and MBA848 (NIL) upon knockdown of candidate genes. Knockdown of some genes on chr. III did not have a significant effect on the SCN in mutant (*egl-18(ga97)*) animals in N2 and MBA848. Knockdown of two genes (*sor-1* and *hsp-110*) had a significant effect on the SCN in N2 and MBA848 ($F(5, 234) = 10.48$, $p = 8.37 \times 10^{-9}$; $F(5, 234) = 19.65$, $p = 2.47 \times 10^{-16}$, respectively). SCN decreased upon knockdown of *sor-1* and *hsp-110* both N2 and MBA848 ($p < 0.001$). Knockdown of *lin-9* had a significant decrease in SCN only in MBA848 ($p < 0.001$). There was a decrease in SCN upon knockdown of *fbn-1* in both N2 and MBA848 ($F(1, 78) = 4.39$, $p = 0.04$; $F(1, 78) = 6.27$, $p = 0.01$, respectively). In both A and B, Vertical black line in the graph distinguishes between different sets of experiments. One-way ANOVA was conducted separately for each strain (N2, CB4856 and MBA848) on different RNAi treatments within each experiment. $n = 40$ per strain. Error bars indicate average SCN \pm 95 % confidence intervals. *** $p < 1 \times 10^{-4}$ corresponds to significant differences compared to control (ht115) RNAi by post hoc Dunnett's multiple comparison test. SCN data from RNAi in wild-type strains was generated in collaboration with Mingke Pan, a Master's student in the lab.

exclusive regions with a small overlap (Fig. 5.12A). I confirmed that *dsh-2* (exons 3 and 8) and *egl-27* (exons 11 and 12) carry multiple variants in CB4856. MBA944 carries natural variation (CB4856 version) in *dsh-2* but not *egl-27* and vice versa in MBA846. Further, all recombinants that convert N2-mutant to CB4856-mutant phenotype carry one or the other or both these two genes (CB4856 version). Mutation in *dsh-2* has been shown to increase SCN variability, by both gain and loss of seam cells (Baldwin, Clemons, and Phillips, 2016). Different mutations of *egl-27* have been shown to specifically affect T seam cell lineage resulting in the loss of T-seam cell (Herman et al., 1999) and result in higher SCN in embryos (Solari, Bateman, and Ahringer, 1999). In order to investigate if *egl-27* RNAi affects seam cells other than T during larval development, I counted SCN in F2 N2 (WT) animals upon *egl-27* RNAi. I found that there was an increase in SCN caused frequently by symmetrisation of T-cell and occasionally anterior cells (H1a, H1p, H2, V1a, V2a) upon *egl-27* knockdown (data not shown). Therefore, *dsh-2* and *egl-27* on chr.II independently may be responsible for the differential expressivity of *egl-18(ga97)* between N2 and CB4856.

SNP-marker based genotyping allowed us to exclude a few candidate genes from the RNAi screen by resolving the boundaries of the NILs containing fragments of chr. III from CB4856. For instance, two NILs (MBA964 and MBA1056) that failed to convert N2-mutant to CB4856-mutant both contained the CB4856 version of *fbn-1* which decreased SCN in mutant condition. Therefore, *fbn-1* could be excluded. Similarly, *lin-36* could be excluded by the fact that it is present in MBA1056, which does not convert N2-mutant to CB4856-mutant phenotype. *lin-9*

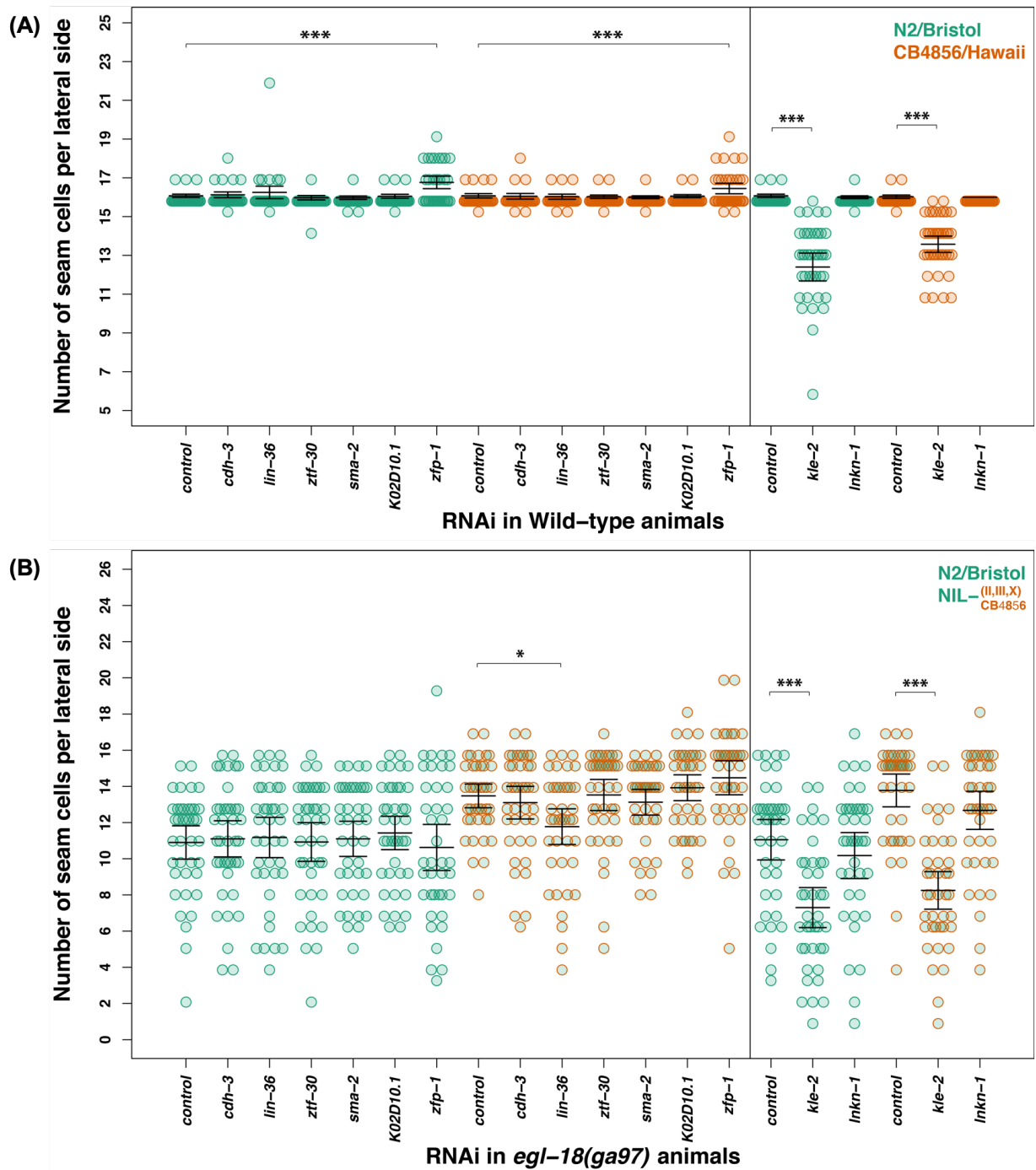


Figure 5.11: RNAi screen to identify candidate genes in QTL on chromosome III (second set). (A) Seam cell number in wild-type N2 and CB4856 upon knockdown of candidate genes. There were no statistically significant differences in SCN between strains knockdown of most genes on chr. III in wild-type N2 and CB4856 animals. Knockdown of *zfp-1* had a significant effect on the SCN in both N2 and CB4856 ($F(6, 267) = 8.01, p = 5.8 \times 10^{-8}$; $F(6, 267) = 4.95, p = 8 \times 10^{-5}$, respectively). SCN increased upon knockdown of *zfp-1* compared to control in wild-type animals ($p < 0.001$). Knockdown of *kle-2* significantly affected SCN in both N2 and CB4856 ($F(2, 117) = 102.03, p < 2.2 \times 10^{-16}$; $F(2, 117) = 131.02, p < 2.2 \times 10^{-16}$, respectively). There was a decrease in SCN upon knockdown of *kle-2* compared to control in wild-type animals ($p < 1 \times 10^{-10}$). Continued on next page.

CHAPTER 5. MAPPING GENETIC VARIATION UNDERLYING DIFFERENCES IN EXPRESSIVITY OF THE *EGL-18(GA97)* MUTATION BETWEEN N2 AND CB4856

Figure 5.11: **RNAi screen to identify candidate genes in QTL on chromosome III (second set)**. (B) Seam cell number in mutant N2 and MBA848 (NIL) upon knockdown of candidate genes. Knockdown of most genes on chr. III did not have a significant effect on the SCN in mutant (*egl-18(ga97)*) animals in N2 and MBA848. However, knockdown of *kle-2* had a significant effect on the SCN in N2 and MBA848 ($F(5, 234) = 10.48$, $p = 8.37 \times 10^{-9}$; $F(5, 234) = 19.65$, $p = 2.47 \times 10^{-16}$, respectively). RNAi against *kle-2* decreased SCN compared to control in both N2 and MBA848 ($p < 1 \times 10^{-10}$). Knockdown of *lin-36* had a significant effect on the SCN in MBA848 ($F(6, 273) = 4.17$, $p = 5 \times 10^{-4}$). SCN decreased slightly in MBA848 upon knockdown of *lin-36* ($p = 0.02$). In both A and B, Vertical black line in the graph distinguishes between different sets of experiments. One-way ANOVA was conducted separately for each strain (N2, CB4856 and MBA848) on different RNAi treatments within each experiment. $n = 40$ per strain. Error bars indicate average SCN \pm 95% confidence intervals. *** $p < 1 \times 10^{-4}$ corresponds to significant differences compared to control (ht115) RNAi by post hoc Dunnett's multiple comparison test. SCN data from RNAi in wild-type strains was generated in collaboration with Mingke Pan, a Master's student in the lab.

could be excluded because NILs (MBA786, MBA819 and MBA945) that convert the N2-mutant phenotype to CB4856-mutant do not carry the CB4856 version of *lin-9*. Interestingly, *sor-1*, *hsp-110* and *zfp-1* were common to NILs that convert N2-mutant to CB4856-mutant phenotype. *sor-1*, *hsp-110* and *zfp-1* have not been studied in the context of seam cell development. *zfp-1* is not a likely candidate gene as RNAi knockdown does not modify *egl-18(ga97)* phenotype. I confirmed the presence of a single non-synonymous polymorphism *sor-1* (exon 7) and an 3 bp in-frame deletion in *hsp-110* (exon 5) in CB4856. *hsp-110* is a chaperone that has been shown to suppress fibrilisation of Huntingtin (Htt) and the disaggregation and Htt fibrils in vitro (Scior et al., 2018). Therefore, *hsp-110* may be of global importance for maintaining protein homeostasis and may modify the mutant phenotype differentially between N2 and CB4856. *sor-1* is known to repress anterior expression of Hox genes *egl-5* and *mab-5* in seam cells and may be an important player in modifying SCN in *egl-18(ga97)* mutant. To investigate which seam cells respond to *sor-1* RNAi to increase SCN, I counted SCN in F2 N2 (WT) animals. There was a significant increase in SCN compared to control RNAi and there was frequent symmetrisation of seam cell divisions in the following seam cells: (V1a, V1p V2a), and occasionally mid/posterior cells (V3a, V3p, V4a, V4p, V5, V6p) upon *sor-1* knockdown (data not shown). Based on the literature and experimental evidence, *sor-1* and *hsp-110* on chr. III are most likely to underlie the difference in expressivity of *egl-18(ga97)* between N2 and CB4856.

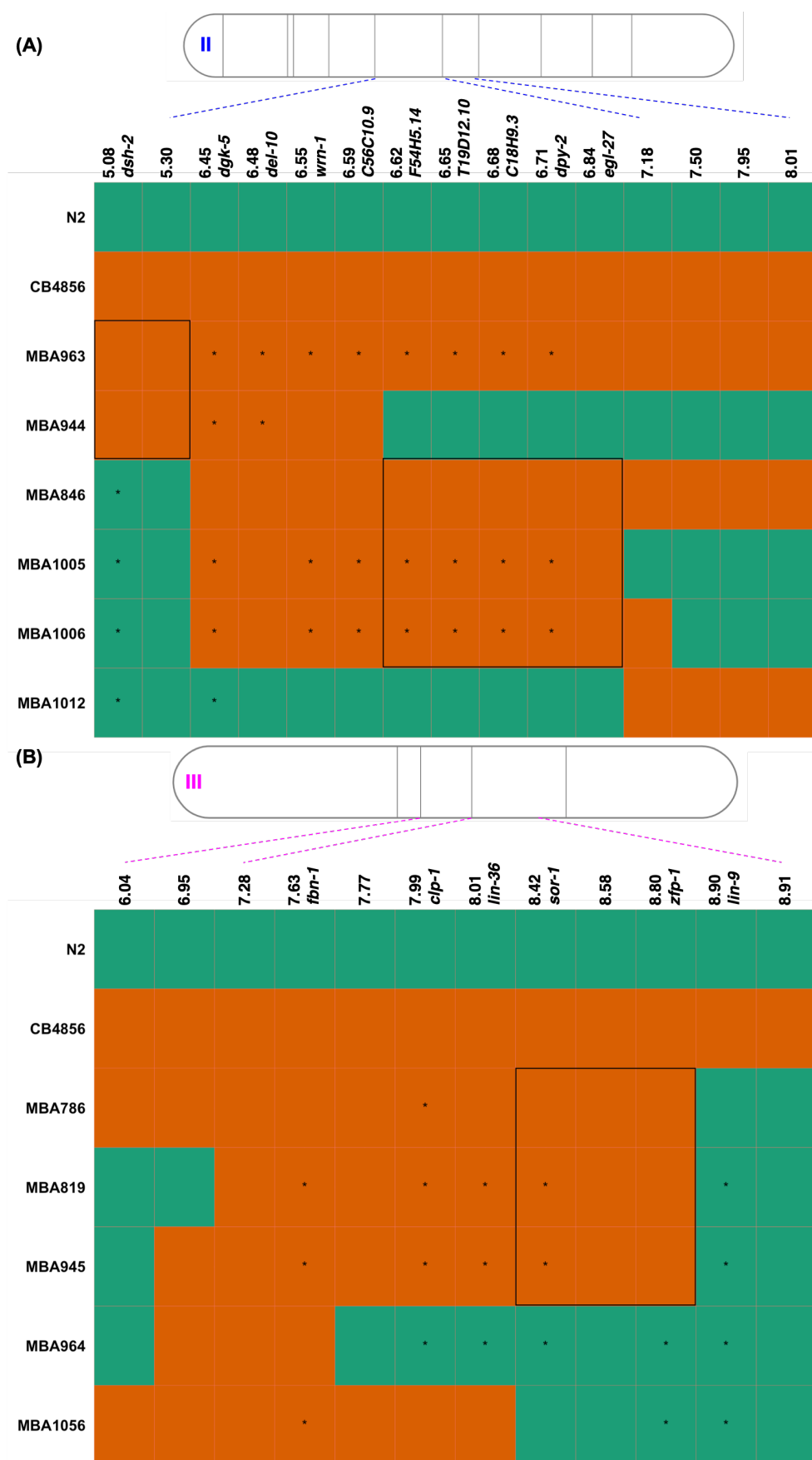


Figure 5.12: Genotyping of NILs carrying genomic fragments of chromosomes II and III from CB4856 using SNP-based genetic markers. Continued on next page.

Figure 5.12: **Genotyping of NILs carrying genomic fragments of chromosomes II and III from CB4856 using SNP-based genetic markers.** (A) Genotyping of NILs carrying genomic fragments of chr. II from CB4856. Black-boxes highlight genomic regions ≈ 6.62 Mb – ≈ 6.84 Mb and ≈ 5.08 Mb – ≈ 5.30 Mb on chr. II that is common to the NILs that convert N2-mutant to CB4856-mutant phenotype and absent in the NILs that do not. (B) Genotyping of NILs carrying genomic fragments of chr. II from CB4856. Black-box highlights genomic region ≈ 8.42 Mb – ≈ 8.80 Mb on chr. III that is common to the NILs that convert N2-mutant to CB4856-mutant phenotype and absent in the NILs that do not. In both A and B, each row and column represents RIL line and its genetic composition at genetic markers, respectively. The location of the genetic markers on the chromosomes is shown above the graph. Green and orange tiles represents N2 and CB4856 genomes, respectively. Black stars inside the tiles are interpolated/inferred. The region enclosed by the black-box may contain the causative genetic variation for differential expressivity of *egl-18(ga97)* mutation.

5.2.9 QTLs may modify the *egl-18(ga97)* phenotype in N2 and CB4856 through the Wnt pathway

egl-18 functions downstream of the Wnt/ β -catenin asymmetry ($W\beta A$) to specify seam cell fate during asymmetric seam cell divisions (Gorrepati, K. W. Thompson, and Eisenmann, 2013). In this paper, they found that animals have lower SCN in an activated Wnt background upon RNAi knockdown of *egl-18*. Therefore, we hypothesised there may be a difference in $W\beta A$ pathway in N2 and CB4856 that results in different *egl-18* loss-of-function phenotypes; specifically that N2 may have higher Wnt pathway activity compared to CB4856 resulting in a lower average SCN. To test possibility, we quantified overall Wnt pathway activity between wild-type N2 and CB4856 animals using POP-1 and HMG-helper optimal promoter (*POPHHOP*) reporter (Bhambhani et al., 2014). *POPHHOP* reporter is expressed in cells where $W\beta A$ signalling is activated. The strongest expression of the *POPHHOP* marker is observed in posterior muscle cells, intestinal cells and seam cells between the vulva and the tail (Fig. 5.13A). We imaged L4 animals that had completed seam cell divisions using confocal microscopy and quantified total fluorescence in Fiji (Fig. 5.13B). We did not find a consistent trend of Wnt pathway activity between N2 and CB4856 (Fig. 5.13B). We did not find any differences in qualitative measures such as total number of cells/seam cells/intestinal cells that expressed the *POPHHOP* marker between N2 and CB4856 (data not shown). These experiments suggest there may not be overall differences between $W\beta A$ pathway between N2 and CB4856.

We then performed somatic RNAi against genes (*lit-1/NLK*, *pop-1/TCF* and *apr-1/APC*) in $W\beta A$ pathway to test whether there are differences between N2 and CB4856. Seam cell

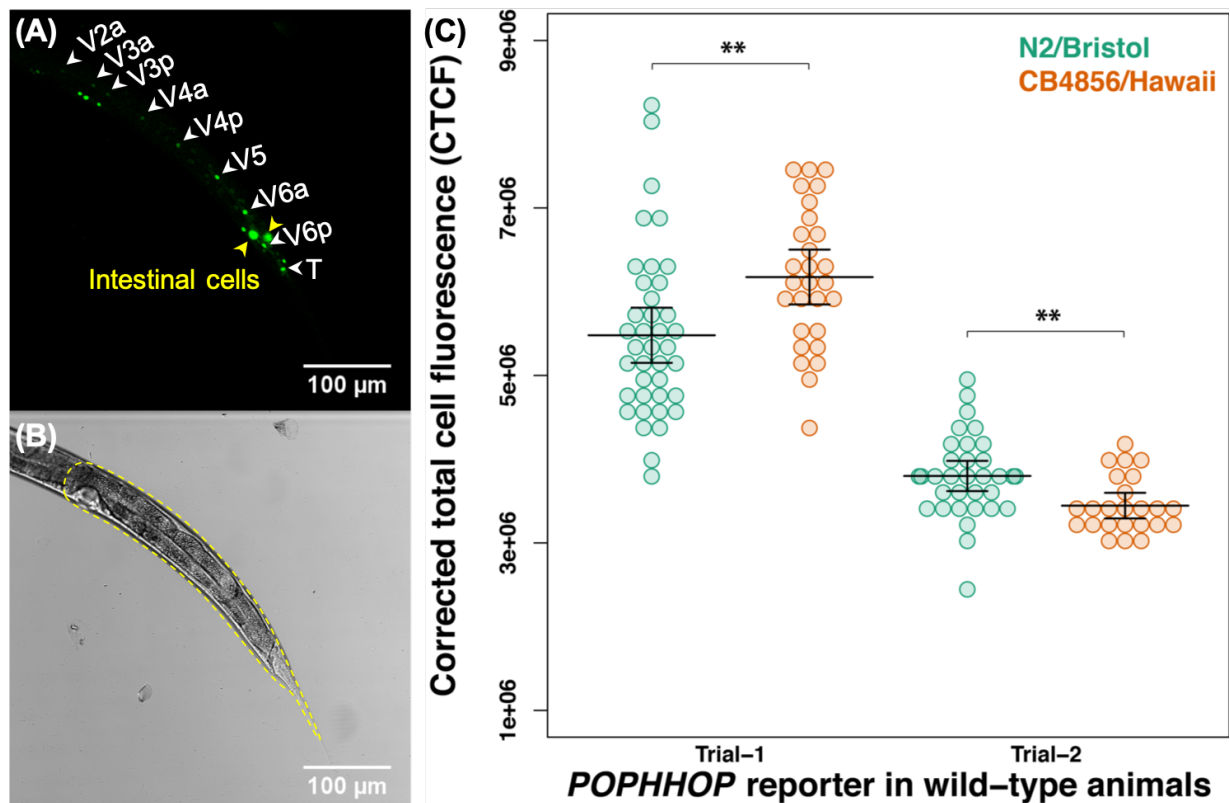


Figure 5.13: No differences in overall Wnt pathway activity between N2 and CB4856. (A) Representative wild-type (N2) animal at L4 stage carrying *POPHHOP* reporter. Note expression in seam cells (white arrowheads) and intestinal cells (yellow arrowheads). (B) Brightfield channel of the wild-type animal in (A). The region of interest in which the fluorescence intensity is quantified is highlighted in dotted yellow line. (C) Quantification of corrected total fluorescence (CTCF) in N2 and CB4856 wild-type animals carrying *POPHHOP* reporter. Two independent trials produced opposing trends. In trial-1, CTCF was significantly higher in CB4856 animals compared to N2 animals ($p = 0.0033$), and vice versa trial-2 ($p = 0.0033$). $24 \leq n \leq 39$ per strain. *** $p < 1 \times 10^{-4}$ corresponds to significant differences by Welch two-sample t-test test.

fate is associated with lower POP-1 levels. POP-1 is phosphorylated and exported out of the nucleus by LIT-1, which is bound and activated by WRM-1, a β -catenin. *apr-1* is a negative regulator of $W\beta A$ pathway that is localised to anterior cortex during seam cell division. Since N2 and CB4856 differ in sensitivity to RNAi, we performed RNAi against GFP as a control (Fig. 5.14A). We also utilised two additional strains of CB4856 (MBA840 and MBA841), which had integrated *SCMp::GFP* on different chromosomes as additional controls. We found that 10% *GFP* RNAi effectiveness was similar between N2 and CB4856 based on SCN ($p = 0.14$). However, 10% *GFP* RNAi was not effective in MBA840 and MBA841 ($p > 0.13$). To avoid differences in RNAi sensitivity, we used undiluted 100% RNAi against genes-of-interest. We found that RNAi against *lit-1* decreased SCN compared to control RNAi in N2, CB4856 and MBA841 ($p < 0.0046$) but not MBA840 ($p = 0.77$). There was a significant difference in SCN upon *lit-1* RNAi between N2 and CB4856/MBA840 ($p < 2.2e - 16$) suggesting that N2 may be sensitive to the loss of *lit-1*. However, there is a possibility that this is due to differences in RNAi sensitivity between N2 and CB4856.

We found that RNAi knockdown of *apr-1* and *pop-1* increased SCN in all strains compared to control RNAi (N2, CB4856, MBA840 and MBA841; $p < 0.0085$). Reduction of POP-1 levels or negative regulator APR-1 in the anterior daughter cell causes symmetrisation of seam cell fate and increases SCN. Surprisingly, SCN was significantly lower in N2 compared to CB4856/MBA840/MBA841 upon RNAi against *pop-1* ($p < 0.04$). There was also a significant difference in SCN between N2 and CB4856 upon knockdown of *apr-1* ($p < 0.0017$) but not MBA840/MBA841 ($p > 0.32$). Seam cells in CB4856 compared to N2 may be more prone to symmetrisation upon RNAi knockdown of *pop-1* and *apr-1*. These results are likely to reflect genuine differences in the $W\beta A$ pathway between N2 and CB4856 as we would expect the opposite trend in SCN given that CB4856 is slightly insensitive to RNAi.

To address the possibility that the QTLs affecting *egl-18(ga97)* phenotype identified in this study may be acting through the $W\beta A$ pathway, we performed RNAi knockdown of *pop-1* and *apr-1* in wild-type NILs (Fig. 5.14B). We found that SCN in NIL containing all three QTLs (II, III and X) was significantly higher compared to N2 ($p < 0.001$) but not to CB4856 ($p > 0.31$) upon knockdown of *pop-1* and *apr-1* suggesting that the QTLs affect $W\beta A$ pathway

activity. To investigate the individual contributions of QTLs to the *pop-1* RNAi phenotype in CB4856, we performed *pop-1* RNAi in NILs carrying one or two QTLs. Upon *pop-1* RNAi, NIL carrying any two of the three QTLs had significantly higher SCN compared to N2 but lower to CB4856 ($p < 0.05$; $p < 0.013$, respectively) suggesting an epistatic interaction between two QTLs. We found that SCN in NILs carrying a single QTL was not different to N2 ($p > 0.14$), but was significantly different to CB4856 ($p < 0.001$). Taken together, any two QTLs sufficiently recapitulate increase in SCN upon *pop-1* RNAi in CB4856 suggesting that the QTLs act additively to increase symmetrisation of seam cell fate in cell divisions. This may be one potential developmental mechanism by which CB4856 has higher SCN compared to N2 in *egl-18(ga97)* background.

5.3 Discussion

5.3.1 Towards understanding genetic basis of differential expressivity of *egl-18(ga97)* by a quantitative genetics approach

In this study, we discovered at least four quantitative trait loci (QTLs) that modify seam cell number (SCN) of *egl-18(ga97)* alone or in combination with each other using bulk segregant analysis. Three QTLs on chromosomes II, III and X contained genomic fragments from CB4856 and one QTL on chr. V contained genomic fragment from N2 increased SCN. We studied the effect of three QTLs containing CB4856 genomic fragments in N2 background already containing QTL on chr.V from N2 genome called near isogenic lines (NILs). QTLs on chromosomes II and III are major QTLs, which increased seam cell number (SCN) in *egl-18(ga97)* animals in N2 background. QTL on chr. X is a minor QTL, which works in combination with III to increase SCN in *egl-18(ga97)* animals in N2 background to that of CB4856. Finally, a combination of three QTLs on chromosomes II, III and X containing CB4856 genome was sufficient to convert SCN in N2-mutant animals to CB4856-mutant. Genotyping RILs indicated that QTL on chr. V may explain parts of the observed transgressive segregation. RILs with higher SCN than CB4856 were likely to carry N2 fragment on chr. V. However, we found that RILs that had lower SCN than N2 were equally likely to carry N2 or CB4856 fragment on chr. V suggesting negative epistasis between CB4856 fragments in QTLs on chromosomes III and X

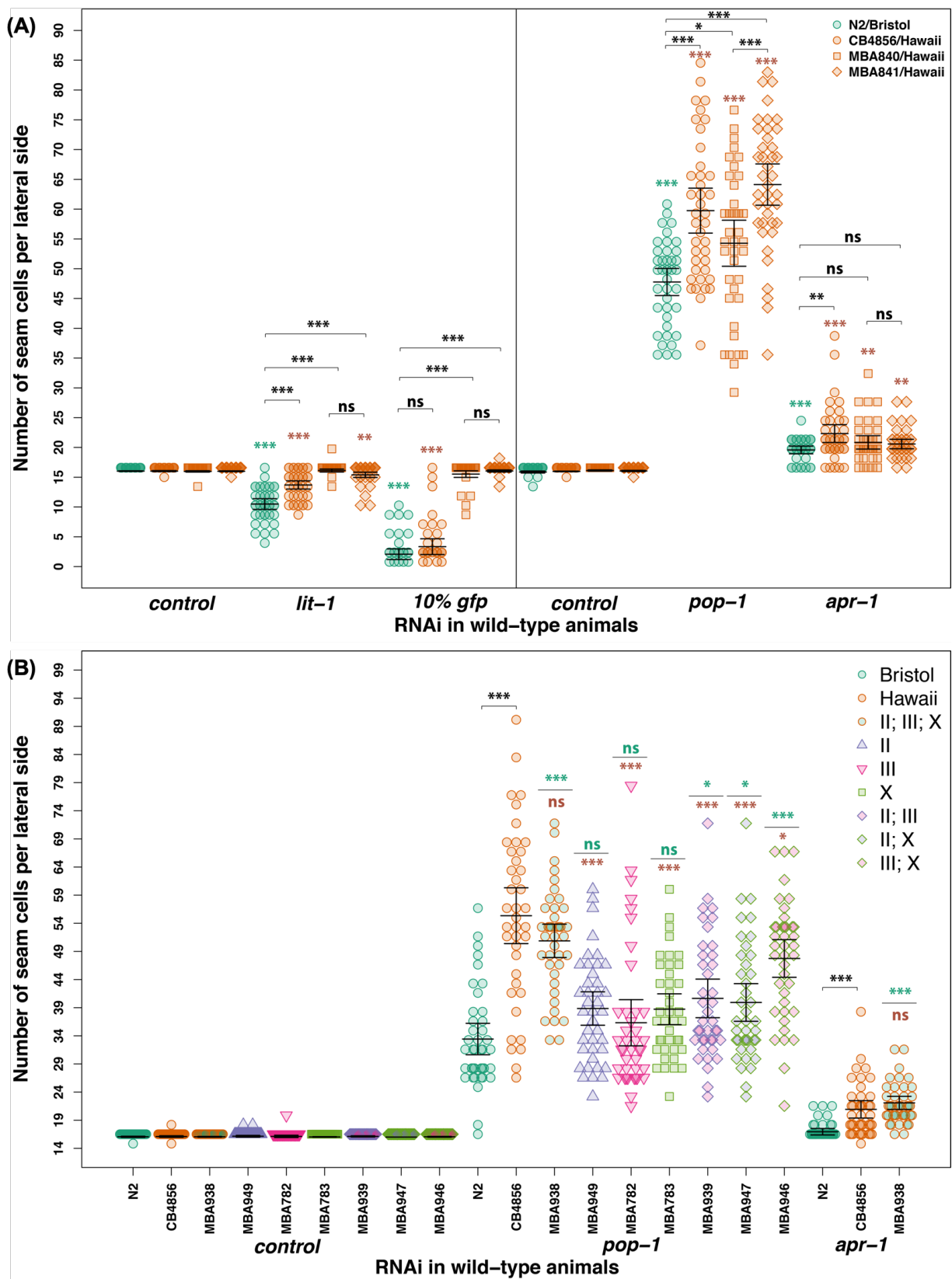


Figure 5.14: N2 and CB4856 show differences in SCN upon RNAi knockdown of Wnt components. Continued on next page.

CHAPTER 5. MAPPING GENETIC VARIATION UNDERLYING DIFFERENCES IN EXPRESSIVITY OF THE *EGL-18(GA97)* MUTATION BETWEEN N2 AND CB4856

Figure 5.14: **N2 and CB4856 show differences in SCN upon RNAi knockdown of Wnt components.** (A) Seam cell number in N2 and CB4856 strains upon knockdown of *NLK/lit-1*, 10% *GFP*, *TCF/pop-1* and *APC/apr-1*. Vertical black line in the graph distinguishes between two sets of experiments. CB4856 carries an introgressed *SCMp::GFP* marker from N2. MBA840 and MBA841 carrying two independently integrated versions of *SCMp::GFP* marker serve as additional controls. In the first experiment, there was a significant effect of *lit-1* and 10% *GFP* RNAi compared to control on SCN in N2, CB4856 and MBA841 ($F(2, 117) = 375.1$, $p = <2.2 \times 10^{-16}$; $F(2, 117) = 251.73$, $p = <2.2 \times 10^{-16}$ and $F(2, 117) = 7$, $p = 0.0013$ respectively). There was a significant difference in SCN between the strains upon *lit-1* and 10% *GFP* ($F(3, 156) = 66.68$, $p = <2.2 \times 10^{-16}$; $F(3, 156) = 325.73$, $p = <2.2 \times 10^{-16}$). In the second experiment, there was a significant effect of *pop-1* and *apr-1* RNAi compared to control on SCN in all strains (N2 $F(2, 117) = 666.02$, $p = <2.2 \times 10^{-16}$; CB4856 $F(2, 117) = 417.43$, $p = <2.2 \times 10^{-16}$, MBA840 $F(2, 117) = 326.7$, $p = <2.2 \times 10^{-16}$ and MBA841 $F(2, 117) = 683.67$, $p = <2.2 \times 10^{-16}$). There was a significant difference in SCN between the strains upon *pop-1* and *apr-1* RNAi ($F(3, 156) = 17.62$, $p = 6.67 \times 10^{-10}$; $F(3, 156) = 4.65$, $p = 0.0038$, respectively). One-way ANOVA was conducted separately for each strain (N2, CB4856, MBA840 and MBA841) on RNAi treatments comparing them to control within each experiment. Stars in green and orange bars correspond to statistical difference when compares to their corresponding ht115 controls.(B) *pop-1* compared to control RNAi increased SCN in all strains ($F(17, 702) = 141.41$, $p = <2.2 \times 10^{-16}$). There was a significant difference in SCN between the strains upon *pop-1* ($F(8, 351) = 17.85$, $p = <2.2 \times 10^{-16}$). *apr-1* compared to control RNAi increased SCN ($F(5, 234) = 44.83$, $p = <2.2 \times 10^{-16}$). There was a significant difference in SCN between the strains upon *apr-1* RNAi ($F(2, 117) = 21.81$, $p = 8.88 \times 10^{-9}$). One-way ANOVA was conducted together for all strains (N2, CB4856, MBA938, MBA949, MBA782, MBA783, MBA939, MBA947, MBA946) on RNAi treatments comparing them to control within each experiment. Green and orange stars correspond to differences with N2 and CB4856, respectively. In both A and B, One-way ANOVA was conducted for each RNAi treatment comparing strains within each experiment. $n = 40$ per strain. Error bars indicate average SCN \pm 95% confidence intervals. *** $p < 1 \times 10^{-4}$ corresponds to significant differences compared to corresponding control (ht115) RNAi by post hoc Dunnett's multiple comparison test or by post hoc Tukey HSD comparing SCN between strains upon RNAi treatment.

and N2 fragment in QTL on chr. V. Taken together, phenotypic analysis suggest a complex genetic architecture underlying seam cell number phenotype.

Due to the limited resolution of QTL mapping, the QTLs found contain large genomic intervals of CB4856 genome. *C. elegans* is a self-fertilising hermaphrodite, with only one crossover per chromosome pair in meiosis reducing the recombination between genomes, which is required for higher resolution of QTLs. Reciprocal crosses followed by multiple rounds of random mating in the F2 generation and advanced breeding designs in the generation of RILs would increase the resolution of QTL mapping (Rockman and Kruglyak, 2008; Burga, Ben-David, et al., 2019). However, the lack of functional vulva and sickly nature of *egl-18* animals were an impediment to our crossing design.

In order to increase resolution of used classical genetics to break the large genomic interval of QTLs, we narrowed down the genomic interval for the two major QTLs by analysing SCN in

NILs containing the smaller genomic fragments on chromosomes II and III. We discovered that there are two independent QTLs on chr. II, one between ≈ 4.78 Mb – ≈ 6.62 Mb (containing 112 genes with natural variation) and the another between ≈ 6.44 Mb – ≈ 7.18 Mb (containing 53 genes with natural variation). The presence of two QTLs on chr.II might also explain the large genomic interval found in QTL detected in our initial RILs. The genomic interval for the QTL on chr. III containing causative natural variation is between ≈ 8.01 Mb – ≈ 8.90 Mb (containing 38 genes with natural variation).

The ultimate goal of quantitative genetics is to quantify the genotypic contributions of individual genes to a phenotype. Prioritising candidate genes in a QTL based on the type of natural variation or its impact on the function of protein is not straightforward because both changes in expression or function may explain the differences in phenotype. In an analysis of the impact of SNPs associated with complex human disease on protein function, their distribution of amino acid substitution scores was indistinguishable from distribution of normal human variation (Thomas and Kejariwal, 2004). There is not yet a clear consensus on the type of natural variation that is most likely to modify phenotypes in *C. elegans*. In one instance, Frézal et al. (2018) discovered that a deletion in *set-24* gene, an ortholog of human KMT2E (lysine methyltransferase 2E) was the genetic basis of differences in Mrt phenotype in wild-isolates MY16 and JU1395, while Duveau and Félix (2012) found a non-synonymous polymorphism in a RNA cytidine acetyltransferase gene *nath-10* modulated differences in vulval induction index in N2 and AB1 strains. In addition to the natural variation in protein coding sequences of genes, there could be polymorphisms in regulatory regions of the candidate genes which affect the *egl-18(ga97)* phenotype.

Going from QTLs to individual genes is difficult using genotype-by-phenotype analysis especially when thousands of genes are involved in modifying the phenotype-of-interest. It is not yet clear if a single gene in the interval or multiple genes acting epistatically in the genomic interval of QTLs are responsible for difference in SCN of *egl-18(ga97)* animals in N2 and CB4856 background. Fortunately, CRISPR-Cas9 has revolutionised gene-editing technology and made it possible to directly test the effect of specific changes on the phenotype once a set of candidate genes are identified. CRISPR-Cas9 genome-editing could be employed to create specific

nucleotide changes in candidate genes in N2-mutant background and vice versa to pin down the molecular nature of the QTLs. Higher genome-editing efficiency has been reported with CRISPR-Cas9 ribonucleoprotein complexes (Paix, Folkmann, Rasoloson, et al., 2015). Use of single stranded DNA oligos with short homology arms (30 bp – 60 bp) were found to be sufficient for precise single nucleotide edits (Prior et al., 2017) and smaller inserts (Paix, Folkmann, and Seydoux, 2017). Further, nucleotide changes at different loci can be produced together in a combinatorial fashion facilitating high-throughput genome-editing (Paix, Folkmann, and Seydoux, 2017; Dokshin et al., 2018).

We performed RNAi against candidate genes in the genetic interval and discovered a new role as modulators of SCN for many genes on chr. III (*fbn-1*, *lin-9*, *kle-2*, *zfp-1*, *lin-36*, *sor-1* and *hsp-110*). *fbn-1* is a putative Wnt target as it was found to be significantly upregulated upon Wnt over-activation (Gorrepati and Eisenmann, 2015). A genome-wide RNAi screen for regulators of SCN did not find these genes that were picked up by our targeted RNAi screen (Hughes et al., 2013). Two pleiotropic genes (*C34F11.1* and *T07F8.1*) were identified to increase SCN by their RNAi screen that are present in the genomic interval of QTLs on chr. II and also contain natural genetic variation, which were not tested in this study.

5.3.2 *dsh-2* and *egl-27* are potential candidate genes on chromosome II affecting *egl-18(ga97)* expressivity

There are two QTLs on chr. II, left-QTL between ≈ 4.78 Mb – ≈ 6.62 Mb and right-QTL ≈ 6.44 Mb – ≈ 7.18 Mb. Left-QTL contains 112 genes with natural variation. Amongst these genes, *dsh-1* and *dsh-2* are candidate genes based on the annotation in wormbase. *dsh-1* and *dsh-2* are both orthologs of human DVL (dishevelled segment polarity protein) 2 and 3, and they are involved in the Wnt signalling pathway. There are three dishevelled proteins (MIG-5, DSH-1 and DSH-2) in *C. elegans*. *dsh-2* and *mig-5* have been shown to act redundantly to regulate seam cell fate by negative regulation of nuclear WRM-1 (β -catenin) and positive regulation of nuclear SYS-1 (β -catenin), while *dsh-1* is not considered a major contributor (Baldwin, Clemons, and Phillips, 2016). *cdc-14* is another promising candidate gene for the left-QTL, which have not been tested in this study. *cdc-14*, an ortholog of human CDC14A (cell division cycle 14A) is a phosphatase that is known to be expressed in seam cells (S. H.

Roy et al., 2011). It is a negative regulator of cell cycle; loss of *cdc-14* causes extra divisions in VPCs and intestine (Saito et al., 2004). A mutation in *cdc-14* in CB4856 may cause extra divisions to compensate for the loss of seam cells in *egl-18(ga97)* in CB4856.

egl-27 represents a strong candidate gene amongst the 53 genes with natural variation in the right-QTL on chr. II. *egl-27*, an ortholog of human RERE (arginine-glutamic acid dipeptide repeats) is a highly pleiotropic gene that is expressed in all somatic cells (Solari, Bateman, and Ahringer, 1999). *egl-27* mutants are egg-laying defective and have abnormal male tail morphology. It has been shown to affect T seam cell polarity where *egl-27* loss-of-function leads to symmetrisation towards the *hyp7* fate and loss of seam cell fate (Herman et al., 1999). Contrarily, *egl-27* loss-of-function mutants have been shown to have higher SCN as compared to wild-type (Solari, Bateman, and Ahringer, 1999). I have observed occasionally doublets of T and other anterior seam cells upon strong RNAi knockdown of *egl-27*, suggesting that it can cause symmetrisation towards seam cell fate as well, thus increasing SCN. Therefore, *egl-27* may be required to both maintain seam cell fate and repress symmetrisation of seam cell divisions.

A small number of chromatin remodelling components have been identified as hub genes in a *C. elegans* genetic interaction network (Lehner et al., 2006). The loss of these genes was found to enhance phenotypes of mutations in unrelated pathways. *egl-27* was found to be a hub gene, which encodes a subunit of the nucleosome remodeling and histone deacetylase (NURD) complex. However, these results are with RNAi knockdown of *egl-27* in N2 and not genetic evidence in other isolates. It is still possible that *egl-27* buffers *egl-18* loss-of-function phenotypes as a global genetic modifiers like *hsp-70* but in a differential fashion in N2 and CB4856. *egl-27* has been shown to affect asymmetric cell division of T seam cell which is governed by the Wnt signalling pathway (Herman et al., 1999). Asymmetric cell divisions most likely involve *egl-27*-mediated chromatin reprogramming to produce associated gene expression changes during cell fate transitions. However, it is not clear if *egl-27* interacts with Wnt signalling during asymmetric cell division and remains to be explored.

5.3.3 Novel role for *sor-1* and *hsp-110* for modifying seam cell number in an *egl-18* mutant background

The narrowed QTL on chr. III is between ≈ 8.01 Mb – ≈ 8.90 Mb. *sor-1* and *hsp-110* are the most promising candidate genes amongst 38 genes in the QTL region with natural variation. *sor-1* is a part of polycomb repressor complex (PRC1), which is involved in global repression of homeobox (Hox) genes (T. Zhang et al., 2006). Proper spatial pattern of Hox gene expression is crucial as patterns of expression of Hox genes along the anteroposterior axis direct cell fates in development of animals. *sor-1* controls the expression of two hox genes *mab-5* and *egl-5*, which are orthologous to *Drosophila* Antennapedia (*Antp*) and Abdominal-B (*Abd-B*), respectively. *mab-5* is expressed in the posterior part of *C. elegans*, in V5 and V6 seam cells. *egl-5* is expressed posterior to *mab-5*, in V6 seam cell. In *C. elegans* males, V5, V6 and T give rise to rays during postembryonic development.

Role of *mab-5* and *egl-5* has been studied in seam cell development in males but their role in seam cells in hermaphrodites is unclear. Ectopic expression of *mab-5* in anterior V1 – V4 has been observed in *lin-22* mutants (L. A. Wrischnik and C. J. Kenyon, 1997; Katsanos et al., 2017) but phenotypic consequences of *mab-5* in hermaphrodite were not studied. Loss of function of *sor-1* leads to ectopic expression of *mab-5* and *egl-5* in the anterior body, which leads to homeotic transformations of anterior cells to produce rays in males (T. Zhang et al., 2006). Unpublished data in the lab shows that a gain-of-function allele of *mab-5* that leads to *mab-5* expression in the anterior V cells and symmetrisation of seam cell divisions phenocopies *sor-1* RNAi phenotype. Thus, inappropriate expression of *mab-5* in anterior seam cells causes symmetrisation of V seam cells in hermaphrodites. Thus, we discovered a novel role for *sor-1* in hermaphroditic seam cell development as a suppressor of symmetrisation of seam cell divisions like *egl-27*.

hsp-110 has not been studied in the context of seam cell development. *hsp-110*, an ortholog of human HSPA4 (heat shock protein family A (Hsp70) member 4) is a co-chaperone that is involved in protein homeostasis. *hsp-110* is required for protein aggregate solubilisation in vivo (Rampelt et al., 2012), and to suppress fibrilisation of Huntingtin (Htt) and the disaggregation of Htt fibrils in vitro (Scior et al., 2018). Molecular chaperones like *Hsp90* are known to

act as phenotypic capacitors that reduce variation in traits (Rutherford and Lindquist, 1998; Queitsch, Carlson, and Girirajan, 2012). Previous work in the lab has shown that RNAi knockdown of *hsp-90* leads to variable SCN suggesting a conserved role for *hsp-90* as a capacitor for phenotypes (Katsanos et al., 2017). Knockdown of *hsp-110*, albeit in a sensitised mutant background reduces SCN in both N2 and the NIL background. Therefore, *hsp-110* may be involved in buffering *C. elegans* SCN in sensitised background, but the mechanism needs to be investigated.

Chapter 6

General Discussion

6.1 The fusogen EFF-1 contributes to robustness of seam cell patterning

This work implements a new experimental framework to study mechanisms of developmental robustness in *Caenorhabditis elegans*. Specifically, I was interested in identifying genes that confer robustness of seam cell patterning to intercellular or intracellular stochastic noise. In the third chapter, I describe the results from an unbiased phenotypic variance-based (not mean-based) mutagenesis screen to identify genes involved in buffering seam cell number (SCN), a phenotype I used throughout this thesis as a proxy of seam cell patterning accuracy. I recovered a nonsense mutation in the fusogen gene *eff-1*, which is required for the fusion of the anterior seam cell daughters to the hyp7 syncytium. In addition to this, I discovered a novel role for *eff-1* in buffering SCN by showing that mutations in *eff-1* lead to increased SCN variability.

An outstanding question in the field is what type of genes and molecular mechanisms confer robustness to a developmental system. There is experimental evidence that highly connected genes (also known as network hubs) are important for developmental robustness. *Hsp90*, a gene encoding a molecular chaperone is one the most well-characterised examples of a network hub that buffers phenotypic variation. When its function is impaired, it leads to increase in phenotypic variability in the development of various organisms like *Drosophila melanogaster*, *Arabidopsis thaliana* and *Caenorhabditis elegans* (Rutherford and Lindquist, 1998; Queitsch, Sangster, and Susan Lindquist, 2002; Katsanos et al., 2017). In addition, a high-throughput study in *Saccharomyces cerevisiae* identified ≈ 300 genes, which upon deletion increased cellular morphological variation to stochastic and microenvironmental variation. These genes were involved in core cellular processes like maintenance of chromosome organisation, DNA integrity and cell cycle and were highly connected in genetic or protein-protein networks (Levy and Siegal, 2008). Surprisingly, I discovered *eff-1*, which is not a hub but a core component in the seam cell gene network to be important for the robustness of SCN in *C. elegans*. *eff-1* is a nematode specific fusogen, which is an essential component in the fusion of anterior seam daughters to the hyp7 syncytium. This is consistent with previous results from a similar screen in the lab that identified *lin-22*, a Hes-related basic helix-loop-helix (bHLH) transcription factor

as a modulator of SCN variance (Katsanos et al., 2017). The *lin-22* transcription factor was shown to act in a cell autonomous manner in the seam and not in a systemic fashion to secure wild-type seam cell patterning. *lin-22* mutants show cell-to-cell variability in Wnt pathway activation, which is thought to drive the SCN variability. Therefore, it would be interesting in the future to assess SCN variability when two of the robustness conferring genes (*lin-22* and *eff-1*) are eliminated, which is the case in a double mutant combination (Katsanos et al., 2017). Intracellular processes involving microRNAs, transcription factors and genetic redundancies in Wnt signalling may also contribute to developmental robustness. Importantly, our seam cell screens are far from saturated, which is also supported by the serendipitous discovery of many new seam cell regulators in chapter 5 while fine mapping the QTLs using an RNAi approach. Together, these results indicate that core components of the seam cell gene network influence developmental variance in this system and more regulators are yet to be discovered.

I found that *eff-1* mutants show SCN variability without changes to mean SCN stemming from random gains and losses of seam cell fate. Gains were explained through defects in anterior cells that retain the seam cell fate post division. Losses were found to be via stochastic terminal differentiation of posterior cells. I show that *eff-1* is not expressed in posterior seam cells that contribute to SCN, therefore, *eff-1* mutations may cause SCN variability acting both autonomously and non cell-autonomously. Local cell-cell communication is essential for cell polarity, proliferation and tissue level homeostasis in development. I propose that such non-cell autonomous events may be mediated by the disrupted physical cell-cell communication in the seam tissue in the *eff-1* mutants. Interestingly, cell-to-cell communication may be important not only for the nematode invariant lineage, but has also been shown to drive invariant embryonic development in ascidians (Guignard et al., 2017).

Robustness of SCN may therefore rely not only on intracellular but also intercellular processes. Seam cells reconnect to each other after every round of cell division and reconstitute apical junctions before the next round of cell divisions. Contact between seam cells through apical junctions is thought to be essential for the stereotyped seam cell patterning and proliferation (Austin and C. Kenyon, 1994; Silhánková, Jindra, and Asahina, 2005). EFF-1 facilitates appropriate cell-cell contacts between seam cells by timely fusion of the anterior seam daugh-

ters. The role of cell contact is exemplified by studying the development of the V5 cell lineage upon laser ablation of its neighbouring cells. The asymmetric cell division (neuroblast-seam) of V5, becomes symmetrised (seam-seam) upon the loss of two or more anterior (V2 – V4) or posterior (V6) cells (Austin and C. Kenyon, 1994). However, it is not clear to what extent each of the seam cells is affected by the loss of contact. A key prediction from my work is that such interactions must be common leading to cell compensation or loss of divisions. My preliminary attempts in the lab to ablate H1/H2 seam cells in L1 animals and assess the phenotypic consequences in their neighbours were unsuccessful, because ablated animals became developmentally arrested at L1 stage. Alternate strategies in the future should involve disrupting cell-cell contacts by knocking down apical junctional components or performing genetic cell ablations and study their influence on seam cell patterning.

6.2 Cell differentiation is uncoupled from cell fusion in the seam

It was previously thought that fusion equals cell differentiation in the seam or that is at least part of the differentiation programme. This observation was based on previous evidence which suggested that fusion is required for differentiation of anterior seam daughters (Brabin, Appleford, and Woollard, 2011; Brabin and Woollard, 2012). However, in the third chapter of this thesis, I present experimental evidence the differentiation programme of the seam daughters may act in parallel to cell fusion. While the cell fusion driven by EFF-1 is essential for the robustness of seam cell patterning, anterior seam cell daughters in its absence continue to differentiate by expressing *hyp7* markers and stop expressing seam cell markers. Vulval precursor cells also adopt a vulval cell fate to form ectopic vulva by responding to neighbouring signals in the absence of cell fusion (Mohler et al., 2002). Previous arguments on the fate of the non-fused cells relied on an apical junction marker. However, apical junction is broken down by EFF-1 during fusion and is not a bona fide seam cell marker as anterior daughters differentiate despite maintaining apical junctions in *eff-1* mutants. The dissolution of apical junctions by *eff-1* in differentiating anterior seam daughters is necessary for limiting cell migration paths for these cells as proposed by Shemer and Podbilewicz (2003). In wild-type worms, anterior seam cell

daughters differentiate, and fuse to the hyp7 syncytium. In contrast, anterior seam cell daughters do not fuse to hyp7 (hyp7 is not formed) in *eff-1* mutants, and they migrate throughout the hypodermis (Mohler et al., 2002). This study resolves a fundamental question in the field by demonstrating that fusion is largely not required for cell differentiation. Nevertheless, some seam divisions, especially in late post-embryonic division, are aberrant which indicates that in the context of a whole organism fusion can be linked to cell fate acquisition.

6.3 Cryptic genetic variation influences seam cell development

Work presented in Chapter 4 of this thesis shows that seam cell number is robust to standing genetic variation and stochastic noise as evidenced by low variance of SCN in a number of divergent wild isolates. Nevertheless, robustness in developmental systems allows accumulation of cryptic genetic variation. This study discovered for the first time CGV affecting seam cell development by studying the effect of genetic variation on SCN upon environmental (temperature increase) and genetic (mutation or RNAi) perturbations.

First, upon environmental perturbation, when wild-type N2 animals developed at 25 °C as opposed to the standard growth temperature of 20 °C, there was an overall increase in SCN due to symmetrisation of asymmetric cell divisions in specific seam cell lineages (V6a, V5, V2a and V1a). Curiously, anterior V cell daughters compared to posterior were more sensitive to temperature increase at 25 °C, evidenced by the frequent seam cell division symmetrisation events observed. The presence of CGV was validated by the discovery of a $G \times E$ interaction, wherein the genetic background influenced the frequency and seam cell position of the symmetrisation event. Notably, the polymorphic CB4856 strain from Honolulu from Hawaii differed from all other strains in completely suppressing frequency of division symmetrisation at 25 °C. In contrast, the XZ1516 strain from a neighbouring island Kekaha from Hawaii, which is the most divergent *C. elegans* strain isolated so far, showed an increase in the frequency of division symmetrisation at 25 °C (Fig. 1.3 and Fig. 1.3). This suggests that genetic variation can both enhance and suppress seam cell division symmetrisation. It also indicates that strains from Hawaii are not necessarily more robust to temperature increase than the Bristol reference

strain, as potentially expected due to difference in the temperature in these two geographic locations. The genetic basis explaining the difference in frequency of symmetrisation remains unknown but can be mapped using a similar quantitative genetics approach to the one presented in this thesis. To facilitate this mapping endeavour, I propose that the frequency of symmetrisation of the V6a cell division at 25 °C would be the best phenotype to focus on, as it shows the greatest difference among strains. It will be intriguing to find out if the same or different genetic loci that explain the differences in frequency of V6a cell division symmetrisation between N2 and CB4856 also underlie the increase in the XZ1516 strain.

Mutations in the GATA transcription factor *egl-18* also revealed the presence of CGV affecting the performance of the seam cell gene network. To characterise the genetic architecture of cryptic genetic variation, I mapped the genetic variation underlying differences in the expressivity of *egl-18* mutations between N2 and CB4856 using a quantitative genetics approach, as presented in Chapter 5. I found that multiple quantitative trait loci (QTLs) on four of the six *C. elegans* chromosomes were likely causing the difference in SCN between N2 and CB4856 carrying *egl-18(ga97)*. QTLs on different chromosomes acted epistatically to modify *egl-18* mutant phenotype suggesting that a complex genetic architecture underlies the seam cell development.

During the course of my experiments, I also discovered that RNAi could be used as a perturbation to reveal CGV. While N2 and CB4856 may not differ in overall Wnt pathway activity, they still differ in their sensitivity to loss of Wnt pathway components by RNAi. SCN is higher in CB4856 compared to N2 upon RNAi knockdown of *pop-1*/TCF, *lit-1*/NLK and *apr-1*/APC. Interestingly, QTLs that were found to modify SCN in *egl-18* mutants also modified SCN upon RNAi knockdown of Wnt pathway components. This suggests that the QTLs discovered in this study may act as broader modifiers of the Wnt signalling pathway. Therefore, an overarching question remains whether these QTLs may also influence the division symmetrisation of seam cells at 25 °C.

6.4 The genetic basis for the differential expressivity of *egl-18* mutations between N2 and CB4856

There are two plausible genetic mechanisms for the differences in the expressivity of *egl-18* mutations between N2 and CB4856. The first mechanism is direct compensation by a gene paralogue or another gene with redundant function. Many experimental studies in *S. cerevisiae* and *C. elegans* have shown that loss-of-function of genes with a duplicated gene is less severe compared to genes without a duplicate (Gu et al., 2003; Conant and Wagner, 2004; Tischler et al., 2006). Indeed, *egl-18* has a gene paralogue (*elt-6*) and the SCN in RNAi knockdown of *egl-18* alone is less severe (higher SCN) compared to a knockdown of both *egl-18* and its paralogue *elt-6* (Gorrepati, K. W. Thompson, and Eisenmann, 2013). However, I believe it is unlikely that this mechanism accounts for the difference in the phenotypic outcome of *egl-18* mutations between isolates. The *egl-18* and *elt-6* homologues are located one next to the other in the genome thus wild-type *elt-6* from N2 was introgressed together with the *egl-18* mutation in CB4856 and I did not find differences in the level of its paralogue *elt-6* between N2 and CB4856. Therefore, while the *egl-18* loss-of-function phenotype depends on *elt-6*, the difference in mutation expressivity between N2 and CB4856 cannot be explained by functional compensation by *elt-6* due to changes in expression levels or protein function.

The second mechanism to explain differences in mutation expressivity is through independent modifiers in the genome. These modifiers could be for example molecular chaperones such as *Hsp90* (Burga, Casanueva, and Lehner, 2011; Casanueva, Burga, and Lehner, 2012) or chromatin regulators (Tischler et al., 2006), which may act to buffer developmental phenotypes. In support of this, I have discovered several QTLs that seem to modify the *egl-18* phenotype. It is also of note that some of the most promising candidate genes within the QTL intervals on chromosomes II and III are chromatin modifiers (*sor-1* and *egl-27*) or protein chaperones (*hsp-110*). Although, these candidates are yet to be validated, it is tempting to speculate that there may be additional chromatin modifiers or chaperones in the genomic QTL interval on chromosome X. Consistent with this prediction, unpublished data in the lab from an RNAi screen against chromatin regulators in *C. elegans* revealed two genes, *set-30* and *C52B9.8*,

within the same interval on chromosome X (≈ 3 Mb – ≈ 5.08 Mb) that influence SCN. Given that these genes harbour genetic variation in CB4856, they are promising candidate genes for the QTL on chromosome X.

The next important step is to discover the molecular nature of QTLs by validating the causative genes in these chromosomal intervals. CRISPR-Cas9 genome editing will be the approach to follow, as we can now replace exact nucleotides and assess their consequence for a particular developmental phenotype of interest. Candidate genes were prioritised in this study based on RNAi knockdown experiments. The most promising candidates to pursue are on chromosome III, where the interval is much smaller. These are *sor-1*, which encodes a component of polycomb repressor complex 1 (PRC1) and *hsp-110*, which is a co-chaperone involved in protein homeostasis. *sor-1* in CB4856 contains a single missense mutation (G–A) that converts methionine to isoleucine (Met760Ile) in CB4856 whereas *hsp-110* in CB4856 contains an in-frame deletion of three nucleotides (GAT) that results in the deletion of a highly conserved amino acid (Asp474del).

6.5 A developmental model for buffering seam cell number in *egl-18* loss-of-function mutants

Most seam cell divisions during larval development are asymmetric, where the anterior cell adopts hyp7 fate and the posterior cell adopts seam cell fate. Activation of Wnt β -catenin asymmetry ($W\beta A$) pathway in the posterior daughter leads to lowering of nuclear POP-1 levels. Higher level of POP-1/TCF acts as a transcription repressor in the anterior daughter, whereas a low level of POP/TCF along with transcriptional coactivator acts to activate Wnt-target genes, such as *egl-18* that specifies seam cell fate. In *egl-18* mutants, posterior seam cells lose the seam cell fate and differentiate to hyp7 fate. However, not all posterior seam cells lose fate because of the function of *elt-6*, which redundantly regulates seam cell fate, as well as embryonic development (Koh and Rothman, 2001). In contrast to the *egl-18* loss of function phenotype, *pop-1* RNAi increases SCN due to symmetrisation of seam cell divisions towards the seam cell fate. The increase in SCN is dependent on the function of *egl-18* in both anterior and posterior seam cells as *pop-1* RNAi in an *egl-18* loss-of-function background not

only abrogates the increase but also leads to further reduction in SCN.

A question which arises is whether buffering SCN in *egl-18* loss-of-function mutants may work via differences in the regulation of the asymmetric cell divisions potentially through the $W\beta A$ pathway. N2 and CB4856 not only differ in SCN upon the loss of *egl-18*, but also upon decrease in levels of the upstream transcriptional activator *pop-1*, suggesting that there may be indeed a difference in the $W\beta A$ pathway between the two isolates. Based on my observation that the discovered QTLs influence the SCN also upon *pop-1* RNAi in the two isolates, I argue that the QTLs are likely to act as modifiers of the $W\beta A$ pathway.

There is a difference of two seam cells on average between N2 and CB4856 carrying *egl-18(ga97)*. There are two possible development mechanisms by which SCN might be buffered in *egl-18* loss-of-function mutants, depending on whether the anterior or posterior cell daughters of the asymmetric divisions are affected. First, candidate gene variants in the QTLs might cause an increase in symmetrisation of the cell divisions towards seam cell fate to increase SCN. Second, a modifying candidate gene in the QTLs might function to reinforce seam cell fate in the posterior cell in the absence of *egl-18*. Either or both of these processes could occur differentially in N2 and CB4856 *egl-18* mutants and drive a difference in SCN. However, I favour the hypothesis which involves reinforcing or stabilising seam cell fate in the posterior cell in CB4856 due to the following reasons. I have not observed symmetrisation of seam cell divisions while counting SCN in thousands of *egl-18(ga97)* animals. Instead, I only observed large spaces between seam cells, which is more consistent with the differentiation and thus loss of seam cells. Second, RNAi knockdown of all QTL candidate genes (*egl-27*, *sor-1* and *hsp-110*) behave like *pop-1* RNAi in *egl-18(ga97)* (Gorrepati, K. W. Thompson, and Eisenmann, 2013), which causes further decrease in SCN by the loss of seam cell fate in the posterior seam daughters. This suggests that the modifiers may buffer the maintenance of posterior seam cell fate in the absence of *egl-18* summarised in 6.1. Further investigation needs to involve precise lineaging making use of newly available technological advances to fully understand the developmental basis of the differential expressivity of *egl-18* mutations between N2 and CB4856 (Gritti et al., 2016).

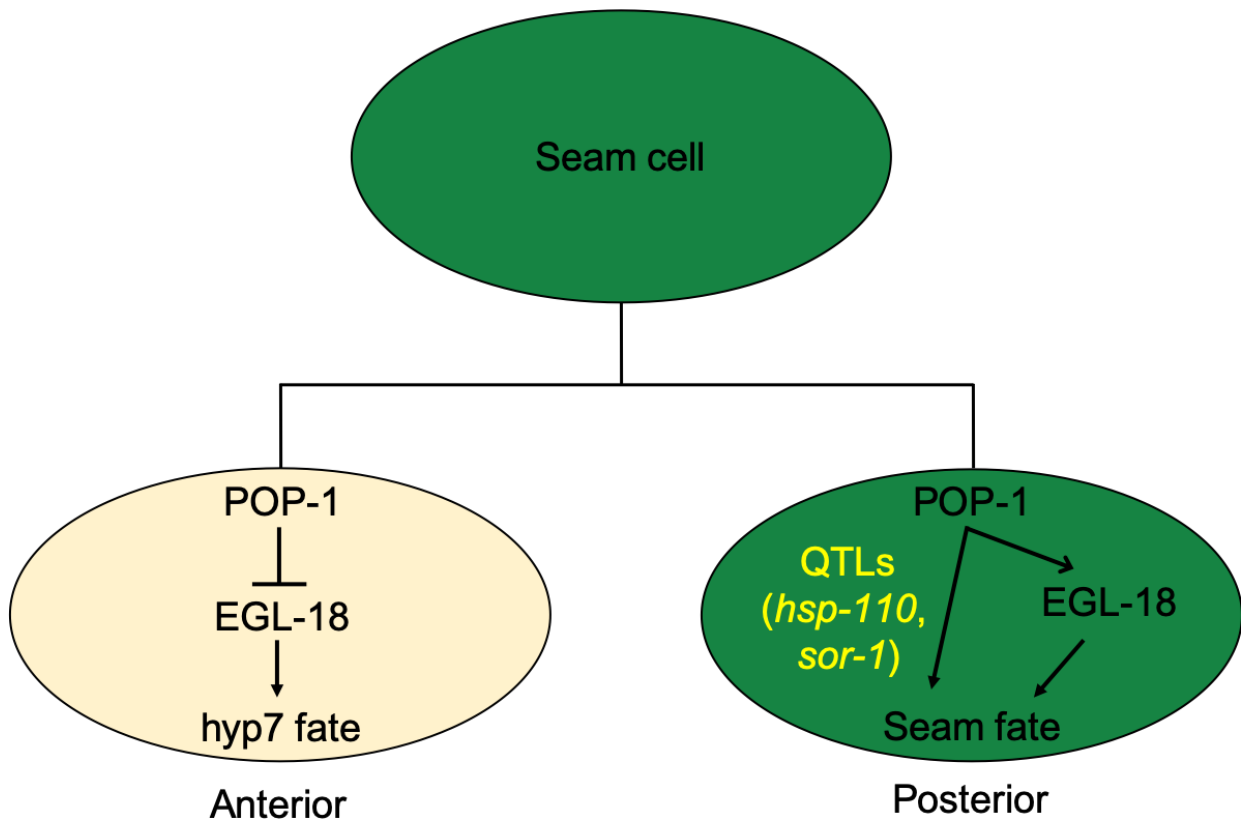


Figure 6.1: A schematic of the developmental model by which seam cell number may be buffered in *egl-18(ls)* mutants. POP-1 acts as a transcriptional repressor in the anterior cell, which adopts a hypodermal (*hyp7*) fate. Through the activation of $W\beta A$ pathway, POP-1 acts as a transcriptional activator and causes the expression of transcription factors such as EGL-18 causing the posterior cell to adopt a seam cell fate. Candidates genes (e.g. *hsp-110* and/or *sor-1*) from QTLs may act to reinforce seam cell fate in the posterior nucleus in the absence of EGL-18.

Bibliography

- [1] Abitbol, Marie, Bernex, Florence, Puy, Hervé, Jouault, Hélène, Deybach, Jean-Charles, Guénet, Jean-Louis, and Montagutelli, Xavier. A mouse model provides evidence that genetic background modulates anemia and liver injury in erythropoietic protoporphyria. *American journal of physiology. Gastrointestinal and liver physiology* 288.6 (2005), G1208–16.
- [2] Alper, Scott and Podbilewicz, Benjamin. Cell Fusion in *Caenorhabditis elegans*. *Cell Fusion*. Totowa, NJ: Humana Press, 2008, pp. 53–74.
- [3] Andersen, Erik C, Bloom, Joshua S, Gerke, Justin P, and Kruglyak, Leonid. A variant in the neuropeptide receptor *npr-1* is a major determinant of *Caenorhabditis elegans* growth and physiology. *PLoS Genetics* 10.2 (2014), e1004156.
- [4] Andersen, Erik C, Gerke, Justin P, Shapiro, Joshua A, Crissman, Jonathan R, Ghosh, Rajarshi, Bloom, Joshua S, Félix, Marie-Anne, and Kruglyak, Leonid. Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. *Nature genetics* 44.3 (2012), pp. 285–290.
- [5] Armenti, Stephen T and Nance, Jeremy. Adherens Junctions in *C. elegans* Embryonic Morphogenesis. *Adherens Junctions: from Molecular Mechanisms to Tissue Development and Disease*. Dordrecht: Springer Netherlands, 2012, pp. 279–299.
- [6] Arribere, Joshua A, Arribere, J A, Bell, Ryan T, Bell, R T, Fu, B X H, Fu, Becky X H, Artiles, Karen L, Artiles, K L, Hartman, P S, Hartman, Phil S, Fire, A Z, and Fire, Andrew Z. Efficient Marker-Free Recovery of Custom Genetic Modifications with CRISPR/Cas9 in *Caenorhabditis elegans*. *Genetics* 198.3 (2014), pp. 837–846.

- [7] Austin, Judith and Kenyon, Cynthia. Cell contact regulates neuroblast formation in the *Caenorhabditis elegans* lateral epidermis. *Development* 120.2 (1994), pp. 313–323.
- [8] Baldwin, Austin T, Clemons, Amy M, and Phillips, Bryan T. Unique and redundant β -catenin regulatory roles of two Dishevelled paralogs during *C. elegans* asymmetric cell division. *Journal of cell science* 129.5 (2016), pp. 983–993.
- [9] Baldwin, Austin T and Phillips, Bryan T. The tumor suppressor APC differentially regulates multiple β -catenins through the function of axin and CKI α during *C. elegans* asymmetric stem cell divisions. *Journal of cell science* 127.Pt 12 (2014), pp. 2771–2781.
- [10] Banerjee, Diya, Chen, Xin, Lin, Shin Yi, and Slack, Frank J. kin-19/casein kinase I α has dual functions in regulating asymmetric division and terminal differentiation in *C. elegans* epidermal stem cells. *Cell Cycle* 9.23 (2010), pp. 4748–4765.
- [11] Barkoulas, Michalis, Zon, Jeroen S van, Milloz, Josselin, Oudenaarden, Alexander van, and Félix, Marie-Anne. Robustness and Epistasis in the *C. elegans* Vulval Signaling Network Revealed by Pathway Dosage Modulation. *Developmental Cell* 24.1 (2013), pp. 64–75.
- [12] Barrière, Antoine and Félix, Marie-Anne. High Local Genetic Diversity and Low Outcrossing Rate in *Caenorhabditis elegans* Natural Populations. *Current Biology* 15.13 (2005), pp. 1176–1184.
- [13] Barrière, Antoine and Félix, Marie-Anne. Natural variation and population genetics of *Caenorhabditis elegans*. *WormBook* (2005). DOI: doi/10.1895/wormbook.1.43.1.
- [14] Bazakos, Christos, Hanemian, Mathieu, Trontin, Charlotte, Jiménez-Gómez, José M, and Loudet, Olivier. New Strategies and Tools in Quantitative Genetics: How to Go from the Phenotype to the Genotype. *Annual review of plant biology* 68.1 (2017), pp. 435–455.
- [15] Ben-David, Eyal, Burga, Alejandro, and Kruglyak, Leonid. A maternal-effect selfish genetic element in *Caenorhabditis elegans*. *Science* 356.6342 (2017), pp. 1051–1055. ISSN: 0036-8075. DOI: 10.1126/science.aan0621. eprint: <https://science.sciencemag.org/content/356/6342/1051.full.pdf>. URL: <https://science.sciencemag.org/content/356/6342/1051>.

- [16] Bergman, Aviv and Siegal, Mark L. Evolutionary capacitance as a general feature of complex gene networks. *Nature* 424.6948 (2003), pp. 549–552.
- [17] Bhambhani, Chandan, Ravindranath, Aditi J, Mentink, Remco A, Chang, Mikyung V, Betist, Marco C, Yang, Yaxuan X, Koushika, Sandhya P, Korswagen, Hendrik C, and Cadigan, Ken M. Distinct DNA Binding Sites Contribute to the TCF Transcriptional Switch in *C. elegans* and *Drosophila*. *PLoS Genetics* 10.2 (2014), e1004133.
- [18] Blazie, Stephen M, Geissel, Heather C, Wilky, Henry, Joshi, Rajan, Newbern, Jason, and Mangone, Marco. Alternative Polyadenylation Directs Tissue-Specific miRNA Targeting in *Caenorhabditis elegans* Somatic Tissues. *Genetics* 206.2 (2017), pp. 757–774.
- [19] Block, Dena Hs and Shapira, Michael. GATA transcription factors as tissue-specific master regulators for induced responses. *Worm* 4.4 (2015), e1118607.
- [20] Blumenthal, Thomas. Trans-splicing and operons. *WormBook* (2005). DOI: doi/10.1895/wormbook.1.5.1.
- [21] Bone, Courtney R, Chang, Yu-Tai, Cain, Natalie E, Murphy, Shaun P, and Starr, Daniel A. Nuclei migrate through constricted spaces using microtubule motors and actin networks in *C. elegans* hypodermal cells. *Development* 143.22 (2016), pp. 4193–4202.
- [22] Brabin, Charles, Appleford, Peter J, and Woollard, Alison. 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), e1002200.
- [23] Brabin, Charles and Woollard, Alison. Finding a niche for seam cells? *Worm* 1.2 (2012), pp. 107–111.
- [24] Braendle, Christian and Félix, Marie-Anne. Plasticity and Errors of a Robust Developmental System in Different Environments. *Developmental Cell* 15.5 (2008), pp. 714–724.
- [25] Brenner, Sydney. The genetics of *Caenorhabditis elegans*. *Genetics* (1974).

- [26] Burga, Alejandro, Ben-David, Eyal, Lemus Vergara, Tzitziki, Boocock, James, and Kruglyak, Leonid. Fast genetic mapping of complex traits in *C. elegans* using millions of individuals in bulk. *Nature communications* 10.1 (2019), p. 2680.
- [27] Burga, Alejandro, Casanueva, M Olivia, and Lehner, Ben. Predicting mutation outcome from early stochastic variation in genetic interaction partners. *Nature* 480.7376 (2011), pp. 250–253.
- [28] Cameron, Ewan R and Neil, James C. The Runx genes: lineage-specific oncogenes and tumor suppressors. *Oncogene* 23.24 (2004), pp. 4308–4314.
- [29] Campo, Jacob J del, Opoku-Serebuoh, Eugene, Isaacson, Ariel B, Scranton, Victoria L, Tucker, Morgan, Han, Min, and Mohler, William A. Fusogenic activity of EFF-1 is regulated via dynamic localization in fusing somatic cells of *C. elegans*. *Current biology : CB* 15.5 (2005), pp. 413–423.
- [30] Cao, Junyue, Packer, Jonathan S, Ramani, Vijay, Cusanovich, Darren A, Huynh, Chau, Daza, Riza, Qiu, Xiaojie, Lee, Choli, Furlan, Scott N, Steemers, Frank J, Adey, Andrew, Waterston, Robert H, Trapnell, Cole, and Shendure, Jay. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science (New York, N.Y.)* 357.6352 (2017), pp. 661–667.
- [31] Casanueva, M Olivia, Burga, Alejandro, and Lehner, Ben. Fitness trade-offs and environmentally induced mutation buffering in isogenic *C. elegans*. *Science (New York, N.Y.)* 335.6064 (2012), pp. 82–85.
- [32] Cassata, Giuseppe, Shemer, Gidi, Morandi, Paolo, Donhauser, Roland, Podbilewicz, Benjamin, and Baumeister, Ralf. *ceh-16*/engrailed patterns the embryonic epidermis of *Caenorhabditis elegans*. *Development* 132.4 (2005), pp. 739–749.
- [33] Chandler, Christopher H. Cryptic intraspecific variation in sex determination in *Caenorhabditis elegans* revealed by mutations. *Heredity* 105.5 (2010), pp. 473–482.
- [34] Chandler, Christopher H, Chari, Sudarshan, and Dworkin, Ian. Does your gene need a background check? How genetic background impacts the analysis of mutations, genes, and evolution. *Trends in genetics : TIG* 29.6 (2013), pp. 358–366.

- [35] Chen, Zhe, Eastburn, Dennis J, and Han, Min. The *Caenorhabditis elegans* nuclear receptor gene *nhr-25* regulates epidermal cell development. *Molecular and cellular biology* 24.17 (2004), pp. 7345–7358.
- [36] Chiorazzi, Michael, Rui, Lixin, Yang, Yandan, Ceribelli, Michele, Tishbi, Nima, Maurer, Carine W, Ranuncolo, Stella M, Zhao, Hong, Xu, Weihong, Chan, Wing-Chung C, Jaffe, Elaine S, Gascoyne, Randy D, Campo, Elias, Rosenwald, Andreas, Ott, German, Delabie, Jan, Rimsza, Lisa M, Shaham, Shai, and Staudt, Louis M. Related F-box proteins control cell death in *Caenorhabditis elegans* and human lymphoma. *Proceedings of the National Academy of Sciences of the United States of America* 110.10 (2013), pp. 3943–3948.
- [37] Chow, Clement Y. Bringing genetic background into focus. *Nature Reviews Genetics* 17.2 (2016), pp. 63–64.
- [38] Coffman, James A. Runx transcription factors and the developmental balance between cell proliferation and differentiation. *Cell biology international* 27.4 (2003), pp. 315–324.
- [39] Conant, Gavin C and Wagner, Andreas. Duplicate genes and robustness to transient gene knock-downs in *Caenorhabditis elegans*. *Proceedings. Biological sciences / The Royal Society* 271.1534 (2004), pp. 89–96.
- [40] Cook, Daniel E, Zdraljevic, Stefan, Roberts, Joshua P, and Andersen, Erik C. CeNDR, the *Caenorhabditis elegans* natural diversity resource. *Nucleic Acids Research* 45.D1 (2017), pp. D650–D657.
- [41] Crombie, Timothy A, Zdraljevic, Stefan, Cook, Daniel E, Tanny, Robyn E, Brady, Shannon C, Wang, Ye, Evans, Kathryn S, Hahnel, Steffen, Lee, Daehan, Rodriguez, Briana C, Zhang, Gaotian, Zwaag, Joost van der, Kiontke, Karin C, and Andersen, Erik C. Deep sampling of Hawaiian *Caenorhabditis elegans* reveals high genetic diversity and admixture with global populations. *bioRxiv* 47 (2019), p. 716928.
- [42] Cutting, Garry R. Modifier genes in Mendelian disorders: the example of cystic fibrosis. *Annals of the New York Academy of Sciences* 1214.1 (2010), pp. 57–69.

- [43] Dickinson, Daniel J, Ward, Jordan D, Reiner, David J, and Goldstein, Bob. Engineering the *Caenorhabditis elegans* genome using Cas9-triggered homologous recombination. *Nature Methods* 10.10 (2013), pp. 1028–1034.
- [44] Doitsidou, Maria and Hobert, Oliver. New alleles of the *lin-22*/Hairy bHLH transcription factor. *microPublication Biology* (2019). DOI: 10.17912/micropub.biology.000111. URL: <https://www.micropublication.org/journals/biology/micropub-biology-000111/>.
- [45] Doitsidou, Maria, Poole, Richard J, Sarin, Sumeet, Bigelow, Henry, and Hobert, Oliver. *C. elegans* Mutant Identification with a One-Step Whole-Genome-Sequencing and SNP Mapping Strategy. *PLOS ONE* 5.11 (2010), e15435.
- [46] Dokshin, Gregoriy A, Ghanta, Krishna S, Piscopo, Katherine M, and Mello, Craig C. Robust Genome Editing with Short Single-Stranded and Long, Partially Single-Stranded DNA Donors in *Caenorhabditis elegans*. *Genetics* 210.3 (2018), pp. 781–787.
- [47] Dowell, Robin D, Ryan, Owen, Jansen, An, Cheung, Doris, Agarwala, Sudeep, Danford, Timothy, Bernstein, Douglas A, Rolfe, P Alexander, Heisler, Lawrence E, Chin, Brian, Nislow, Corey, Giaever, Guri, Phillips, Patrick C, Fink, Gerald R, Gifford, David K, and Boone, Charles. Genotype to Phenotype: A Complex Problem. *Science (New York, N.Y.)* 328.5977 (2010), pp. 469–469.
- [48] Duveau, Fabien and Félix, Marie-Anne. Role of Pleiotropy in the Evolution of a Cryptic Developmental Variation in *Caenorhabditis elegans*. *PLoS Biology* 10.1 (2012), e1001230–19.
- [49] Dworkin, Ian, Kennerly, Erin, Tack, David, Hutchinson, Jennifer, Brown, Julie, Mahaffey, James, and Gibson, Greg. Genomic consequences of background effects on scalloped mutant expressivity in the wing of *Drosophila melanogaster*. *Genetics* 181.3 (2009), pp. 1065–1076.
- [50] Dworkin, Ian, Palsson, Arnar, Birdsall, Kelli, and Gibson, Greg. Evidence that *Egfr* contributes to cryptic genetic variation for photoreceptor determination in natural populations of *Drosophila melanogaster*. *Current biology : CB* 13.21 (2003), pp. 1888–1893.

- [51] Eisenmann, David M and Kim, Stuart K. Protruding Vulva Mutants Identify Novel Loci and Wnt Signaling Factors That Function During *Caenorhabditis elegans* Vulva Development. *Genetics* 156.3 (2000), pp. 1097–1116.
- [52] El Mouridi, Sonia, Lecroisey, Claire, Tardy, Philippe, Mercier, Marine, Leclercq-Blondel, Alice, Zariohi, Nora, and Boulin, Thomas. Reliable CRISPR/Cas9 Genome Engineering in *Caenorhabditis elegans* Using a Single Efficient sgRNA and an Easily Recognizable Phenotype. *G3 (Bethesda, Md.)* 7.5 (2017), pp. 1429–1437.
- [53] Elowitz, Michael B, Levine, Arnold J, Siggia, Eric D, and Swain, Peter S. Stochastic Gene Expression in a Single Cell. *Science (New York, N. Y.)* 297.5584 (2002), pp. 1183–1186.
- [54] Evans, Thomas. Transformation and microinjection. *WormBook* (2006). DOI: doi/10.1895/wormbook.1.108.1.
- [55] Farboud, Behnom and Meyer, Barbara J. Dramatic enhancement of genome editing by CRISPR/Cas9 through improved guide RNA design. *Genetics* 199.4 (2015), pp. 959–971.
- [56] Félix, Marie-Anne and Barkoulas, Michalis. Robustness and flexibility in nematode vulva development. *Trends in genetics : TIG* 28.4 (2012), pp. 185–195.
- [57] Félix, Marie-Anne and Barkoulas, Michalis. Pervasive robustness in biological systems. *Nature Reviews Genetics* 16.8 (2015), pp. 483–496.
- [58] Félix, Marie-Anne and Wagner, Andreas. Robustness and evolution: concepts, insights and challenges from a developmental model system. *Heredity* 100.2 (2008), pp. 132–140.
- [59] Fournier, Téo and Schacherer, Joseph. Genetic backgrounds and hidden trait complexity in natural populations. *Current opinion in genetics & development* 47 (2017), pp. 48–53.
- [60] Frézal, Lise, Demoinet, Emilie, Braendle, Christian, Miska, Eric, and Félix, Marie-Anne. Natural Genetic Variation in a Multigenerational Phenotype in *C. elegans*. *Current biology : CB* 28.16 (2018), 2588–2596.e8.

- [61] Friedland, Ari E, Tzur, Yonatan B, Esvelt, Kevin M, Colaiácovo, Monica P, Church, George M, and Calarco, John A. Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nature Methods* 10.8 (2013), pp. 741–743.
- [62] Friedlander-Shani, Lilach and Podbilewicz, Benjamin. Heterochronic Control of AFF-1-Mediated Cell-to-Cell Fusion in *C. elegans*. *Cell Fusion in Health and Disease*. Dordrecht: Springer Netherlands, 2011, pp. 5–11.
- [63] Fujii, Takashi, Nakao, Fumi, Shibata, Yukimasa, Shioi, Go, Kodama, Eiji, Fujisawa, Hajime, and Takagi, Shin. *Caenorhabditis elegans* PlexinA, PLX-1, interacts with transmembrane semaphorins and regulates epidermal morphogenesis. *Development* 129.9 (2002), pp. 2053–2063.
- [64] Gasch, Audrey P, Payseur, Bret A, and Pool, John E. The Power of Natural Variation for Model Organism Biology. *Trends in genetics : TIG* 32.3 (2016), pp. 147–154.
- [65] Gattegno, Tamar, Mittal, Aditya, Valansi, Clari, Nguyen, Ken C Q, Hall, David H, Chernomordik, Leonid V, and Podbilewicz, Benjamin. Genetic control of fusion pore expansion in the epidermis of *Caenorhabditis elegans*. *Molecular Biology of the Cell* 18.4 (2007), pp. 1153–1166.
- [66] Ghose, Piya, Rashid, Alina, Insley, Peter, Trivedi, Meera, Shah, Pavak, Singhal, Anupriya, Lu, Yun, Bao, Zhirong, and Shaham, Shai. EFF-1 fusogen promotes phagosome sealing during cell process clearance in *Caenorhabditis elegans*. *Nature Cell Biology* 20.4 (2018), pp. 393–399.
- [67] Giaever, Guri, Chu, Angela M, Ni, Li, Connelly, Carla, Riles, Linda, Véronneau, Steve, Dow, Sally, Lucau-Danila, Ankuta, Anderson, Keith, André, Bruno, Arkin, Adam P, Astromoff, Anna, El-Bakkoury, Mohamed, Bangham, Rhonda, Benito, Rocio, Brachat, Sophie, Campanaro, Stefano, Curtiss, Matt, Davis, Karen, Deutschbauer, Adam, Entian, Karl-Dieter, Flaherty, Patrick, Foury, Francoise, Garfinkel, David J, Gerstein, Mark, Gotte, Deanna, Güldener, Ulrich, Hegemann, Johannes H, Hempel, Svenja, Herman, Zelek, Jaramillo, Daniel F, Kelly, Diane E, Kelly, Steven L, Kötter, Peter, LaBonte, Darlene, Lamb, David C, Lan, Ning, Liang, Hong, Liao, Hong, Liu, Lucy, Luo, Chuanyun,

- Lussier, Marc, Mao, Rong, Menard, Patrice, Ooi, Siew Loon, Revuelta, Jose L, Roberts, Christopher J, Rose, Matthias, Ross-Macdonald, Petra, Scherens, Bart, Schimmack, Greg, Shafer, Brenda, Shoemaker, Daniel D, Sookhai-Mahadeo, Sharon, Storms, Reginald K, Strathern, Jeffrey N, Valle, Giorgio, Voet, Marleen, Volckaert, Guido, Wang, Ching-yun, Ward, Teresa R, Wilhelmy, Julie, Winzeler, Elizabeth A, Yang, Yonghong, Yen, Grace, Youngman, Elaine, Yu, Kexin, Bussey, Howard, Boeke, Jef D, Snyder, Michael, Philippsen, Peter, Davis, Ronald W, and Johnston, Mark. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418.6896 (2002), pp. 387–391.
- [68] Gibson, Greg and Dworkin, Ian. Uncovering cryptic genetic variation. *Nature Reviews Genetics* 5.9 (2004), pp. 681–690.
- [69] Gibson, Greg and Helden, Sylvie van. Is function of the *Drosophila* homeotic gene *Ultrabithorax* canalized? *Genetics* 147.3 (1997), pp. 1155–1168.
- [70] Gilleard, J S, Barry, J D, and Johnstone, I L. cis regulatory requirements for hypodermal cell-specific expression of the *Caenorhabditis elegans* cuticle collagen gene *dpy-7*. *Molecular and cellular biology* 17.4 (1997), pp. 2301–2311.
- [71] Ginzburg, Val E, Roy, Peter J, and Culotti, Joseph G. Semaphorin 1a and semaphorin 1b are required for correct epidermal cell positioning and adhesion during morphogenesis in *C. elegans*. *Development* 129.9 (2002), pp. 2065–2078.
- [72] Gleason, Julie E and Eisenmann, David M. Wnt signaling controls the stem cell-like asymmetric division of the epithelial seam cells during *C. elegans* larval development. *Developmental Biology* 348.1 (2010), pp. 58–66.
- [73] Gorrepati, Lakshmi and Eisenmann, David M. The *C. elegans* embryonic fate specification factor EGL-18 (GATA) is reutilized downstream of Wnt signaling to maintain a population of larval progenitor cells. *Worm* 4.1 (2015), e996419.
- [74] Gorrepati, Lakshmi, Thompson, Kenneth W, and Eisenmann, David M. *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 (2013), pp. 2093–2102.

- [75] Grimbert, Stéphanie and Braendle, Christian. Cryptic genetic variation uncovers evolution of environmentally sensitive parameters in *Caenorhabditis vulval* development. *Evolution & Development* 16.5 (2014), pp. 278–291.
- [76] Gritti, Nicola, Kienle, Simone, Filina, Olga, and Zon, Jeroen Sebastiaan van. Long-term time-lapse microscopy of *C. elegans* post-embryonic development. *Nature communications* 7 (2016), p. 12500.
- [77] Gu, Zhenglong, Steinmetz, Lars M, Gu, Xun, Scharfe, Curt, Davis, Ronald W, and Li, Wen-Hsiung. Role of duplicate genes in genetic robustness against null mutations. *Nature* 421.6918 (2003), pp. 63–66.
- [78] Guignard, Léo, Fiuza, Ulla-Maj, Leggio, Bruno, Faure, Emmanuel, Laussu, Julien, Hufnagel, Lars, Malandain, Grégoire, Godin, Christophe, and Lemaire, Patrick. Contact-dependent cell-cell communications drive morphological invariance during ascidian embryogenesis. *bioRxiv* (2017). DOI: 10.1101/238741. eprint: <https://www.biorxiv.org/content/early/2017/12/24/238741.full.pdf>. URL: <https://www.biorxiv.org/content/early/2017/12/24/238741>.
- [79] Haber, Markus, Schüngel, Manuela, Putz, Annika, Müller, Sabine, Hasert, Barbara, and Schulenburg, Hinrich. Evolutionary history of *Caenorhabditis elegans* inferred from microsatellites: evidence for spatial and temporal genetic differentiation and the occurrence of outbreeding. *Molecular biology and evolution* 22.1 (2005), pp. 160–173.
- [80] Hall, David H and Altun, Zeynep F. C. *Elegans Atlas*. CSHL Press, 2008.
- [81] Harterink, Martin, Kim, Dong Hyun, Middelkoop, Teije C, Doan, Thang Dinh, Oude-naarden, Alexander van, and Korswagen, Hendrik C. 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 (2011), pp. 2915–2924.
- [82] Hedgecock, Edward M and White, John G. Polyploid tissues in the nematode *Caenorhabditis elegans*. *Developmental Biology* 107.1 (1985), pp. 128–133.

- [83] Herman, Michael A, Ch'ng, QueeLim, Hettenbach, Susan M, Ratliff, Thomas M, Kenyon, Cynthia, and Herman, Robert K. EGL-27 is similar to a metastasis-associated factor and controls cell polarity and cell migration in *C. elegans*. *Development* 126.5 (1999), pp. 1055–1064.
- [84] Hodgkin, Jonathan and Doniach, Tabitha. Natural variation and copulatory plug formation in *Caenorhabditis elegans*. *Genetics* 146.1 (1997), pp. 149–164.
- [85] Huang, Xinxin, Tian, E, Xu, Yanhua, and Zhang, Hong. The *C. elegans* engrailed homolog *ceh-16* regulates the self-renewal expansion division of stem cell-like seam cells. *Developmental Biology* 333.2 (2009), pp. 337–347.
- [86] Hughes, Samantha, Brabin, Charles, Appleford, Peter J, and Woollard, Alison. 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 (2013), pp. 718–727.
- [87] Joshi, Pradeep M, Riddle, Misty R, Djabrayan, Nareg J V, and Rothman, Joel H. *Caenorhabditis elegans* as a model for stem cell biology. *Developmental Dynamics* 239.5 (2010), pp. 1539–1554.
- [88] Kageyama, Ryoichiro, Ohtsuka, Toshiyuki, and Kobayashi, Taeko. The Hes gene family: repressors and oscillators that orchestrate embryogenesis. *Development* 134.7 (2007), pp. 1243–1251.
- [89] Kagoshima, H, Nimmo, R, Saad, N, Tanaka, J, Miwa, Y, Mitani, S, Kohara, Y, and Woollard, A. The *C. elegans* CBF homologue BRO-1 interacts with the Runx factor, RNT-1, to promote stem cell proliferation and self-renewal. *Development* 134.21 (2007), pp. 3905–3915.
- [90] Kagoshima, Hiroshi, Sawa, Hitoshi, Mitani, Shohei, Bürglin, Thomas R, Shigesada, Katsuya, and Kohara, Yuji. The *C. elegans* RUNX transcription factor RNT-1/MAB-2 is required for asymmetrical cell division of the T blast cell. *Developmental Biology* 287.2 (2005), pp. 262–273.
- [91] Kamath, Ravi S and Ahringer, Julie. Genome-wide RNAi screening in *Caenorhabditis elegans*. *Methods (San Diego, Calif.)* 30.4 (2003), pp. 313–321.

- [92] Kanamori, Takahiro, Inoue, Takao, Sakamoto, Taro, Gengyo-Ando, Keiko, Tsujimoto, Masafumi, Mitani, Shohei, Sawa, Hitoshi, Aoki, Junken, and Arai, Hiroyuki. Beta-catenin asymmetry is regulated by PLA1 and retrograde traffic in *C. elegans* stem cell divisions. *The EMBO journal* 27.12 (2008), pp. 1647–1657.
- [93] Katsanos, Dimitris, Koneru, Sneha L, Mestek Boukhibar, Lamia, Gritti, Nicola, Ghose, Ritobrata, Appleford, Peter J, Doitsidou, Maria, Woollard, Alison, Zon, Jeroen S van, Poole, Richard J, and Barkoulas, Michalis. Stochastic loss and gain of symmetric divisions in the *C. elegans* epidermis perturbs robustness of stem cell number. *PLoS Biology* 15.11 (2017), e2002429.
- [94] Kenyon, Cynthia. A gene involved in the development of the posterior body region of *C. elegans*. *Cell* 46.3 (1986), pp. 477–487.
- [95] Kidd, Ambrose R, Miskowski, Jennifer A, Siegfried, Kellee R, Sawa, Hitoshi, and Kimble, Judith. A beta-catenin identified by functional rather than sequence criteria and its role in Wnt/MAPK signaling. *Cell* 121.5 (2005), pp. 761–772.
- [96] Kim, UnKyung, Jorgenson, Eric, Coon, Hilary, Leppert, Mark, Risch, Neil, and Drayna, Dennis. Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. *Science (New York, N.Y.)* 299.5610 (2003), pp. 1221–1225.
- [97] Kitano, Hiroaki. Biological robustness. *Nature Reviews Genetics* 5.11 (2004), pp. 826–837.
- [98] Koh, Kyunghhee, Peyrot, Sara M, Wood, Cricket G, Wagmaister, Javier A, Maduro, Morris F, Eisenmann, David M, and Rothman, Joel H. Cell fates and fusion in the *C. elegans* vulval primordium are regulated by the EGL-18 and ELT-6 GATA factors — apparent direct targets of the LIN-39 Hox protein. *Development* 129.22 (2002), pp. 5171–5180.
- [99] Koh, Kyunghhee and Rothman, Joel H. ELT-5 and ELT-6 are required continuously to regulate epidermal seam cell differentiation and cell fusion in *C. elegans*. *Development* 128.15 (2001), pp. 2867–2880.

- [100] Kuzmanov, Aleksandra, Yochem, John, and Fay, David S. Analysis of PHA-1 reveals a limited role in pharyngeal development and novel functions in other tissues. *Genetics* 198.1 (2014), pp. 259–268.
- [101] Lam, Arielle Koonyee and Phillips, Bryan T. Wnt Signaling Polarizes *C. elegans* Asymmetric Cell Divisions During Development. *Results and problems in cell differentiation* 61.Chapter 4 (2017), pp. 83–114.
- [102] Landrum, Melissa J and Kattman, Brandi L. ClinVar at five years: Delivering on the promise. *Human mutation* 39.11 (2018), pp. 1623–1630.
- [103] Lee, Rosalind C, Feinbaum, Rhonda L, and Ambros, Victor. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75.5 (1993), pp. 843–854.
- [104] Lehner, Ben, Crombie, Catriona, Tischler, Julia, Fortunato, Angelo, and Fraser, Andrew G. Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. *Nature genetics* 38.8 (2006), pp. 896–903.
- [105] Levy, Sasha F and Siegal, Mark L. Network Hubs Buffer Environmental Variation in *Saccharomyces cerevisiae*. *PLoS Biology* 6.11 (2008), e264.
- [106] Li, Heng, Handsaker, Bob, Wysoker, Alec, Fennell, Tim, Ruan, Jue, Homer, Nils, Marth, Gabor, Abecasis, Goncalo, Durbin, Richard, and 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* 25.16 (2009), pp. 2078–2079.
- [107] Lo, Miao-Chia, Gay, Frédérique, Odom, Raanan, Shi, Yang, and Lin, Rueyling. 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 (2004), pp. 95–106.
- [108] Ma, Zhipeng, Zhu, Peipei, Shi, Hui, Guo, Liwei, Zhang, Qinghe, Chen, Yanan, Chen, Shuming, Zhang, Zhe, Peng, Jinrong, and Chen, Jun. PTC-bearing mRNA elicits a genetic compensation response via Upf3a and COMPASS components. *Nature* 568.7751 (2019), pp. 259–263.

- [109] Mackay, Trudy F C, Stone, Eric A, and Ayroles, Julien F. The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics* 10.8 (2009), pp. 565–577.
- [110] Maduro, Morris F. Developmental robustness in the *Caenorhabditis elegans* embryo. *Molecular Reproduction and Development* 82.12 (2015), pp. 918–931.
- [111] Mallick, Avijit, Ranawade, Ayush, and Gupta, Bhagwati P. Role of PRY-1/Axin in heterochronic miRNA-mediated seam cell development. *BMC developmental biology* 19.1 (2019), pp. 17–12.
- [112] Masel, Joanna and Siegal, Mark L. Robustness: mechanisms and consequences. *Trends in Genetics* 25.9 (2009), pp. 395–403.
- [113] Mestek Boukhibar, Lamia and Barkoulas, Michalis. The developmental genetics of biological robustness. *Annals of Botany* 117.5 (2016), pp. 699–707.
- [114] Michaux, Grègoire, Legouis, Renaud, and Labouesse, Michel. Epithelial biology: lessons from *Caenorhabditis elegans*. *Gene* 277.1-2 (2001), pp. 83–100.
- [115] Michelmore, Richard W, Paran, Ilan, and Kesseli, Richard V. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences* 88.21 (1991), pp. 9828–9832.
- [116] Milloz, Josselin, Duveau, Fabien, Nuez, Isabelle, and Félix, Marie-Anne. Intraspecific evolution of the intercellular signaling network underlying a robust developmental system. *Genes & development* 22.21 (2008), pp. 3064–3075.
- [117] Minevich, Gregory, Park, Danny S, Blankenberg, Daniel, Poole, Richard J, and Hobert, Oliver. CloudMap: a cloud-based pipeline for analysis of mutant genome sequences. *Genetics* 192.4 (2012), pp. 1249–1269.
- [118] Miyabayashi, Tomoyuki, Palfreyman, Mark T, Sluder, Ann E, Slack, Frank, and Sengupta, Piali. Expression and Function of Members of a Divergent Nuclear Receptor Family in *Caenorhabditis elegans*. *Developmental Biology* 215.2 (1999), pp. 314–331.

- [119] Mizumoto, Kota and Sawa, Hitoshi. Cortical β -Catenin and APC Regulate Asymmetric Nuclear β -Catenin Localization during Asymmetric Cell Division in *C. elegans*. *Developmental Cell* 12.2 (2007), pp. 287–299.
- [120] Mizumoto, Kota and Sawa, Hitoshi. Two betas or not two betas: regulation of asymmetric division by beta-catenin. *Trends in Cell Biology* 17.10 (2007), pp. 465–473.
- [121] Mohler, William A, Shemer, Gidi, Campo, Jacob J del, Valansi, Clari, Opoku-Serebuoh, Eugene, Scranton, Victoria, Assaf, Nirit, White, John G, and Podbilewicz, Benjamin. The type I membrane protein EFF-1 is essential for developmental cell fusion. *Developmental Cell* 2.3 (2002), pp. 355–362.
- [122] Montagutelli, Xavier. Effect of the Genetic Background on the Phenotype of Mouse Mutations. *Journal of the American Society of Nephrology* 11.suppl 2 (2000), S101–S105.
- [123] Moss, Eric G, Lee, Rosalind C, and Ambros, Victor. The cold shock domain protein LIN-28 controls developmental timing in *C. elegans* and is regulated by the *lin-4* RNA. *Cell* 88.5 (1997), pp. 637–646.
- [124] Nimmo, Rachael A and Slack, Frank J. An elegant miRror: microRNAs in stem cells, developmental timing and cancer. *Chromosoma* 118.4 (2009), pp. 405–418.
- [125] Nimmo, Rachael, Antebi, Adam, and Woollard, Alison. *mab-2* encodes RNT-1, a *C. elegans* Runx homologue essential for controlling cell proliferation in a stem cell-like developmental lineage. *Development* 132.22 (2005), pp. 5043–5054.
- [126] Nimmo, Rachael and Woollard, Alison. Worming out the biology of Runx. *Developmental Biology* 313.2 (2008), pp. 492–500.
- [127] Oren-Suissa, Meital, Hall, David H, Treinin, Millet, Shemer, Gidi, and Podbilewicz, Benjamin. The fusogen EFF-1 controls sculpting of mechanosensory dendrites. *Science (New York, N.Y.)* 328.5983 (2010), pp. 1285–1288.
- [128] Oren-Suissa, Meital and Podbilewicz, Benjamin. Cell fusion during development. *Trends in Cell Biology* 17.11 (2007), pp. 537–546.

- [129] Oren-Suissa, Meital and Podbilewicz, Benjamin. Evolution of programmed cell fusion: Common mechanisms and distinct functions. *Developmental Dynamics* 239.5 (2010), pp. 1515–1528.
- [130] Paaby, Annalise B and Rockman, Matthew V. Cryptic genetic variation: evolution’s hidden substrate. *Nature Reviews Genetics* 15.4 (2014), pp. 247–258.
- [131] Paaby, Annalise B, White, Amelia G, Riccardi, David D, Gunsalus, Kristin C, Piano, Fabio, Rockman, Matthew V, and Flint, Jonathan. Wild worm embryogenesis harbors ubiquitous polygenic modifier variation. *eLife* 4 (2015), e09178.
- [132] Paix, Alexandre, Folkmann, Andrew, Rasoloson, Dominique, and Seydoux, Geraldine. High Efficiency, Homology-Directed Genome Editing in *Caenorhabditis elegans* Using CRISPR-Cas9 Ribonucleoprotein Complexes. *Genetics* 201.1 (2015), pp. 47–54.
- [133] Paix, Alexandre, Folkmann, Andrew, and Seydoux, Geraldine. Precision genome editing using CRISPR-Cas9 and linear repair templates in *C. elegans*. *Methods (San Diego, Calif.)* 121-122 (2017), pp. 86–93.
- [134] Petersen, Carola, Dirksen, Philipp, and Schulenburg, Hinrich. Why we need more ecology for genetic models such as *C. elegans*. *Trends in genetics : TIG* 31.3 (2015), pp. 120–127.
- [135] Podbilewicz, Benjamin. Membrane fusion as a morphogenetic force in nematode development. *Nematology* 2.1 (2000), pp. 99–111.
- [136] Podbilewicz, Benjamin. Cell fusion. *WormBook* (2006). DOI: doi/10.1895/wormbook.1.52.1.
- [137] Podbilewicz, Benjamin, Leikina, Evgenia, Sapir, Amir, Valansi, Clari, Suissa, Meital, Shemer, Gidi, and Chernomordik, Leonid V. The *C. elegans* developmental fusogen EFF-1 mediates homotypic fusion in heterologous cells and in vivo. *Developmental Cell* 11.4 (2006), pp. 471–481.
- [138] Podbilewicz, Benjamin and White, John G. Cell Fusions in the Developing Epithelia of *C. elegans*. *Developmental Biology* 161.2 (1994), pp. 408–424.

- [139] Pollard, Daniel A and Rockman, Matthew V. Resistance to germline RNA interference in a *Caenorhabditis elegans* wild isolate exhibits complexity and nonadditivity. *G3 (Bethesda, Md.)* 3.6 (2013), pp. 941–947.
- [140] Prior, Harriet, Jawad, Ali K, MacConnachie, Lauren, and Beg, Asim A. Highly Efficient, Rapid and Co-CRISPR-Independent Genome Editing in *Caenorhabditis elegans*. *G3 (Bethesda, Md.)* 7.11 (2017), pp. 3693–3698.
- [141] Queitsch, Christine, Carlson, Keisha D, and Girirajan, Santhosh. Lessons from model organisms: phenotypic robustness and missing heritability in complex disease. *PLoS Genetics* 8.11 (2012), e1003041.
- [142] Queitsch, Christine, Sangster, Todd a, and Lindquist, Susan. Hsp90 as a capacitor of phenotypic variation. *Nature* 417.6889 (2002), pp. 618–624.
- [143] Raj, Arjun and Oudenaarden, Alexander van. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* 135.2 (2008), pp. 216–226.
- [144] Raj, Arjun, Rifkin, Scott A, Andersen, Erik, and Oudenaarden, Alexander van. Variability in gene expression underlies incomplete penetrance. *Nature* 463.7283 (2010), pp. 913–918.
- [145] Rampelt, Heike, Kirstein-Miles, Janine, Nillegoda, Nadinath B, Chi, Kang, Scholz, Sebastian R, Morimoto, Richard I, and Bukau, Bernd. Metazoan Hsp70 machines use Hsp110 to power protein disaggregation. *The EMBO journal* 31.21 (2012), pp. 4221–4235.
- [146] Reinhart, Brenda J, Slack, Frank J, Basson, Michael, Pasquinelli, Amy E, Bettinger, Jill C, Rougvie, Ann E, Horvitz, H Robert, and Ruvkun, Gary. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 403.6772 (2000), pp. 901–906.
- [147] Ren, Haiyan and Zhang, Hong. Wnt signaling controls temporal identities of seam cells in *Caenorhabditis elegans*. *Developmental Biology* 345.2 (2010), pp. 144–155.

- [148] Rijnberk, Lotte M van, Horst, Suzanne E M van der, Heuvel, Sander van den, and Ruijtenberg, Suzan. A dual transcriptional reporter and CDK-activity sensor marks cell cycle entry and progression in *C. elegans*. *PLOS ONE* 12.2 (2017), e0171600.
- [149] Rocheleau, Christian E, Downs, William D, Lin, Rueyling, Wittmann, Claudia, Bei, Yoon-Hee, Cha, Y H, Ali, Mussa, Priess, James R, and Mello, Craig C. Wnt signaling and an APC-related gene specify endoderm in early *C. elegans* embryos. *Cell* 90.4 (1997), pp. 707–716.
- [150] Rocheleau, Christian E, Yasuda, Jun, Shin, Tae H, Lin, Rueyling, Sawa, Hitoshi, Okano, Hideyuki, Priess, James R, Davis, Roger J, and Mello, Craig C. WRM-1 activates the LIT-1 protein kinase to transduce anterior/posterior polarity signals in *C. elegans*. *Cell* 97.6 (1999), pp. 717–726.
- [151] Rockman, Matthew V and Kruglyak, Leonid. Breeding designs for recombinant inbred advanced intercross lines. *Genetics* 179.2 (2008), pp. 1069–1078.
- [152] Rockman, Matthew V and Kruglyak, Leonid. Recombinational landscape and population genomics of *Caenorhabditis elegans*. *PLoS Genetics* 5.3 (2009), e1000419.
- [153] Rossi, Andrea, Kontarakis, Zacharias, Gerri, Claudia, Nolte, Hendrik, Hölper, Soraya, Krüger, Marcus, and Stainier, Didier Y R. Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature* 524.7564 (2015), pp. 230–233.
- [154] Roy, Peter J, Zheng, Hong, Warren, Charles E, and Culotti, Joseph G. mab-20 encodes Semaphorin-2a and is required to prevent ectopic cell contacts during epidermal morphogenesis in *Caenorhabditis elegans*. *Development* 127.4 (2000), pp. 755–767.
- [155] Roy, Sarah H, Clayton, Joseph E, Holmen, Jenna, Beltz, Eleanor, and Saito, R Mako. Control of Cdc14 activity coordinates cell cycle and development in *Caenorhabditis elegans*. *Mechanisms of development* 128.7-10 (2011), pp. 317–326.
- [156] Rual, Jean-François, Ceron, Julian, Koreth, John, Hao, Tong, Nicot, Anne-Sophie, Hirozane-Kishikawa, Tomoko, Vandenhaute, Jean, Orkin, Stuart H, Hill, David E, Heuvel, Sander van den, and Vidal, Marc. Toward improving *Caenorhabditis elegans*

- phenome mapping with an ORFeome-based RNAi library. *Genome research* 14.10B (2004), pp. 2162–2168.
- [157] Rutherford, S L and Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* 396.6709 (1998), pp. 336–342.
- [158] Saito, R Mako, Perreault, Audrey, Peach, Bethan, Satterlee, John S, and Heuvel, Sander van den. The CDC-14 phosphatase controls developmental cell-cycle arrest in *C. elegans*. *Nature Cell Biology* 6.8 (2004), pp. 777–783.
- [159] Sapir, Amir, Choi, Jaebok, Leikina, Evgenia, Avinoam, Ori, Valansi, Clari, Chernomordik, Leonid V, Newman, Anna P, and Podbilewicz, Benjamin. AFF-1, a FOS-1-regulated fusogen, mediates fusion of the anchor cell in *C. elegans*. *Developmental Cell* 12.5 (2007), pp. 683–698.
- [160] Sawa, Hitoshi and Korswagen, Hendrik C. Wnt signaling in *C. elegans*. *WormBook* (2013), pp. 1–30.
- [161] Schindelin, Johannes, Arganda-Carreras, Ignacio, Frise, Erwin, Kaynig, Verena, Longair, Mark, Pietzsch, Tobias, Preibisch, Stephan, Rueden, Curtis, Saalfeld, Stephan, Schmid, Benjamin, Tinevez, Jean-Yves, White, Daniel James, Hartenstein, Volker, Eliceiri, Kevin, Tomancak, Pavel, and Cardona, Albert. Fiji: an open-source platform for biological-image analysis. *Nature Methods* 9.7 (2012), pp. 676–682.
- [162] Schulenburg, Hinrich and Félix, Marie-Anne. The Natural Biotic Environment of *Caenorhabditis elegans*. *Genetics* 206.1 (2017), pp. 55–86.
- [163] Schuler, Gregory D. Sequence mapping by electronic PCR. *Genome research* 7.5 (1997), pp. 541–550.
- [164] Scior, Annika, Buntru, Alexander, Arnsburg, Kristin, Ast, Anne, Iburg, Manuel, Juene-mann, Katrin, Pigazzini, Maria Lucia, Mlody, Barbara, Puchkov, Dmytro, Priller, Josef, Wanker, Erich E, Prigione, Alessandro, and Kirstein, Janine. Complete suppression of Htt fibrilization and disaggregation of Htt fibrils by a trimeric chaperone complex. *The EMBO journal* 37.2 (2018), pp. 282–299.

- [165] Segev, Nadav, Avinoam, Ori, and Podbilewicz, Benjamin. Fusogens. *Current biology : CB* 28.8 (2018), R378–R380.
- [166] Seidel, Hannah S, Ailion, Michael, Li, Jialing, Oudenaarden, Alexander van, Rockman, Matthew V, and Kruglyak, Leonid. A novel sperm-delivered toxin causes late-stage embryo lethality and transmission ratio distortion in *C. elegans*. *PLoS Biology* 9.7 (2011), e1001115.
- [167] Seidel, Hannah S, Rockman, Matthew V, and Kruglyak, Leonid. Widespread genetic incompatibility in *C. elegans* maintained by balancing selection. *Science (New York, N.Y.)* 319.5863 (2008), pp. 589–594.
- [168] Shaye, Daniel D and Greenwald, Iva. OrthoList: a compendium of *C. elegans* genes with human orthologs. *PLOS ONE* 6.5 (2011), e20085.
- [169] Shemer, Gidi and Podbilewicz, Benjamin. Fusomorphogenesis: Cell fusion in organ formation. *Developmental Dynamics* 218.1 (2000), pp. 30–51.
- [170] Shemer, Gidi and Podbilewicz, Benjamin. The story of cell fusion: Big lessons from little worms. *BioEssays : news and reviews in molecular, cellular and developmental biology* 25.7 (2003), pp. 672–682.
- [171] Shemer, Gidi, Suissa, Meital, Kolotuev, Irina, Nguyen, Ken C Q, Hall, David H, and Podbilewicz, Benjamin. EFF-1 Is Sufficient to Initiate and Execute Tissue-Specific Cell Fusion in *C. elegans*. *Current Biology* 14.17 (2004), pp. 1587–1591.
- [172] Shetty, Premnath, Lo, Miao-Chia, Robertson, Scott M, and Lin, Rueyling. *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 (2005), pp. 584–592.
- [173] Shinn-Thomas, Jessica H, Campo, Jacob J del, Wang, Jianjun, and Mohler, William A. The EFF-1A Cytoplasmic Domain Influences Hypodermal Cell Fusions in *C. elegans* But Is Not Dependent on 14-3-3 Proteins. *PLOS ONE* 11.1 (2016), e0146874.
- [174] Siegal, Mark L and Bergman, Aviv. Waddington’s canalization revisited: developmental stability and evolution. *Proceedings of the National Academy of Sciences of the United States of America* 99.16 (2002), pp. 10528–10532.

- [175] Silhánková, Marie, Jindra, Marek, and Asahina, Masako. Nuclear receptor NHR-25 is required for cell-shape dynamics during epidermal differentiation in *Caenorhabditis elegans*. *Journal of cell science* 118.Pt 1 (2005), pp. 223–232.
- [176] Sittig, Laura J, Carbonetto, Peter, Engel, Kyle A, Krauss, Kathleen S, Barrios-Camacho, Camila M, and Palmer, Abraham A. Genetic Background Limits Generalizability of Genotype-Phenotype Relationships. *Neuron* 91.6 (2016), pp. 1253–1259.
- [177] Sivasundar, Arjun and Hey, Jody. Population Genetics of *Caenorhabditis elegans*: The Paradox of Low Polymorphism in a Widespread Species. *Genetics* 163.1 (2003), pp. 147–157.
- [178] Slack, Frank J, Basson, Michael, Liu, Zhongchi, Ambros, Victor, Horvitz, H Robert, and Ruvkun, Gary. The *lin-41* RBCC gene acts in the *C. elegans* heterochronic pathway between the *let-7* regulatory RNA and the LIN-29 transcription factor. *Molecular Cell* 5.4 (2000), pp. 659–669.
- [179] Smurova, Ksenia and Podbilewicz, Benjamin. RAB-5- and DYNAMIN-1-Mediated Endocytosis of EFF-1 Fusogen Controls Cell-Cell Fusion. *Cell Reports* 14.6 (2016), pp. 1517–1527.
- [180] Solari, F, Bateman, A, and Ahringer, J. The *Caenorhabditis elegans* genes *egl-27* and *egr-1* are similar to MTA1, a member of a chromatin regulatory complex, and are redundantly required for embryonic patterning. *Development* 126.11 (1999), pp. 2483–2494.
- [181] Song, Kai, Li, Li, and Zhang, Guofan. Coverage recommendation for genotyping analysis of highly heterologous species using next-generation sequencing technology. *Scientific Reports* 6.1 (2016), p. 35736.
- [182] Sterken, Mark G, Snoek, L Basten, Kammenga, Jan E, and Andersen, Erik C. The laboratory domestication of *Caenorhabditis elegans*. *Trends in genetics : TIG* 31.5 (2015), pp. 224–231.
- [183] Stiernagle, Theresa. Maintenance of *C. elegans*. *WormBook* (2006). DOI: [doi/10.1895/wormbook.1.101.1](https://doi.org/10.1895/wormbook.1.101.1).

- [184] Sugioka, Kenji, Mizumoto, Kota, and Sawa, Hitoshi. Wnt regulates spindle asymmetry to generate asymmetric nuclear β -catenin in *C. elegans*. *Cell* 146.6 (2011), pp. 942–954.
- [185] Sulston, John E and Horvitz, H Robert. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Developmental Biology* 56.1 (1977), pp. 110–156.
- [186] Sulston, John E, Schierenberg, Einhard, White, John G, and Thomson, J Nichol. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Developmental Biology* 100.1 (1983), pp. 64–119.
- [187] Symmons, Orsolya and Raj, Arjun. What’s Luck Got to Do with It: Single Cells, Multiple Fates, and Biological Nondeterminism. *Molecular Cell* 62.5 (2016), pp. 788–802.
- [188] Takeshita, Hisako and Sawa, Hitoshi. Asymmetric cortical and nuclear localizations of WRM-1/ β -catenin during asymmetric cell division in *C. elegans*. *Genes & development* 19.15 (2005), pp. 1743–1748.
- [189] Thomas, Paul D and Kejariwal, Anish. Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects. *Proceedings of the National Academy of Sciences* 101.43 (2004), pp. 15398–15403.
- [190] Thompson, Owen A, Snoek, L Basten, Nijveen, Harm, Sterken, Mark G, Volkers, Rita J M, Brenchley, Rachel, Hof, Arjen van’t, Bevers, Roel P J, Cossins, Andrew R, Yanai, Itai, Hajnal, Alex, Schmid, Tobias, Perkins, Jaryn D, Spencer, David, Kruglyak, Leonid, Andersen, Erik C, Moerman, Donald G, Hillier, LaDeana W, Kammenga, Jan E, and Waterston, Robert H. Remarkably Divergent Regions Punctuate the Genome Assembly of the *Caenorhabditis elegans* Hawaiian Strain CB4856. *Genetics* 200.3 (2015), pp. 975–989.
- [191] Tijsterman, Marcel, Okihara, Kristy L, Thijssen, Karen, and Plasterk, Ronald H A. PPW-1, a PAZ/PIWI protein required for efficient germline RNAi, is defective in a natural isolate of *C. elegans*. *Current biology : CB* 12.17 (2002), pp. 1535–1540.

- [192] Tischler, Julia, Lehner, Ben, Chen, Nansheng, and Fraser, Andrew G. Combinatorial RNA interference in *Caenorhabditis elegans* reveals that redundancy between gene duplicates can be maintained for more than 80 million years of evolution. *Genome biology* 7.8 (2006), R69.
- [193] Tremblay, Mathieu, Sanchez-Ferras, Oraly, and Bouchard, Maxime. GATA transcription factors in development and disease. *Development* 145.20 (2018), dev164384.
- [194] Trent, Carol, Tsuing, Nancy, and Horvitz, H Robert. Egg-laying defective mutants of the nematode *Caenorhabditis elegans*. *Genetics* 104.4 (1983), pp. 619–647.
- [195] Visser, J Arjan G M de, Hermisson, Joachim, Wagner, Günter P, Ancel Meyers, Lauren, Meyers, Lauren Ancel, Bagheri-Chaichian, Homayoun, Blanchard, Jeffrey L, Chao, Lin, Cheverud, James M, Elena, Santiago F, Fontana, Walter, Gibson, Greg, Hansen, Thomas F, Krakauer, David, Lewontin, Richard C, Ofria, Charles, Rice, Sean H, Dassow, George von, Wagner, Andreas, and Whitlock, Michael C. Perspective: Evolution and detection of genetic robustness. *Evolution* 57.9 (2003), pp. 1959–1972.
- [196] Vu, Victoria, Verster, Adrian J, Schertzberg, Michael, Chuluunbaatar, Tungalag, Spensley, Mark, Pajkic, Djina, Hart, G Traver, Moffat, Jason, and Fraser, Andrew G. Natural Variation in Gene Expression Modulates the Severity of Mutant Phenotypes. *Cell* 162.2 (2015), pp. 391–402.
- [197] Waddington, Conrad H. Canalization of development and the inheritance of acquired characters. *Nature* (1942).
- [198] Wagner, Andreas. Distributed robustness versus redundancy as causes of mutational robustness. *BioEssays : news and reviews in molecular, cellular and developmental biology* 27.2 (2005), pp. 176–188.
- [199] Wall, Jeffrey D, Tang, Ling Fung, Zerbe, Brandon, Kvale, Mark N, Kwok, Pui-Yan, Schaefer, Catherine, and Risch, Neil. Estimating genotype error rates from high-coverage next-generation sequence data. *Genome research* 24.11 (2014), pp. 1734–1739.

- [200] Wang, Han, Park, Heenam, Liu, Jonathan, and Sternberg, Paul W. An Efficient Genome Editing Strategy To Generate Putative Null Mutants in *Caenorhabditis elegans* Using CRISPR/Cas9. *G3 (Bethesda, Md.)* 8.11 (2018), pp. 3607–3616.
- [201] Ward, Jordan D, Bojanala, Nagagireesh, Bernal, Teresita, Ashrafi, Kaveh, Asahina, Masako, and Yamamoto, Keith R. Sumoylated NHR-25/NR5A Regulates Cell Fate during *C. elegans* Vulval Development. *PLoS Genetics* 9.12 (2013), e1003992.
- [202] Waring, David A, Wrischnik, Lisa, and Kenyon, Cynthia. Cell signals allow the expression of a pre-existent neural pattern in *C. elegans*. *Development* 116.2 (1992), pp. 457–466.
- [203] Whitacre, James Michael. Biological robustness: paradigms, mechanisms, and systems principles. *Frontiers in Genetics* 3 (2012), p. 67. DOI: 10.3389/fgene.2012.00067. URL: <https://www.frontiersin.org/article/10.3389/fgene.2012.00067>.
- [204] Wicks, Stephen R, Yeh, Raymond T, Gish, Warren R, Waterston, Robert H, and Plasterk, Ronald H. Rapid gene mapping in *Caenorhabditis elegans* using a high density polymorphism map. *Nature genetics* 28.2 (2001), pp. 160–164.
- [205] Wightman, Bruce, Ha, Ilha, and Ruvkun, Gary. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* 75.5 (1993), pp. 855–862.
- [206] Wildwater, Marjolein, Sander, Nicholas, Vreede, Geert de, and Heuvel, Sander van den. Cell shape and Wnt signaling redundantly control the division axis of *C. elegans* epithelial stem cells. *Development* 138.20 (2011), pp. 4375–4385.
- [207] Wood, Andrew R., Esko, Tonu, Yang, Jian, Vedantam, Sailaja, Pers, Tune H., Gustafsson, Stefan, Chu, Audrey Y., Estrada, Karol, Luan, Jian'an, Kutalik, Zoltán, Amin, Najaf, Buchkovich, Martin L., Croteau-Chonka, Damien C., Day, Felix R., Duan, Yanan, Fall, Tove, Fehrmann, Rudolf, Ferreira, Teresa, Jackson, Anne U., Karjalainen, Juha, Lo, Ken Sin, Locke, Adam E., Mägi, Reedik, Mihailov, Evelin, Porcu, Eleonora, Randall, Joshua C., Scherag, André, Vinkhuyzen, Anna A. E., Westra, Harm-Jan, Winkler, Thomas W., Workalemahu, Tsegaselassie, Zhao, Jing Hua, Absher, Devin, Albrecht,

Eva, Anderson, Denise, Baron, Jeffrey, Beekman, Marian, Demirkan, Ayse, Ehret, Georg B., Feenstra, Bjarke, Feitosa, Mary F., Fischer, Krista, Fraser, Ross M., Goel, Anuj, Gong, Jian, Justice, Anne E., Kanoni, Stavroula, Kleber, Marcus E., Kristiansson, Kati, Lim, Unhee, Lotay, Vaneet, Lui, Julian C., Mangino, Massimo, Leach, Irene Mateo, Medina-Gomez, Carolina, Nalls, Michael A., Nyholt, Dale R., Palmer, Cameron D., Pasko, Dorota, Pechlivanis, Sonali, Prokopenko, Inga, Ried, Janina S., Ripke, Stephan, Shungin, Dmitry, Stancáková, Alena, Strawbridge, Rona J., Sung, Yun Ju, Tanaka, Toshiko, Teumer, Alexander, Trompet, Stella, Laan, Sander W. van der, Setten, Jessica van, Van Vliet-Ostaptchouk, Jana V., Wang, Zhaoming, Yengo, Loïc, Zhang, Weihua, Afzal, Uzma, Ärnlöv, Johan, Arscott, Gillian M., Bandinelli, Stefania, Barrett, Amy, Bellis, Claire, Bennett, Amanda J., Berne, Christian, Blüher, Matthias, Bolton, Jennifer L., Böttcher, Yvonne, Boyd, Heather A., Bruinenberg, Marcel, Buckley, Brendan M., Buyske, Steven, Caspersen, Ida H., Chines, Peter S., Clarke, Robert, Claudi-Boehm, Simone, Cooper, Matthew, Daw, E. Warwick, De Jong, Pim A., Deelen, Joris, Delgado, Graciela, Denny, Josh C., Dhonukshe-Rutten, Rosalie, Dimitriou, Maria, Doney, Alex S. F., Dörr, Marcus, Eklund, Niina, Eury, Elodie, Folkersen, Lasse, Garcia, Melissa E., Geller, Frank, Giedraitis, Vilmantas, Go, Alan S., Grallert, Harald, Grammer, Tanja B., Gräßler, Jürgen, Grönberg, Henrik, Groot, Lisette C. P. G. M. de, Groves, Christopher J., Haessler, Jeffrey, Hall, Per, Haller, Toomas, Hallmans, Goran, Hannemann, Anke, Hartman, Catharina A., Hassinen, Maija, Hayward, Caroline, Heard-Costa, Nancy L., Helmer, Quinta, Hemani, Gibran, Henders, Anjali K., Hillege, Hans L., Hlatky, Mark A., Hoffmann, Wolfgang, Hoffmann, Per, Holmen, Oddgeir, Houwing-Duistermaat, Jeanine J., Illig, Thomas, Isaacs, Aaron, James, Alan L., Jeff, Janina, Johansen, Berit, Johansson, Åsa, Jolley, Jennifer, Juliusdottir, Thorhildur, Junttila, Juhani, Kho, Abel N., Kinnunen, Leena, Klopp, Norman, Kocher, Thomas, Kratzer, Wolfgang, Lichtner, Peter, Lind, Lars, Lindström, Jaana, Lobbens, Stéphane, Lorentzon, Mattias, Lu, Yingchang, Lyssenko, Valeriya, Magnusson, Patrik K. E., Mahajan, Anubha, Maillard, Marc, McArdle, Wendy L., McKenzie, Colin A., McLachlan, Stela, McLaren, Paul J., Menni, Cristina, Merger, Sigrun, Milani, Lili, Moayyeri, Alireza, Monda, Keri L., Morken, Mario A., Müller,

Gabriele, Müller-Nurasyid, Martina, Musk, Arthur W., Narisu, Narisu, Nauck, Matthias, Nolte, Ilja M., Nöthen, Markus M., Oozageer, Laticia, Pilz, Stefan, Rayner, Nigel W., Renstrom, Frida, Robertson, Neil R., Rose, Lynda M., Roussel, Ronan, Sanna, Serena, Scharnagl, Hubert, Scholtens, Salome, Schumacher, Fredrick R., Schunkert, Heribert, Scott, Robert A., Sehmi, Joban, Seufferlein, Thomas, Shi, Jianxin, Silventoinen, Karri, Smit, Johannes H., Smith, Albert Vernon, Smolonska, Joanna, Stanton, Alice V., Stirrups, Kathleen, Stott, David J., Stringham, Heather M., Sundström, Johan, Swertz, Morris A., Syvänen, Ann-Christine, Tayo, Bamidele O., Thorleifsson, Gudmar, Tyrer, Jonathan P., Dijk, Suzanne van, Schoor, Natasja M. van, Velde, Nathalie van der, Heemst, Diana van, Oort, Floor V. A. van, Vermeulen, Sita H., Verweij, Niek, Vonk, Judith M., Waite, Lindsay L., Waldenberger, Melanie, Wennauer, Roman, Wilkens, Lynne R., Willenborg, Christina, Wilsgaard, Tom, Wojczynski, Mary K., Wong, Andrew, Wright, Alan F., Zhang, Qunyuan, Arveiler, Dominique, Bakker, Stephan J. L., Beilby, John, Bergman, Richard N., Bergmann, Sven, Biffar, Reiner, Blangero, John, Boomsma, Dorret I., Bornstein, Stefan R., Bovet, Pascal, Brambilla, Paolo, Brown, Morris J., Campbell, Harry, Caulfield, Mark J., Chakravarti, Aravinda, Collins, Rory, Collins, Francis S., Crawford, Dana C., Cupples, L. Adrienne, Danesh, John, Faire, Ulf de, Ruijter, Hester M. den, Erbel, Raimund, Erdmann, Jeanette, Eriksson, Johan G., Farrall, Martin, Ferrannini, Ele, Ferrières, Jean, Ford, Ian, Forouhi, Nita G., Forrester, Terrence, Gansevoort, Ron T., Gejman, Pablo V., Gieger, Christian, Golay, Alain, Gottesman, Omri, Gudnason, Vilmundur, Gyllensten, Ulf, Haas, David W., Hall, Alistair S., Harris, Tamara B., Hattersley, Andrew T., Heath, Andrew C., Hengstenberg, Christian, Hicks, Andrew A., Hindorff, Lucia A., Hingorani, Aroon D., Hofman, Albert, Hovingh, G. Kees, Humphries, Steve E., Hunt, Steven C., Hypponen, Elina, Jacobs, Kevin B., Jarvelin, Marjo-Riitta, Jousilahti, Pekka, Jula, Antti M., Kaprio, Jaakko, Kastelein, John J. P., Kayser, Manfred, Kee, Frank, Keinanen-Kiukaanniemi, Sirkka M., Kiemeny, Lambertus A., Kooner, Jaspal S., Kooperberg, Charles, Koskinen, Seppo, Kovacs, Peter, Kraja, Aldi T., Kumari, Meena, Kuusisto, Johanna, Lakka, Timo A., Langenberg, Claudia, Le Marchand, Loic, Lehtimäki, Terho, Lupoli, Sara, and Madden,

- Pamela A. F. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics* 46.11 (2014), pp. 1173–1186. ISSN: 1546-1718. DOI: 10.1038/ng.3097. URL: <https://doi.org/10.1038/ng.3097>.
- [208] Woollard, Alison. Gene duplications and genetic redundancy in *C. elegans*. *WormBook* (2005). DOI: [doi/10.1895/wormbook.1.2.1](https://doi.org/10.1895/wormbook.1.2.1).
- [209] Wrischnik, Lisa A and Kenyon, Cynthia J. The role of *lin-22*, a hairy/enhancer of split homolog, in patterning the peripheral nervous system of *C. elegans*. *Development* 124.15 (1997), pp. 2875–2888.
- [210] Xia, Dan, Zhang, Yuxia, Huang, Xinxin, Sun, Yinyan, and Zhang, Hong. The *C. elegans* CBF β homolog, BRO-1, regulates the proliferation, differentiation and specification of the stem cell-like seam cell lineages. *Developmental Biology* 309.2 (2007), pp. 259–272.
- [211] Yamamoto, Yuko, Takeshita, Hisako, and Sawa, Hitoshi. Multiple Wnts Redundantly Control Polarity Orientation in *Caenorhabditis elegans* Epithelial Stem Cells. *PLoS Genetics* 7.10 (2011), e1002308.
- [212] Yang, Yihong, Zhang, Yan, Li, Wen-Jun, Jiang, Yuxiang, Zhu, Zhiwen, Hu, Huifang, Li, Wei, Wu, Jia-Wei, Wang, Zhi-Xin, Dong, Meng-Qiu, Huang, Shanjin, and Ou, Guangshuo. Spectraplakins Induces Positive Feedback between Fusogens and the Actin Cytoskeleton to Promote Cell-Cell Fusion. *Developmental Cell* 41.1 (2017), 107–120.e4.
- [213] Zhang, Tingting, Sun, Yinyan, Tian, E, Deng, Hansong, Zhang, Yuxia, Luo, Xin, Cai, Qingchun, Wang, Huayi, Chai, Jijie, and Zhang, Hong. RNA-binding proteins SOP-2 and SOR-1 form a novel PcG-like complex in *C. elegans*. *Development* 133.6 (2006), pp. 1023–1033.
- [214] Zhao, Pei, Zhang, Zhe, Lv, Xiaoying, Zhao, Xiang, Suehiro, Yuji, Jiang, Yinan, Wang, Xinquan, Mitani, Shohei, Gong, Haipeng, and Xue, Ding. One-step homozygosity in precise gene editing by an improved CRISPR/Cas9 system. *Cell research* 26.5 (2016), pp. 633–636.
- [215] Zheng, Rena and Blobel, Gerd A. GATA Transcription Factors and Cancer. *Genes & Cancer* 1.12 (2011), pp. 1178–1188.

Appendix

Appendix A

Resources

A.1 List of strains

Table A.1: Strains used in this thesis

Strain	Genotype	Genetic background
JR667	<i>unc-119(e2498::Tc1) III; wIs51[SCMp::GFP + unc-119(+)] V</i>	N2
MBA19	<i>icbIR1(V, N2>JU2007); wIs51[SCMp::GFP + unc-119(+)] V</i>	JU2007
MBA21	<i>wIs51; msc(icb4)</i>	N2
MBA79	<i>wIs51; eff-1(icb4) II (4x BC)</i>	N2
MBA171	<i>201508EMS20c; wIs51[SCMp::GFP + unc-119(+)] V</i>	N2
MBA190	<i>201508EMS23c; wIs51[SCMp::GFP + unc-119(+)] V</i>	N2
MBA191	<i>201508EMS50; wIs51[SCMp::GFP + unc-119(+)] V</i>	N2
MBA205	<i>201508EMS29(icb61); wIs51[SCMp::GFP + unc-119(+)] V</i>	N2
MBA156	<i>eff-1(hy21)II; wIs51[SCMp::GFP + unc-119(+)] V</i>	N2
MBA159	<i>eff-1(op55)II; wIs51[SCMp::GFP + unc-119(+)] V</i>	N2
MBA173	<i>bro-1(icb44)</i>	N2
MBA250	<i>icbIs2[arf-3::GFP:CAAX::unc-54]I; icbS12[dpy-7::mCherry::H2B::unc-54 3'UTR+cb-unc-119] IV; wIs51[SCMp::GFP + unc-119(+)] V</i>	N2
MBA252	<i>icbIs2[arf-3::GFP:CAAX::unc-54]I; eff-1(icb4) II; icbS12[dpy-7::mCherry::H2B::unc-54 3'UTR+cb-unc-119] IV; wIs51[SCMp::GFP + unc-119(+)] V</i>	N2
MBA290	<i>egl-18(ga97) IV; wIs51[SCMp::GFP + unc-119(+)] V</i>	N2
AW298	<i>wIs78[scm::gfp + qjm-1::gfp + unc-119+] IV; him-5(e1490) V</i>	N2
MBA187	<i>eff-1(icb4) II; wIs78[scm::gfp + qjm-1::gfp + unc-119+] IV</i>	N2
MBA235	<i>icbIR3(V, N2>JU2519); wIs51[SCMp::GFP + unc-119(+)] V</i>	N2
MBA237	<i>icbIs3[arf-3::GFP:CAAX::unc-54] III; wIs51[scm::gfp] V</i>	N2
MBA243	<i>bro-1(icb45); wIs51[SCMp::GFP + unc-119(+)] V</i>	JU2519
MBA244	<i>bro-1(icb46); wIs51[SCMp::GFP + unc-119(+)] V</i>	N2
MBA246	<i>eff-1(icb4) II; icbIs3[arf-3::GFP:CAAX::unc-54]III; wIs51[scm::gfp]V</i>	N2
MBA248	<i>icbIR4(V, N2>JU775); wIs51[SCMp::GFP + unc-119(+)] V</i>	JU775
MBA251	<i>eff-1(icb4) II; icbS12[dpy-7::mCherry::H2B::unc-54 3'UTR+cb-unc-119] IV; wIs51[SCMp::GFP + unc-119(+)] V</i>	N2
MBA256	<i>icbIR2(V, N2>CB4856); wIs51[SCMp::GFP + unc-119(+)] V</i>	CB4856
MBA267	<i>egl-18(ok290)IV; wIs51[scm::gfp]V</i>	N2
MBA268	<i>lin-22(qcb49) IV; egl1[dat-1::gfp]IV; wIs51[SCMp::GFP + unc-119(+)] V</i>	N2
MBA269	<i>lin-22(icb50) IV; egl1[dat-1::gfp]IV; wIs51[SCMp::GFP + unc-119(+)] V</i>	N2
MBA271	<i>lin-22(qcb49) IV; vtl1 V</i>	CB4856
MBA272	<i>bro-1(icb47) I; wIs51[SCMp::GFP + unc-119(+)] V</i>	JU2007
MBA273	<i>eff-1(ok1021) II; wIs51[SCMp::GFP + unc-119(+)] V</i>	N2
MBA274	<i>bro-1(icb45) I; wIs51[SCMp::GFP + unc-119(+)] V</i>	CB4856
MBA275	<i>lin-22(qcb52) IV; vtl1 V</i>	CB4856
MBA333	<i>icbIR9(IV, N2>JU2007); icbIR1(V, N2>JU2007); egl-18(ok290) IV; wIs51[scm::gfp + unc-119+] V</i>	JU2007
MBA368	<i>icbIR10(II, N2>JU2007); icbIR1(V, N2>JU2007); eff-1(icb4); wIs51[SCMp::GFP + unc-119(+)] V</i>	JU2007
MBA374	<i>icbExc116[pdpy-7::egl-18::unc-54 3'UTR]; wIs51[SCMp::GFP + unc-119(+)] V</i>	N2
MBA411	<i>icbIR7(IV; N2>CB4856); icbIR2(V; N2>CB4856); egl-18(ok290) IV; wIs51[scm::gfp + unc-119+] V</i>	CB4856

Table A.1: Strains used in this thesis

Strain	Genotype	Genetic background
MBA412	<i>icbIR8(IV;N2>JU2007); icbIR1(V;N2>JU2007); egl-18(ga97) IV; wsls51 [scm::gfp + unc-119+] V</i>	JU2007
MBA413	<i>icbIR6(IV;N2>CB4856); icbIR2(V;N2>CB4856); egl-18(ga97) IV; wsls51 [scm::gfp + unc-119+] V</i>	CB4856
MBA459	<i>icbIs2[arf-3::GFP::CAA X::unc-54]II; egl-1(icb4) II; egl-18(ga97) IV; wsls51 [SCMP::GFP+unc-119(+)] V</i>	N2
MBA499	<i>ycIs11[odr-1::gfp;phlh-3::vab-10-abd::venus;phlh-3::mIs::tdTOMATO] V; icbEx125 [qjm-1::mcherry::unc-54; dpy-7::gfp::unc-54]</i>	N2
MBA500	<i>egl-1(icb4); ycIs11 [odr-1::gfp;phlh-3::vab-10-abd::venus;phlh-3::mIs::tdTOMATO] V; icbEx125 [qjm-1::mcherry::unc-54; dpy-7::gfp::unc-54]</i>	N2
MBA782	<i>icbIR15(III;CB4856>N2); wsls51 [scm::gfp + unc-119+] V</i>	N2
MBA783	<i>icbIR16(X;CB4856>N2); wsls51 [scm::gfp + unc-119+] V</i>	N2
MBA784	<i>icbIR13(II;CB4856>N2); egl-18(ga97) IV; wsls51 [scm::gfp + unc-119+] V</i>	N2
MBA785	<i>icbIR14(II;CB4856>N2); egl-18(ga97) IV; wsls51 [scm::gfp + unc-119+] V</i>	N2
MBA786	<i>icbIR15(III;CB4856>N2); egl-18(ga97) IV; wsls51 [scm::gfp + unc-119+] V</i>	N2
MBA790	<i>icbIR15(III;CB4856>N2); icbIR16(X;CB4856>N2); egl-18(ga97) IV; wsls51 [scm::gfp + unc-119+] V</i>	N2
MBA789	<i>icbIR22(II;CB4856>N2); icbIR16(X;CB4856>N2); egl-18(ga97) IV; wsls51 [scm::gfp + unc-119+] V</i>	N2
MBA819	<i>icbIR19(III;CB4856>N2); egl-18(ga97) IV; wsls51 [scm::gfp + unc-119+] V</i>	N2
MBA840	<i>icbIs16 [arf-3::GFP::unc-54]</i>	CB4856
MBA841	<i>icbIs18 [arf-3::GFP::unc-54]</i>	CB4856
MBA846	<i>icbIR21(II;CB4856>N2); egl-18(ga97) IV; wsls51 [scm::gfp + unc-119+] V</i>	N2
MBA848	<i>icbIR22(II;CB4856>N2); icbIR15(III;CB4856>N2); icbIR16(X;CB4856>N2); egl-18(ga97) IV; wsls51 [scm::gfp + unc-119+] V</i>	N2
MBA862	<i>icbIR22(II;CB4856>N2); icbIR15(III;CB4856>N2); egl-18(ga97) IV; wsls51 [scm::gfp + unc-119+] V</i>	N2
MBA938	<i>icbIR22(II;CB4856>N2); icbIR31(III;CB4856>N2); icbIR16(X;CB4856>N2); wsls51 [scm::gfp + unc-119+] V</i>	N2
MBA939	<i>icbIR22(II;CB4856>N2); icbIR15(III;CB4856>N2); wsls51 [scm::gfp + unc-119+] V</i>	N2
MBA944	<i>icbIR18(II;CB4856>N2); egl-18(ga97) IV; wsls51 [scm::gfp + unc-119+] V</i>	N2
MBA945	<i>icbIR27(III;CB4856>N2); egl-18(ga97) IV; wsls51 [scm::gfp + unc-119+] V</i>	N2
MBA946	<i>icbIR15(III;CB4856>N2); icbIR16(X;CB4856>N2); wsls51 [scm::gfp + unc-119+] V</i>	N2
MBA947	<i>icbIR22(II;CB4856>N2); icbIR16(X;CB4856>N2); wsls51 [scm::gfp + unc-119+] V</i>	N2
MBA949	<i>icbIR22(II;CB4856>N2); wsls51 [scm::gfp + unc-119+] V</i>	N2
MBA951	<i>icbIR23(II;CB4856>N2); egl-18(ga97) IV; wsls51 [scm::gfp + unc-119+] V</i>	N2
MBA963	<i>icbIR22(II;CB4856>N2); egl-18(ga97) IV; wsls51 [scm::gfp + unc-119+] V</i>	N2
MBA964	<i>icbIR30(III;CB4856>N2); egl-18(ga97) IV; wsls51 [scm::gfp + unc-119+] V</i>	N2
MBA971	<i>icbIR26(V; N2>XZ1516); wsls51 [scm::gfp + unc-119+] V</i>	XZ1516
MBA1005	<i>icbIR33(II;CB4856>N2); egl-18(ga97) IV; wsls51 [scm::gfp + unc-119+] V</i>	N2
MBA1006	<i>icbIR34(II;CB4856>N2); egl-18(ga97) IV; wsls51 [scm::gfp + unc-119+] V</i>	N2
MBA1012	<i>icbIR35(II;CB4856>N2); egl-18(ga97) IV; wsls51 [scm::gfp + unc-119+] V</i>	N2
MBA1056	<i>icbIR31(III;CB4856>N2); egl-18(ga97) IV; wsls51 [scm::gfp + unc-119+] V</i>	N2

A.2 Lists of primers

Table A.2: Cloning and sequencing primers

Oligo	Sequence(5'-3')
<i>pes-10_egl-18_F</i>	ttgcttgagggtaccgagtttaaacatttATGTCGATCAGCATAATGAC
<i>unc-54_egl-18_R</i>	gtaattggacttagaagtcagaggcaatttTAAAAATGTGGCACTGCTAT
<i>dpy-7_egl-18_F</i>	acattttgtccagataagtttaaacatttATGTCGATCAGCATAATGAC
<i>dpy-10CrispF1</i>	GCTACCATAGGCACCACGAGGTTTTAGAGCTAGAAATAGCAAGTTA
<i>dpy-10CrispR1</i>	CTCGTGGTGCCATGGTAGCAAACATTTAGATTTGCAATTCAATTATATAG
<i>dpy-10repair</i>	CACTTGAACCTCAATACGGCAAGATGAGAATGACTGGAAACCGTACCGCAT GCGGTGCCATGGTAGCGGAGCTTCACATGGCTTCAGACCAACAGCCTAT
<i>dpy-10_seqF</i>	GTCAGATGATCTACCGGTGTGTCCAC
<i>dpy-10_seqR</i>	GTCTCTCCTGGTGCTCCGTCTTCAC
<i>egl-18CRISPF1</i>	AATGATGCAATTATATCAAGTTTTAGAGCTAGAAATAGCAAGTTA
<i>egl-18CRISPR1</i>	TTGATAATAATTGCATCATTAAACATTTAGATTTGCAATTCAATTATATAG
<i>egl-18CRISPF2</i>	GGAGCGATCCGATATCCCGAGTTTTAGAGCTAGAAATAGCAAGTTA
<i>egl-18CRISPR2</i>	TCGGGATATCGGATCGTCCAAACATTTAGATTTGCAATTCAATTATATAG
<i>egl-18repair2</i>	CAGGATGTGAAACAGGAAGAGTCGGAGTGATCCGATATCCCCACTGCTACAG AGGCGCAGAACCCTTTTGGATGCACTAACAGCACAGTTTAGCAGCAACG
<i>egl-18seqF1</i>	AATGACGAAAACGAGACCAG
<i>egl-18seqR1</i>	AAAGTTCTGCGCCTCTGTA
<i>bro-1crispF1</i>	aatcaatatacctgtcaagtGTTTTAGAGCTAGAAATAGCAAGTTA
<i>bro-1crispR1</i>	acttgacaggtatattgattAAACATTTAGATTTGCAATTCAATTATATAG
<i>bro-1repair</i>	tgtgtgtttcagacttctgAaagacgggaatcaatatacctgtcaagCtTgaatgtcccaactaatggaaatgaggtgggttagtataaatgttg
<i>bro1seqF1</i>	AATGCCCTTGGTGAGTGTTTC
<i>bro1seqF2</i>	GGCCCTAATGAAAATGTGGTT
<i>nhr-25CRISPF1</i>	GTTTTGTGGTGATCGAGTCTCGTTTTAGAGCTAGAAATAGCAAGTTA
<i>nhr-25CRISPR1</i>	GAGACTCGATCACCACAAACAAACATTTAGATTTGCAATTCAATTATATAG
<i>nhr-25repair</i>	ggttctgacccaatcatgaagcgagatgtgccggtttgtggCgatcgTgtGtctggatcactacggcTtctgacgtgtgaaagttgcaaggtt
<i>nhr25seqF1</i>	TGACTGACGTCGAGAGGATG
<i>nhr25seqR1</i>	GAGGCACTTCTGGAATCGAC
<i>rnt-1CRISPF1</i>	AGCAAAAGTGCATCGACAAGGTTTTAGAGCTAGAAATAGCAAGTTA
<i>rnt-1CRISPR1</i>	CTTGTCGATGCACTTTTGTCTAAACATTTAGATTTGCAATTCAATTATATAG
<i>rnt-1repair</i>	gacgaaaaaccaTGAGAAGAAGTTCGAAATGAGAAAGCAAAAGTGCATCGGC AGGTCGCAAAATTTAACGATTTGCGGTTTGTTCGGGCGGTCCGGCAGAG
<i>rnt-1seqF1</i>	ATACACAGCTCTTCCGAAGCAT
<i>rnt1seqF2</i>	CCAACCGGTTATTTGGCTAC
<i>lin-22gRNAF</i>	ACTGAAATTGAATCCGATGGGTTTTAGAGCTAGAAATAGCAAGTTA
<i>lin-22gRNAR</i>	CCATCGGATTCATTTTCAGTAAACATTTAGATTTGCAATTCAATTATATAG
<i>lin22-23F</i>	ATGACGTCATTCCTGTGCTCCGA
<i>lin22-22R</i>	GTAACAAGATTCACAGGATGCG
<i>eff-1gRNAF</i>	CCACACTGTTCCAAGACACCGTTTTAGAGCTAGAAATAGCAAGTTA
<i>eff-1gRNAR</i>	GGTGTCTTGGAAACAGTGTGGAAACATTTAGATTTGCAATTCAATTATATAG
<i>eff-1_seqF</i>	CCTAGCAGTTACAACCTACGG
<i>eff-1_seqR</i>	CGGATGGTGATGTTCAAGCT
<i>nath-10_snp_F</i>	GCCGGGAACGAGGAAAAGTCAAATG
<i>nath-10_snp_R</i>	TTCCGACTCACTGTTCC
<i>wrn-1_RNAi_F</i>	AGACCGGCAGATCTGATATCATCGATGCCGGATCTGTGAATGAGGA
<i>wrn-1_RNAi_R</i>	TCGACGGTATCGATAAGCTTGATATCGCGACCATTTGTTTCGTTTGGC
<i>utp-20_RNAi_F</i>	AGACCGGCAGATCTGATATCATCGATGACGATGAAGACGAAAGTGTGTTGCC
<i>utp-20_RNAi_R</i>	TCGACGGTATCGATAAGCTTGATATCGCTGCATGCTGCTCCAACCTCAA

Table A.2 continued from previous page

Oligo	Sequence(5'-3')
<i>hsp-110_WB2894_F</i>	TTCTCCCACTTTCCGTGTTTCGT
<i>hsp-110_WB2894_R</i>	CTCAGCTGGAACCTCCTCGACA
<i>sor-1_WB8058_F</i>	ATCCCAAGTGGCTACTGCGATT
<i>sor-1_WB8058_R</i>	TGGAACCCAACGGTAACCTTGA

Table A.3: Fluorescent labelled oligos used in smFISH

<i>eff-1</i>	<i>nhr-73</i>	<i>elt-6</i>
aactggggagaccactcaaa	tatttcattactcggctctcc	cttgtgaagctgagcgcac
aatccgtaggttgtaactgc	ctgtgacaaacttggcagga	ttttgaaacatcctgtcga
atcgaatttctcctcgagtg	aaagtacagggcgggttcggt	taagagcttcaagcctccg
tgtcttggacagtggtg	aagctgcgcaggaggtgatg	aggacttttcttttact
gagatgttgagcacggaca	cggatcttcggaagaatgca	gctctcggacatttgctcaa
gaattgcatttgattcctc	cagtgacatgtagcgaatc	tgatgggggatacttctcct
atctgaagcagactgcagtg	gaaatagtgcagctgttagt	gagaagcacggctttcagtt
tcattgatctcttgggatgc	tgcatgctctgcagaagaac	attcgactttccgactgatg
atgtctgatttcagcattt	ggctcgaatacaactggtgt	acagattctttggtgggttt
gttcaagctttccaatcga	cattacggctcatgaccatc	tccattctccaaatgtcgat
aagtgtaccgctgagttatc	gcttttctagcagtagctaa	ttctcaaaaggcagctact
tggcatgaacttcagggatt	ccgtggtattatgttgttcg	gcgtcgaaatattccttgtg
tggcatcacactcacagata	gttacttctcaagtgacg	aaacggagagctcgcggaaa
agattctgcggtacatgtg	tgttcaggatcttctgtaac	atcgaacacaaacccttga
agactctggacaagcggtaa	acttgagcaaatccccgttg	tattttgggattggccatg
gcggtagcatgaagatgttt	ctctccacttctttcacata	gcacaaggtttagcaagttc
tgggtctgatttgggaaga	atatctctgctgagccatta	gcggcttgatgttgttgtg
tcgaacgtcacagcaaagct	tggagagcgttcttttctc	ctggtgagcatggtgatgag
gtttgactgcgaggaatgct	aagatcgttactgacttca	tttgactcttcggttttac
atgttgcatacgtttaggt	cttggtaggcgggtgacaatc	attggtttctgttctccgac
tttatcttttccaccaat	tgaagcacgattccagggat	ctgctcaacagacgcacttg
tgtgttccaccatctaattg	acagtatatctactccaggg	tagagagctgatcgagcaga
cgacgttttggctcgagatg	cacaacctcaacgttctcat	atggtgatttcccgttgaac
ggcagttacagccaatgaaa	aatccagggttctcgagaact	ggtggatgcgtgtactgttg
cagttgatgagatgctcgtc	gaagtaggtgcaaggagtga	gacgagttggcacttgtatc
ctccattactgttcttgag	gttacagctaggatccagtg	gcaattggagcatttcgaga
tcaatggttgcattctcagt	ggatgaactcagagattctt	acgctgtcgttttattgtg
accatccaagacggtcaaag	cccagagttgacttgagtag	cgcgttacaacaagcttcc
atgaccagaatcgtccatc	ttcagggttaagctgggcaag	cctatgcagcctataataca

APPENDIX A. RESOURCES

Table A.3: Fluorescent labelled oligos used in smFISH

<i>eff-1</i>	<i>nhr-73</i>	<i>elt-6</i>
tccattttcacaactccatt	gaaggctgcgtattcaatga	atcttttcgcatgtgaaccg
gcaatttttacttttagcct	gcttcagatgcaaaatgac	tacggaatcgctgttgata
tcgagcagattgaatccacg	atgcttgtttgtgatacac	atcttcatcttccctcattc
gtgttacaacagcttgtctg	ctgagccacaatcttcattg	aaaacggctgcttgactggt
tgaagattagttccttcggc	gcacttctaactccttcata	cggtaggccgagaagttgac
gacaaggttttgactttcca	gttcttctataataacttcc	tttgctcgagaaaggtcagc
gatggatccactgaagtca	ccatttccaattcactcatt	cctgtgactgattcaactga
agcctcatatactgtcaagt	tgtgatatccccgattcttg	
ctgatccatcaatttttcca	tcgatacagtgatgatttgc	
tccaaatccagttgacatct	cgagccatgtcattgtagag	
atggctggcagtggaataat	aagcccagtttgatgataca	
tgaatctgctcggagacaga		
tttcaaagggtcttcgaaa		
ccattttccctcaacaagat		
attacaagttgggcaggttc		
gatccaatgagctggattca		
caccaatcgaactgattcct		
atccgtagcaatcataacgc		
gacaaatttgtccaacgga		

Table A.4: List of genetic markers based on SNPs in CB4856 genome

PRIMER NAME	TYPE	CHR.	LOCATION	FORWARD PRIMER	REVERSE PRIMER	AMPLICON N2/CB4856	GENE	VARIATION
<i>dsh-2</i> -splice-framshift	SNPs	II	5078640 – 5079750	ACTCCCACCTCGCTTGACTTCTC	TGGAATGACGCCTTCAGGACAT	1110	dsh-2	WBVar02048904 WBVar01372400 WBVar01372411
<i>dsh-2</i> -WBVar00223936	SNP	II	5080117 – 5080704	TTTGACGATGGTGATGTGGCG	TTGCGAGATGTCAAACGGGAAGC	588	dsh-2	WBVar00223936
<i>dgk-5</i> -6458520	SNP	II	6458186 – 6458750	GGGGAGGTGTTTTCTCGGATGA	ATGGGATTGTTCCACCGGCATA	565	dgk-5	WBVar00172647
<i>wrn1.II</i> :6558315-6558953	SNPs	II	6558315 – 6558953	GCCAACTTGAGACCTAOCCTCGT	CGACTCCCATAAGCTTCAACCCA	639	wrn-1	WBVar01372570 WBVar01372571
<i>C56C10.9</i> -6591468	SNP	II	6591209 – 6591640	GCACAGGGTCTCGTGGAATTG	TCGAGOGACAGTTTCTCGTCAT	432	C56C10.9	WBVar00550941
<i>F54H5.14</i> -6625176	SNP	II	6624909 – 6625371	CCTCTGCTCATATATCCTGTACCGT	ACAAAGACCAACCAAGAGACACGT	463	F54H5.14	WBVar01372585
<i>T19D12.10</i> -6658746	SNP	II	6658451 – 6658863	TCGGTCCACTTGGTGTCACTTT	GTGAGTGGCACTGACGGTATCA	413	T19D12.10	WBVar00172687
<i>C18H9.3</i> -6686689	SNP	II	6686380 – 6687037	GATCGTCCAAAAGCGGTGAATC	CGCCTCTCTGAACCTCGAAGTGA	658	C18H9.3	WBVar00225066
<i>dpy2.II</i> :6715373-6716050	SNP	II	6715373 – 6716050	CAGGAACACCTGGAGAGAAGGG	TCTCCGAAAACAGTTACGGACGT	678	dpy-2	WBVar01372609
<i>egl-27</i> -WBVar00172876	SNP	II	6845564 – 6846083	ACACCACCAACCTCTGACAA	TTAAATCGGTTTCGTGCACACCG	520	egl-27	WBVar00172876
<i>egl-27</i> -CB-SNV-DEL	SNPs, indel	II	6846431 – 6847523	GCAGCATTCCAAACATCAAATGGC	ACAGAATTGAACGAAATGCACGCT	1093	egl-27	WBVar01393720 WBVar00104370 WBVar00104369 WBVar00225208 WBVar00225210 WBVar00104371
								WBVar00172889
<i>dnyj-5</i> -WBVar00172889	SNP	II	6853929 – 6854238	CGAGCACACACTCAAAATGTCC	ATCCTCCTCACGTGGTTTCACAG	310	dnyj-5	WBVar00172889
<i>T07F8.1</i> -WBVar01413585	SNP	II	7047269 – 7047582	CTGGCACTAAAACCAGCGTCAG	ACCAGTGTACCATCCAAGAGCC	633	T07F8.1	WBVar01413585
<i>fbn1</i> -00061000-563999	SNPs	III	7634256 – 7634696	CCGTCTACACCAAGTACGACGT	TTCGGACGTTGACTCGACAGAA	441	fbn-1	WBVar00061000 WBVar00563999

Table A.4: List of genetic markers based on SNPs in CB4856 genome

PRIMER NAME	TYPE	CHR.	LOCATION	FORWARD PRIMER	REVERSE PRIMER	AMPLICON N2/CB4856	GENE	VARIATION
SNPs- <i>lin-36</i>	SNPs	III	8017488 – 8018685	AAGAAGCAGCAACGAAGAAGCC	GTCCTTTTCGCCCAAACAGACGAG	1198	<i>lin-36</i>	WBVar01952398 WBVar01566934 WBVar000065079
<i>sort-1</i> - WBVar00068058	SNP	III	8412989 – 8413577	ATCCCAAAGTGGCTACTGCGATT	TGGAACCCCAACGGTAACCTTGA	589	<i>sort-1</i>	WBVar00068058
<i>sort-1</i> -WB	SNPs	III	8422218 – 8422391	TGGGATGAGGCAGATGTCAAGA	AGAAAGTCGTTGGTTCGGTTCTGT	425	<i>sort-1</i>	WBVar00095026 WBVar00068068
<i>hsp-110</i> - WBVar01962894	SNP	III	8445106 – 8445536	TTCTCCCACTTTCCCGTGTTCGT	CTCAGCTGGAACCTTCCTCGACA	431	<i>hsp-110</i>	WBVar01962894
<i>zfp-1</i> - WBVar00245081	SNP	III	8804664 – 8805037	TGGCTGAAGTGAACACACTGGA	ATGTTCAGAAGAGCTTGTGCGC	502	<i>zfp-1</i>	WBVar00245081
<i>lin-9</i> - WBVar01331763	SNP	III	8903249 – 8903712	CTCGCAAAGGCAGCTCTTTGATT	AGAAGAACTCGCACATGACCCA	464	<i>lin-9</i>	WBVar01331763

Table A.5: List of genetic markers based on snip-SNPs in CB4856 genome

PRIMER NAME/ GENE	snip-SNP	CHR.	LOCATION	FORWARD PRIMER	REVERSE PRIMER	AMPLICON N2	AMPLICON CB4856	RESTRICTION ENZYME	VARIATION
<i>del-10</i>	snip-SNP	II	6481307 – 6481881	AATTCCCAGCTTCGTTGGGAT	ACATCGTGACCCCGTAGAAAGCA	575	157/418	PvuII	WBVar00172651
<i>T13C2.3</i>	snip-SNP	II	6788927 – 6789383	TCCACACTATTTCCCTCGTG	GAGCAATCAAGAACCCGGATC	494	300/126	DraI	WBVar00172772
<i>wrm-1</i>	snip-SNP	III	5722786 – 5723367	CCCCTCATCAATGGCGTTTTTCT	GTAACGCGTAATTGAGCCAGTG	214/368	582	PsyI	WBVar 00061487
<i>clp-1</i>	snip-SNP	III	7998068 – 7998553	TGCCCACTTTTGCTGATTGGAA	AGTGTAACAAAATCCAGTCAGGT	244/172	486	AluI/ HindIII	WBVar00067338

APPENDIX A. RESOURCES

Table A.6: List of genetic markers based on indels in CB4856 genome

PRIMER NAME	TYPE	LOCATION	FORWARD PRIMER	REVERSE PRIMER	AMPLICON N2	AMPLICON CB4856	AMPLICON SIZE DIFFERENCE	GENE
II:1.08	deletion	1082937 – 1083261	AATTCCTCCGCGGAAATTTGTGCT	CAGGGGATCTGGTTCTCCTGATGT	325	183	142	K02E7.1
II:2.88	deletion	2882104 – 2882391	CCGCTTTTGGAGCATAATTTGGC	CCCGAGCAAATTTTGTGTGCC	288	No band	288	pigw-1
II:3.04	deletion	3046437 – 3046767	TGTACCGCTTCAGACGTGAACA	AGGGCAGAGTTCTCAAACGTGAT	331	274	57	sri-30
II:4.02	deletion	4026939 – 4027160	ACTCAAACATCGGAGTGCATGA	CCTAGCTTTTCCACGAAACCAAGT	222	142	80	intergenic region
II:4.48	deletion	4488502 – 4489297	TTTCGCCCCCGATTAAATGGC	AAGTCTCATTTGCCATCCGTCCA	796	726	70	T05C1.3
II:4.78	deletion	4783042 – 4783783	GGTATTTTGGCCACGGCTGAAA	TAAACTTGTCAAACACTGGGGGCC	742	70	672	F26G1.12
II:5.30	deletion	5304326 – 5304673	GTGAGACCCATTTCCAGGGATA	AGACGCCGAAAAGAATTTGGA	348	257	91	intergenic region
II:5.75	deletion	5753884 – 5754521	GGGACAAGGAGGACAAGTTGA	GGTCTGGCGTTGGAATTTGGT	638	575	63	ifb-2, F10C1.8
II:6.44	deletion	6442580 – 6443043	GCCCAACACTCAGTGGAAACATG	TATGTTGAGCTTCCGCATGTGC	464	406	58	cutl-16
II:6.84	deletion	6843749 – 6844800	TGACCTTTTCGAGCTCCATTTGGT	AGAGCATPAGCGGGATGATGACC	1052	929	123	egl-27
II:7.18	deletion	7180442 – 7180779	TCCAAATTCAGAAAGCTTAACTCAGA	TGAACTGTCAAGAGCGCAACAT	338	212	126	intergenic region
II:7.50	deletion	7506230 – 7507336	GGCTCCAGGAAATTTCCCAAAGT	TCCGCCAGTCTTCAATGTGAT	1107	882	225	C28F5.1, g1b-9
II:7.92	deletion	7922472 – 7923156	AACTCAACGCCATTTGCAGATG	GGACTTCGAAGGACACGTCGTA	685	607	78	spv-1
II:8.01	deletion	8015740 – 8016130	CCCAGTTTCATTTCAOCCGTGC	TGCAGATCAGGAGAGTTCTGCA	395	346	49	intergenic region
II:8.18	deletion	8186973 – 8187303	TTTAGTGACAAAACCTCCGCGCG	CCTCAGTGGCGAATTCACAGAGA	332	267	65	cct-1
II:9.91	deletion	9917609 – 9918287	TAGCGCTGTTGGTATATGGCGA	GTCAAAGTCCCAATGTGCCAAGG	679	386	293	C05C10.8, pho-10
II:11.33	deletion	11339188 – 11339428	GCCTTTGTTCGAGACGCTGAAAA	CATGATAAGCACATGGCCCTGC	241	177	64	rmd-3
II:12.44	deletion	12441031 – 12441293	AATCTCTCTGTAGCTGCTCCC	CGAACTGCCGAGTCCGTATCAT	263	207	56	Y57A10C.10
III:5.46	deletion	5469726 – 5470227	GATGAGAGGCCGCTGTGACAGAT	GGTGGTAGGTAGGAAGGCCACAT	502	376	126	intergenic region
III:5.74	deletion	5747166 – 5747466	CAGCCGGTGTGCAATAAAATCCT	CACGACACTCTCGACACGCTGTA	301	212	89	B0244.5
III:6.04	deletion	6040642 – 6040911	TGCGTGTCAAACCTGAGAGAGCAT	AGCAAAGAGGTTCTAGGTCCTCGT	270	210	60	F40H6.2
III:6.42	insertion	6421171 – 6421587	AACGGGGTGTGATAGAGAGCGCA	GGCTGGGACTCATCTTTTGGTCC	417	494	-77	atgl-1
III:6.67	insertion	6670172 – 6670671	AATTTCTGCCTTGATGCCGAGC	CCACTATACTGGGCGGTCAACA	249	379	-130	intergenic region
III:6.73	deletion	6736563 – 6736870	TGACAGGGACATATGGCACCA	CAAAACATGCCGAATGAAGATCA	500	309	191	Y102E9.6
III:7.03	insertion	7033926 – 7034327	GGAGAACCCAGGACCCTCCATAT	GTCCATTCACCTGGTCCGCAAG	405	482	-77	intergenic region
III:7.29	deletion	7290537 – 7290905	TGTGTCCGACATTTCAATTTGTGGA	CTAGGGGTCTCTTGTCTCTCTCC	369	245	124	algn-11
III:7.54	deletion	7542457 – 7543107	CGTTTCTGCGAACGGACGATAG	GGCACAATCGGAACTTTTGTGCA	651	424	227	intergenic region
III:7.66	deletion	7669365 – 7669861	GCGACGGAATCCAAAGATTTCT	GTAACATCTTCGCCATCCTCCA	497	374	123	hido-18
III:7.77	deletion	7774203 – 7774599	AAAACTGTTGGCCAGGCTGTTC	GCGCGCACTTGCATTTTTGTTA	397	318	79	intergenic region
III:8.26	insertion	8269291 – 8269608	ACGAATTTTGTGCTGCCCTT	TTAACCATCTGGCTCAATCCGG	318	371	-53	crn-7
III:8.58	deletion	8589045 – 8589413	CTTCATTTTCCGGACACGGAGG	GGAGGCAAGAAAGTGTGGAGGAT	369	292	77	C06E1.12, fip-3
III:8.91	deletion	8915504 – 8915891	CCGTTTGAGTATGTTGTGCCACA	AAAGTGCACACTTCTTTTGCA	388	305	83	trxt-2

Table A.6: List of genetic markers based on indels in CB4856 genome

PRIMER NAME	TYPE	LOCATION	FORWARD PRIMER	REVERSE PRIMER	AMPLICON N2	AMPLICON CB4856	AMPLICON SIZE DIFFERENCE	GENE
III:9.34	deletion	9344002 – 9344391	GCAAACACATGGTGGAGCACAT	GCTTATCGGTTCTCGGCTTGAC	390	301	89	emb-9
III:9.59	deletion	9593454 – 9593743	ATTTGGAAGTTGGCTGGGCATG	ACGCAGAAAAACGGAAACAAGCT	281	175	106	rom-2
V:18.66	deletion	18662060 – 18662822	GCCCGAAAAATTCACTCGAGACC	CCAAATTTCCAGCCACCAACCTC	763	284	479	Y69H2.3
V:19.02	deletion	19024949 – 19025536	GACAATAACGCCCAACGGTCAG	TGTATCAAGCAATGGCCCAAGG	588	511	77	intergenic region
V:19.59	deletion	19596561 – 19597042	ACGGCAGAGTGGCTATGGAATT	AATTTGCCAGCAGTGGGATTG	482	349	133	kem-1, Y43F8C.t1
V:20.09	deletion	20091314 – 20092009	GCGTCCAAATTCACGTGGAACAT	GTTTCAGCAGGCCCAAAATTG	696	583	113	sec-23
X:3.00	deletion	3005547 – 3005910	CAAAAGGCACTGAAGGTGGACC	TGTTTGGACGTAAACGGAGGTGT	364	264	100	intergenic region
X:4.14	deletion	4149534 – 4149944	TAGCCTCTTGCCAGCCTTCATT	AGCAACAGCCGTGGTATTTTGG	411	252	159	nck-1
X:5.08	deletion	5087928 – 5088442	TTCGATGCCTGTTCACAAAACCG	GGTCGTTCCCTTTTGAGCAATGGG	515	342	173	pnk-4

Appendix B

Additional tables

B.1 ANOVA

table

Table B.1: ANOVA table pertaining to cell division angles in *eff-1(icb4)* compared to wild type

Cell division angle	Cell	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Sig
a-p	H1	1	3.66	1.25	2.93	0.01	**
a-p	H2	1	5.78	1.99	2.9	0.01	**
a-p	V1	1	250.62	250.63	6.75	0.02	*
a-p	V2	1	901.23	901.23	34.17	2.43E-06	***
a-p	V3	1	648.93	648.93	40.43	9.82E-07	***
a-p	V4	1	621.99	621.99	21.68	7.09E-05	***
a-p	V6	1	19.09	19.09	4.1	0.05	.
aa-ap	V1-V4, V6	1	951.1	951.15	11.01	0.0012	**
ap-pa	V1-V4, V6	1	8352.7	8352.7	85.56	5.59E-15	***
pa-pp	V1-V4, V6	1	2941.8	2941.84	17.68	5.82E-05	***
a-p versus ap-pa in mutant	V1-V4, V6	1	3572.8	3572.8	34.9	3.26E-08	***
a-p versus ap-pa in wt	V1-V4, V6	1	1.76	1.76	0.2	0.66	ns

APPENDIX B. ADDITIONAL TABLES

Table B.2: PCA table for seam cell shape

CELL	PC1	PC2	PERCENT OF VARIANCE EXPLAINED
H0	63.53	26.11	89.64
H1	82.76	15.48	98.24
H2	72.19	25.49	97.68
V1 – V4, V6	71.4	25.7	97.1
V5	65.95	28.84	94.79
T	55.68	40.96	96.64

B.2 PCA

table

B.3 List

of

RILs

Table B.3: List of RILs

RIL	Number of generations	mean SCN Trial-1	var SCN Trial-1	mean SCN Trial-2	var SCN Trial-2	mean SCN Trial-3	var SCN Trial-3	low bulk	high bulk	low bulk stringent	high bulk stringent
1	10	13.68	5.92	13.03	4.03						
2	10	12.75	7.53	13.05	4.15						
3	10	12.55	4.97	10.55	9.79						
4	12	12.28	6.26	12.23	6.08	12.38	7.47				
5	12	13.05	6.05	13.1	7.78						
6	10	13.83	5.79	13.9	3.43				selected		selected
7	12	13.65	4.28	13.28	3.95	13.23	7.67		selected		
8	12	12.3	5.19	13.18	6.71						
9	10	14.93	1.87	14.3	4.01				selected		selected
10	14	12.78	6.44	11.55	8.15						
11	10	14.2	4.32	14.25	3.83				selected		selected
12	14	14.4	5.27	14.38	4.04				selected		selected
13	10	11.88	6.27	12.93	4.84						
14	10	13.3	4.78	12.48	7.64						
15	12	13.08	2.53	12.45	4.51						
16	12	13.48	6.97	13.53	6.15				selected		
18	12	12.55	6.97	11.75	6.91						
19	12	13.4	6.09	13.63	7.52				selected		
20	12	13.83	2.76	14.05	10.72	14.93	2.38		selected		selected
21	10	10.55	7.33	9.55	7.74			selected		selected	
22	10	13.33	3.51	12.83	3.64						
23	12	12.45	6	10.23	8.74						
24	12	12.25	7.53	12.33	9.25						
25	12	12.5	5.23	11.2	6.57						
26	12	12.63	2.91	12.68	5.4						
27	12	14.15	3.46	13.98	4.38	13.19	9.65		selected		selected
28	10	14.18	6.15	14.8	2.63				selected		selected
29	14	11.98	8.23	11.63	5.37	11.2	9.24	selected			
30	12	12.4	6.91	12.05	10.41						
31	12	12.98	6.59	13.85	3.52						
32	14	13.23	6.03	12.88	7.29						

APPENDIX B. ADDITIONAL TABLES

Table B.3: List of RILs

RIL	Number of generations	mean SCN Trial-1	var SCN Trial-1	mean SCN Trial-2	var SCN Trial-2	mean SCN Trial-3	var SCN Trial-3	low bulk	high bulk	low bulk stringent	high bulk stringent
33	12	12.88	5.29	13.43	4.05						
34	12	13.95	5.18	13.43	8.1	12.63	9.73		selected		
35	12	11.6	11.12	10.98	8.13			selected		selected	
36	12	11.85	4.85	11.85	7.11			selected			
37	12	12.08	9.2	11.48	14.15	12.93	9.81				
38	12	12.85	5.93	12.63	5.73	10.7	8.78				
39	12	9.08	12.02	10.85	8.13	9.53	16.2	selected		selected	
40	12	12.18	9.28	13.3	5.14						
41	10	12.88	6.93	12.73	11.64						
42	12	11.88	6.57	11.1	3.99	12.35	6.8	selected			
43	10	10.88	12.01	11.05	8.82			selected		selected	
44	12	14.38	2.24	14	5.28				selected		selected
45	14	12.65	8.39	12.88	8.57						
46	12	12.9	8.81	13.1	6.4	12.68	8.17				
47	12	13.3	5.65	12.98	6.49						
48	14	13.45	4.92	11.15	6.23	9.94	17.86				
49	12	13.03	5.87	12.23	8.38	12.33	9.97				
50	12	10.63	13.68	9.5	12.41			selected		selected	
51	12	13.1	5.73	11.95	10.05	11.53	8.01				
52	14	11.6	4.25	11.65	7.05	9.68	8.94	selected			
53	12	12.98	5.97	13	7.13	13.18	4.3				
54	12	12.95	6.92	11.55	7.74						
55	10	13	4.31	11.88	9.7						
56	12	13.63	9.57	13.9	7.58				selected		
57	12	13.98	4.13	12.7	4.73						
58	12	11.78	9.05	11.58	7.17	11.88	5.91	selected			
59	12	12.98	4.33	12.98	5.82						
60	12	13.2	5.34	12.75	7.47						
61	12	12.55	4.31	13.25	1.58						
62	14	13.4	4.3	12.68	7.15						
63	10	13.2	5.55	13.4	3.37	12.85	5.46		selected		
64	12	11.23	8.64	10.6	9.07			selected		selected	
65	12	11.33	6.69	10.4	9.12	10.95	6.2	selected		selected	
66	12	13.28	3.9	12	8	11.7	8.01				
67	12	13.25	5.17	13.15	9.62						
68	10	10	9.33	11.85	8.49	11.83	5.58	selected			
69	12	12.7	8.63	11.55	8.41						
70	12	12.65	4.18	11.33	8.94						
71	10	13.55	6.66	12.53	5.95						
72	14	12.43	5.89	9.6	10.45						
73	10	13.65	4.85	14.28	4.05	14.5	6.82		selected		selected
74	10	12.38	4.65	12.7	4.57	11.88	9.75				
75	12	12.68	7.35	13.4	5.53						
76	14	11.23	6.28	11.85	4.08			selected			
77	12	12.68	7.56	12.28	7.69	10.75	12.35				
78	14	12.63	8.09	10.8	8.57						
79	12	13.43	5.43	13.93	3.71				selected		
80	12	12.83	4.51	13.48	6						
81	12	12.75	4.76	14.33	2.22	13.83	3.84				

APPENDIX B. ADDITIONAL TABLES

Table B.3: List of RILs

RIL	Number of generations	mean SCN Trial-1	var SCN Trial-1	mean SCN Trial-2	var SCN Trial-2	mean SCN Trial-3	var SCN Trial-3	low bulk	high bulk	low bulk stringent	high bulk stringent
82	12	11.18	7.84	11.73	8.05			selected			
83	12	14.23	2.44	13	6.26						
84	12	13.53	4.82	12.58	6.92						
85	12	13.98	3.92	13.95	4.36				selected		selected
86	14	11.43	7.02	11.03	7.97			selected		selected	
87	10	14.1	2.55	13.78	6.13	14.08	5.1		selected		
88	10	13.8	3.91	12.78	6.28						
89	12	13.73	5.64	12.25	8.35						
90	12	12.3	4.27	11.85	5.31						
91	14	13.65	4.8	13.95	3.28				selected		
92	12	13.5	6.62	12.55	8.15						
93	12	14.18	2.97	13.1	6.71						
94	12	12.9	6.71	13.3	5.29	11.33	11.4				
95	12	9.92	10.91	11.25	8.86			selected		selected	
96	12	11.98	5.1	11.2	10.57	10.4	7.53	selected			
97	14	11.78	7	10.68	6.12	9.4	9.02	selected			
98	12	14.1	4.19	13.73	7.18				selected		
99	12	12.28	5.49	10.88	5.8	11.85	5.82				
100	14	13	4.26	11.6	11.22						
101	12	12.15	6.44	10.4	5.27						
102	10	13.35	4.34	13.53	2.87	12.53	6.92				
103	12	13.35	7.46	14.73	2.46				selected		
104	12	13.6	3.89	13.45	4.72				selected		
105	14	11.78	5.46	10.53	12.51			selected			
106	10	12.28	5.33	11.23	5.77						
107	12	12.95	4.31	11.95	5.84	10.63	12.19				
108	12	11.35	8.44	12.6	9.27						
109	14	11.6	6.81	11.4	7.78			selected			
110	12	12.85	6.49	14.33	5.1						
111	12	13.63	4.91	12.25	8.45						
112	14	12.5	5.38	12.13	5.14						
113	12	13.5	3.23	13.85	3.52	12.28	7.85		selected		
114	14	12.8	4.32	12.73	3.79	12.48	6.41				
115	12	11.55	11.23	11.08	10.94	10.6	13.63	selected		selected	
116	12	14.05	5.43	13.65	4.28	13.25	3.01		selected		
117	12			12.73	6.31	13.18	5.79				
Number of lines in bulk								22	24	10	10

B.4 List of genes with natural variation on chromosome II

APPENDIX B. ADDITIONAL TABLES

Table B.4: List of genes with natural variation in the genomic interval chr. II (4794415 - 8173823)

CHR	POS	REF	ALT	effect	gene_name	feature_id	nt_change	aa_change
II	4794415	G	T	missense_variant	F26G1.1	F26G1.1	c.2009G>T	p.Ser670Ile
II	4800611	A	G	missense_variant	rig-6	C33F10.5c.1	c.373T>C	p.Ser125Pro
II	4800881	T	A	missense_variant	rig-6	C33F10.5c.1	c.103A>T	p.Asn35Tyr
II	4821518	A	G	missense_variant	lact-5	C33F10.7a	c.47T>C	p.Leu16Ser
II	4841643	T	C	missense_variant	T23B7.2	T23B7.2	c.391A>G	p.Thr131Ala
II	4862774	C	CT	frameshift_variant	mpst-4	F11G11.9	c.583dupA	p.Arg195fs
II	4862775	T	TG	frameshift_variant	mpst-4	F11G11.9	c.582_583insC	p.Arg195fs
II	4875399	T	G	missense_variant	F11G11.13	F11G11.13	c.55A>C	p.Ile191Leu
II	4892454	C	G	missense_variant	msp-33	R05F9.8	c.241G>C	p.Gly81Arg
II	4892487	A	G	missense_variant	msp-33	R05F9.8	c.208T>C	p.Phe70Leu
II	4892640	C	T	missense_variant	msp-33	R05F9.8	c.55G>A	p.Val19Met
II	4917423	G	A	missense_variant	aagr-2	R05F9.12	c.2264C>T	p.Ser755Leu
II	4937554	A	T	missense_variant	ZK546.5	ZK546.5	c.1614A>T	p.Gln538His
II	4956498	A	G	missense_variant	ulp-2	Y38A8.3b.1	c.1694T>C	p.Val565Ala
II	4987232	A	C	missense_variant	shc-2	T27F7.2c.2	c.395A>C	p.Asp132Ala
II	4996160	T	G	missense_variant	rnh-1.0	F59A6.6a	c.375T>G	p.Asn125Lys
II	5001662	G	A	missense_variant	F59A6.5	F59A6.5	c.2040G>A	p.Met680Ile
II	5010996	C	A	missense_variant	F59A6.3	F59A6.3	c.2228C>A	p.Pro743Gln
II	5038751	C	G	missense_variant	golg-2	F33G12.5.2	c.1048G>C	p.Gln350Gln
II	5080428	C	G	missense_variant	dsh-2	C27A2.6	c.601G>C	p.Asp201His
II	5082264	G	GATGTCTCATCGA	conservative_inframe_insertion	C04G6.4	C04G6.4	c.802_803ins TCGATGAGACAT	p.Thr267_Ser268ins PheAspGluThr
II	5082265	A	ATGTCTCATCGAC, ATGTCTCATCGAT, ATGTCCGATCGAC, ATGTCTCATCGAG	conservative_inframe_insertion	C04G6.4	C04G6.4	c.801_802ins ATCGATGAGACA	p.Thr267_Ser268ins IleAspGluThr
II	5088670	T	C	missense_variant	p1d-1	C04G6.3	c.2206A>G	p.Asn736Asp
II	5103413	GAA	GA,G	frameshift_variant	C04G6.11	C04G6.11	c.322delT	p.Phe108fs
II	5106264	G	A	missense_variant	mpk-2	C04G6.1c.2	c.805G>A	p.Gln269Lys
II	5110150	C	G	missense_variant	F09C12.2	F09C12.2	c.122C>G	p.Ala41Gly
II	5112072	C	A	missense_variant	F09C12.2	F09C12.2	c.1097C>A	p.Ala366Asp
II	5157019	G	T	missense_variant	msp-45	F58A6.8	c.65C>A	p.Ala22Glu
II	5162042	C	T	missense_variant	srb-5	C27D6.6	c.991G>A	p.Asp331Asn
II	5162814	C	A	missense_variant	srb-5	C27D6.6	c.308G>T	p.Cys103Phe
II	5165618	C	G	missense_variant	srb-3	C27D6.8	c.723G>C	p.Arg241Ser
II	5165841	G	A	missense_variant	srb-3	C27D6.8	c.500C>T	p.Pro167Leu
II	5172093	A	T	missense_variant	C27D6.1	C27D6.1	c.133A>T	p.Thr45Ser
II	5172375	T	C	missense_variant	C27D6.1	C27D6.1	c.332T>C	p.Leu111Pro
II	5173456	G	A	missense_variant	C27D6.1	C27D6.1	c.976G>A	p.Val326Ile

APPENDIX B. ADDITIONAL TABLES

Table B.4: Genes with natural variation in the genomic interval chr. II (4794415 – 8173823)

CHR	POS	REF	ALT	effect	gene_name	feature_id	nt_change	aa_change
II	5176828	A	G	missense_variant	C27D6.1	C27D6.1	c.2111A>G	p.Lys704Arg
II	5194511	T	C	missense_variant	msp-49	C34F11.6	c.146T>C	p.Ile49Thr
II	5196142	T	A	missense_variant	C34F11.5	C34F11.5	c.238T>A	p.Phe80Ile
II	5196623	A	G	missense_variant	C34F11.5	C34F11.5	c.514A>G	p.Thr172Ala
II	5196927	C	A	missense_variant	C34F11.5	C34F11.5	c.715C>A	p.Leu239Met
II	5197239	T	C	missense_variant	C34F11.5	C34F11.5	c.988T>C	p.Val313Ala
II	5197289	G	C	missense_variant	C34F11.5	C34F11.5	c.988G>C	p.Val330Leu
II	5197697	A	G	missense_variant	C34F11.5	C34F11.5	c.1198A>G	p.Lys400Glu
II	5198024	G	A	missense_variant	C34F11.5	C34F11.5	c.1525G>A	p.Glu509Lys
II	5204185	A	G	missense_variant	ampd-1	C34F11.3d	c.301A>G	p.Ile101Val
II	5204487	T	A	missense_variant& splice_region_variant	ampd-1	C34F11.3d	c.510T>A	p.Asn170Lys
II	5210320	G	C	missense_variant	C34F11.8	C34F11.8	c.314C>G	p.Thr105Arg
II	5210685	T	C	missense_variant& splice_region_variant	C34F11.8	C34F11.8	c.97A>G	p.Asn33Asp
II	5212297	T	C	missense_variant	C34F11.2	C34F11.2	c.5T>C	p.Leu2Pro
II	5213614	C	A	missense_variant	C34F11.1	C34F11.1	c.446C>A	p.Ala149Asp
II	5213616	C	G	missense_variant	C34F11.1	C34F11.1	c.448C>G	p.Leu150Val
II	5215677	C	A	missense_variant	dsh-1	C34F11.9f	c.1269G>T	p.Met423Ile
II	5216446	T	C	missense_variant	dsh-1	C34F11.9f	c.840A>G	p.Ile280Met
II	5225281	T	C	missense_variant	dsh-1	C34F11.9a	c.88A>G	p.Arg30Gly
II	5230122	T	G	stop_gained	EGAP2.2	EGAP2.2	c.42T>G	p.Tyr14*
II	5230229	A	G	missense_variant	EGAP2.2	EGAP2.2	c.71A>G	p.Glu24Gly
II	5230294	C	CAAAG	frameshift_variant	EGAP2.2	EGAP2.2	c.136_137ins AAAG	p.Leu46fs
II	5231594	G	A	missense_variant	EGAP2.1	EGAP2.1	c.163G>A	p.Asp55Asn
II	5235792	T	G	missense_variant	pho-1	EGAP2.3.2	c.41A>C	p.Glu144Ala
II	5237804	A	G	missense_variant	pho-4	T16D1.2.2	c.434T>C	p.Val145Ala
II	5245301	A	G	missense_variant	ptr-6	C54A12.1	c.971T>C	p.Met324Thr
II	5312181	G	A	missense_variant	gcy-12	F08B1.2a	c.62G>A	p.Arg21Gln
II	5315809	C	A	missense_variant	gcy-12	F08B1.2a	c.2558C>A	p.Pro853Gln
II	5317538	C	T	missense_variant	gcy-12	F08B1.2a	c.3785C>T	p.Ala1262Val
II	5336425	A	G	missense_variant	vhp-1	F08B1.1b.2	c.119A>G	p.Asn40Ser
II	5348469	C	A	missense_variant	F09E5.10	F09E5.10	c.1535G>T	p.Cys512Phe
II	5350094	G	GATTC	frameshift_variant&stop_gained	F09E5.10	F09E5.10	c.292_295dup GAAT	p.Ser99fs
II	5372895	G	A	missense_variant	agt-2	F09E5.13	c.352C>T	p.Pro118Ser
II	5374410	C	T	missense_variant	F09E5.3	F09E5.3	c.347C>T	p.Ala116Val
II	5389831	A	C	missense_variant	ndx-6	EED8.8	c.7T>G	p.Phe3Val

Table B.4: Genes with natural variation in the genomic interval chr. II (4794415 – 8173823)

CHR	POS	REF	ALT	effect	gene_name	feature_id	nt_change	aa_change
II	5393928	C	T	missense_variant	pink-1	EEED8.9	c.1465G>A	p.Val1489Met
II	5396567	T	A	missense_variant	EEED8.10	EEED8.10a	c.1787A>T	p.Lys596Met
II	5398402	C	T	missense_variant	EEED8.10	EEED8.10a	c.277G>A	p.Gly93Arg
II	5423524	T	C	missense_variant	F07F6.7	F07F6.7	c.212A>G	p.Asp71Gly
II	5441028	G	T	missense_variant	nmr-1	F07F6.6.2	c.2647C>A	p.His883Asn
II	5468394	AGAT	A	conservative_inframe_deletion	clr-1	F56D1.4b	c.925_927del GAT	p.Asp309del
II	5470083	C	A	missense_variant	clr-1	F56D1.4e	c.1591C>A	p.Pro531Thr
II	5477678	A	G	missense_variant	ilcr-2	F56D1.2	c.1714A>G	p.Thr572Ala
II	5479061	T	A	missense_variant	F56D1.1	F56D1.1	c.390T>A	p.His130Cln
II	5490242	C	CT	frameshift_variant	nep-5	F12A10.4	c.1697dupT	p.Asp567fs
II	5507477	A	C	missense_variant	npp-24	ZK177.4.3	c.950A>C	p.Lys317Thr
II	5520003	CA	C	frameshift_variant	ZK177.2	ZK177.2	c.738delA	p.Val247fs
II	5525052	C	T	missense_variant	tbx-35	ZK177.10	c.280G>A	p.Glu94Lys
II	5529171	A	G	missense_variant	ZK177.1	ZK177.1	c.898A>G	p.Ile300Val
II	5533118	A	G	missense_variant	hlh-26	C17C3.8	c.494T>C	p.Val165Ala
II	5533685	G	T	missense_variant	hlh-26	C17C3.8	c.38C>A	p.Pro13His
II	5533710	G	A	missense_variant	hlh-26	C17C3.8	c.13C>T	p.Pro5Ser
II	5539310	A	G	missense_variant	hlh-27	C17C3.10	c.797T>C	p.Ile266Thr
II	5539434	C	T	missense_variant	hlh-27	C17C3.10	c.673G>A	p.Val225Ile
II	5539464	T	A	missense_variant	hlh-27	C17C3.10	c.643A>T	p.Ile215Phe
II	5544712	C	G	missense_variant	ins-13	C17C3.18	c.95G>C	p.Gly32Ala
II	5593078	C	T	missense_variant	cdc-14	C17G10.4g	c.239G>A	p.Gly80Asp
II	5598555	G	C	missense_variant	C17G10.10	C17G10.10	c.348C>G	p.Phe116Leu
II	5598634	A	C	missense_variant	C17G10.10	C17G10.10	c.269T>G	p.Val90Gly
II	5598704	C	G,T	missense_variant	C17G10.10	C17G10.10	c.245G>A	p.Cys82Tyr
II	5601365	A	C	missense_variant	C17G10.6	C17G10.6b	c.1042T>G	p.Tyr348Asp
II	5602173	G	A	missense_variant	C17G10.6	C17G10.6b	c.506C>T	p.Thr169Ile
II	5623212	T	C	missense_variant	np1-4.2	F59E12.5a	c.265T>C	p.Phe89Leu
II	5626628	CT	C	frameshift_variant	F59E12.15	F59E12.15	c.585delA	p.Val196fs
II	5627218	C	A	missense_variant	F59E12.15	F59E12.15	c.46G>T	p.Ala16Ser
II	5627241	ATTTCACT	A	frameshift_variant&stop_lost&splice_region_variant	F59E12.6	F59E12.6b	c.1355_*2del AGTGAAA	p.Lys452fs
II	5627244	TCACTTTTG	T	frameshift_variant&stop_lost&splice_region_variant	F59E12.6	F59E12.6b	c.1351_1358del CAAAAAGTG	p.Gln451fs
II	5632741	A	C	missense_variant	np1-4.1	F59E12.4a	c.1223A>C	p.Lys408Thr
II	5645243	T	C	missense_variant	F59E12.9	F59E12.9	c.1129A>G	p.Ile377Val
II	5650771	G	T	missense_variant	zyg-1	F59E12.2.2	c.862G>T	p.Gly288Cys
II	5660182	C	T	missense_variant	C25H3.12	C25H3.12	c.409G>A	p.Asp137Asn

APPENDIX B. ADDITIONAL TABLES

Table B.4: Genes with natural variation in the genomic interval chr. II (4794415 – 8173823)

CHR	POS	REF	ALT	effect	gene_name	feature_id	nt_change	aa_change
II	5660641	C	A	missense_variant	C25H3.12	C25H3.12	c.55G>T	p.Ala19Ser
II	5660784	C	T	start_lost	C25H3.12	C25H3.12	c.3G>A	p.Met1?
II	5684571	T	C	missense_variant	C25H3.11	C25H3.11	c.2752A>G	p.Arg918Gly
II	5731114	A	AT,ATT	frameshift_variant	C18A3.11	C18A3.11	c.88_89dupTT	p.Tyr31fs
II	5756214	G	A	missense_variant	F10C1.8	F10C1.8	c.199G>A	p.Asp67Asn
II	5756235	C	G	missense_variant	F10C1.8	F10C1.8	c.220C>G	p.Arg74Gly
II	5795298	A	G	missense_variant	acdh-8	K05F1.3	c.547A>G	p.Lys183Glu
II	5795371	T	A	missense_variant	acdh-8	K05F1.3	c.620T>A	p.Ile207Asn
II	5804908	G	A	missense_variant& splice_region_variant	K05F1.10	K05F1.10	c.232C>T	p.Pro78Ser
II	5810103	A	ATTTTTT	disruptive_inframe_insertion	msp-64	ZK1248.6	c.227_228ins AAAAAA	p.Asp76delins GluLysAsn
II	5810105	C	A	missense_variant	msp-64	ZK1248.6	c.226G>T	p.Asp76Tyr
II	5814342	A	T	missense_variant	wago-5	ZK1248.7	c.1049T>A	p.Phe350Tyr
II	5821301	T	C	missense_variant	ehs-1	ZK1248.3a	c.406T>C	p.Tyr136His
II	5832627	C	CAAAA	disruptive_inframe_insertion	ZK1248.13	ZK1248.13	c.413_414ins TTT	p.Thr138dup
II	5849882	G	C	missense_variant	H41C03.3	H41C03.3.3	c.1276G>C	p.Val426Leu
II	5864731	T	A	missense_variant	tag-234	F55C12.7.3	c.179T>A	p.Ile60Asn
II	5877228	A	C	missense_variant	F55C12.2	F55C12.2	c.41A>C	p.Asp144Ala
II	5903212	A	G	missense_variant	F59G1.4	F59G1.4a	c.1118A>G	p.Asn373Ser
II	5903596	T	C	missense_variant	F59G1.4	F59G1.4b	c.116T>C	p.Val39Ala
II	5903602	G	A	missense_variant	F59G1.4	F59G1.4b	c.122G>A	p.Gly41Glu
II	5912662	T	C	missense_variant	cgt-3	F59G1.1b.3	c.59T>C	p.Leu20Pro
II	5923700	G	A	missense_variant	vrk-1	F28B12.3.2	c.1480G>A	p.Val494Ile
II	5968426	T	G	missense_variant	B0034.4	B0034.4	c.407A>C	p.Glu136Ala
II	5968707	G	T	missense_variant	B0034.4	B0034.4	c.171C>A	p.Asp57Glu
II	5974584	T	C	missense_variant	B0034.5	B0034.5	c.311A>G	p.Lys104Arg
II	5998565	C	A	missense_variant	ins-5	ZK84.3	c.82C>A	p.Leu28Met
II	6055733	T	C	missense_variant	R12C12.7	R12C12.7.2	c.202A>G	p.Lys68Glu
II	6071957	C	T	missense_variant	H12H13.3	H12H13.3	c.1124G>A	p.Arg375Lys
II	6072062	A	T	missense_variant	H12H13.3	H12H13.3	c.1019T>A	p.Phe340Tyr
II	6080272	G	A	missense_variant	fbf-1	H12H13.4	c.355C>T	p.Pro119Ser
II	6084049	C	A	missense_variant	ptc-2	F21H12.4	c.1583G>T	p.Gly528Val
II	6084945	A	AG	frameshift_variant& splice_region_variant	ptc-2	F21H12.4	c.863dupC	p.Ser289fs
II	6085384	C	A	missense_variant	ptc-2	F21H12.4	c.475G>T	p.Gly159Cys
II	6096046	T	C	missense_variant	tpp-2	F21H12.6	c.1198A>G	p.Asn400Asp
II	6096429	T	C	missense_variant	tpp-2	F21H12.6	c.862A>G	p.Asn288Asp

APPENDIX B. ADDITIONAL TABLES

Table B.4: Genes with natural variation in the genomic interval chr. II (4794415 – 8173823)

CHR	POS	REF	ALT	effect	gene_name	feature_id	nt_change	aa_change
II	6096726	G	A	missense_variant	tpp-2	F21H12.6	c.565C>T	p.Pro189Ser
II	6099145	G	A	missense_variant	rbbp-5	F21H12.1	c.1075G>A	p.Asp359Asn
II	6220129	T	C	missense_variant	C30B5.6	C30B5.6b	c.1528A>G	p.Ile510Val
II	6246829	C	T	missense_variant	T24H7.3	T24H7.3.3	c.368C>T	p.Thr123Met
II	6254663	C	A	splice_donor_variant&intron_variant	F13H8.5	F13H8.5b	c.690+1G>T	NA
II	6261461	G	A	missense_variant&splice_region_variant	nmgp-1	F13H8.4	c.193G>A	p.Asp65Asn
II	6272362	A	G	missense_variant	F13H8.2	F13H8.2	c.1720A>G	p.Asn574Asp
II	6276993	C	T	missense_variant	bpl-1	F13H8.10a	c.1679C>T	p.Thr560Ile
II	6305239	T	G	missense_variant	C29F5.1	C29F5.1	c.577T>G	p.Ser193Ala
II	6322947	A	T	missense_variant	C32D5.4	C32D5.4	c.133T>A	p.Cys45Ser
II	6381487	G	GT	frameshift_variant	WBGene00271812	F58F12.12	c.135dupT	p.Ile46fs
II	6403833	G	A	missense_variant	T25D10.1	T25D10.1	c.143G>A	p.Arg48Gln
II	6406372	C	T	missense_variant	spp-11	T25D10.3	c.688G>A	p.Val1230Ile
II	6407058	G	A	missense_variant	spp-11	T25D10.3	c.175C>T	p.Pro59Ser
II	6407064	T	C	missense_variant	spp-11	T25D10.3	c.169A>G	p.Arg57Gly
II	6423956	T	C	missense_variant	K03H9.3	K03H9.3	c.665T>C	p.Leu222Ser
II	6426971	C	T	missense_variant	col-75	K03H9.2	c.503C>T	p.Thr168Ile
II	6427156	T	C	missense_variant	col-75	K03H9.2	c.688T>C	p.Ser230Pro
II	6442818	G	A	missense_variant	cutl-16	K06A1.3	c.949G>A	p.Ala317Thr
II	6458520	G	T	missense_variant	dgt-5	K06A1.6	c.1721C>A	p.Thr574Asn
II	6481464	T	C	missense_variant	del-10	T28D9.7	c.418A>G	p.Thr140Ala
II	6493120	G	A	missense_variant	T28D9.1	T28D9.1	c.277G>A	p.Glu93Lys
II	6529580	C	G	missense_variant	abch-1	C56E6.5	c.1317G>C	p.Lys439Asn
II	6529757	T	G	missense_variant	abch-1	C56E6.5	c.1186A>C	p.Ser396Arg
II	6530687	C	A	missense_variant	abch-1	C56E6.5	c.683G>T	p.Cys228Phe
II	6537617	C	A	missense_variant	utp-20	F18C5.3	c.986C>A	p.Ala329Glu
II	6539078	A	G	missense_variant&splice_region_variant	utp-20	F18C5.3	c.2351A>G	p.Gln784Gly
II	6558596	G	A	missense_variant	wrm-1	F18C5.2	c.2716G>A	p.Val906Ile
II	6558597	T	A	missense_variant	wrm-1	F18C5.2	c.2717T>A	p.Val906Asp
II	6591468	A	G	missense_variant	C56C10.9	C56C10.9	c.674T>C	p.Leu225Ser
II	6625176	A	G	missense_variant	F54H5.14	F54H5.14	c.169T>C	p.Phe57Leu
II	6658746	C	A	stop_gained&splice_region_variant	T19D12.10	T19D12.10	c.67G>T	p.Gly23*
II	6663213	G	A	missense_variant	T19D12.2	T19D12.2c.3	c.875G>A	p.Arg292Gln
II	6668861	G	A	missense_variant	T19D12.1	T19D12.1d	c.1403G>A	p.Cys468Tyr
II	6670301	A	T	missense_variant	T19D12.1	T19D12.1d	c.2695A>T	p.Thr899Ser

APPENDIX B. ADDITIONAL TABLES

Table B.4: Genes with natural variation in the genomic interval chr. II (4794415 – 8173823)

CHR	POS	REF	ALT	effect	gene_name	feature_id	nt_change	aa_change
II	6686689	G	A	missense_variant	C18H9.3	C18H9.3c	c.1141G>A	p.Ala381Thr
II	6715640	A	G	missense_variant	dpy-2	T14B4.6	c.991A>G	p.Thr331Ala
II	6721804	CA	C	frameshift_variant	T14B4.t1	T14B4.t1	c.41delA	p.Asn14fs
II	6721805	A	C	missense_variant	T14B4.t1	T14B4.t1	c.40A>C	p.Asn14His
II	6721812	A	C	stop_lost	T14B4.t1	T14B4.t1	c.47A>C	p.Ter16Serext*?
II	6721813	G	C	stop_lost	T14B4.t1	T14B4.t1	c.48G>C	p.Ter16Tyrext*?
II	6736534	A	C	missense_variant	T14B4.2	T14B4.2	c.280A>C	p.Asn94His
II	6736536	C	A	missense_variant	T14B4.2	T14B4.2	c.282C>A	p.Asn94Lys
II	6736537	A	C	missense_variant	T14B4.2	T14B4.2	c.283A>C	p.Lys95Gln
II	6738532	A	G	missense_variant	T14B4.1	T14B4.1	c.1088A>G	p.Lys363Arg
II	6740774	G	A	missense_variant	T14B4.1	T14B4.1	c.3106G>A	p.Glu1036Lys
II	6747805	G	A	missense_variant	F41G3.5	F41G3.5	c.356G>A	p.Arg119Gln
II	6749969	A	G	missense_variant	F41G3.20	F41G3.20	c.578T>C	p.Ile193Thr
II	6755188	C	T	missense_variant	fs-1	F41G3.4	c.7C>T	p.Pro3Ser
II	6759299	TC	T	frameshift_variant	F41G3.2	F41G3.2	c.14delC	p.Pro5fs
II	6759972	C	G	missense_variant	F41G3.2	F41G3.2	c.569C>G	p.Pro190Arg
II	6759977	C	G	missense_variant	F41G3.2	F41G3.2	c.574C>G	p.Arg192Gly
II	6759980	A	G	missense_variant	F41G3.2	F41G3.2	c.577A>G	p.Asn193Asp
II	6768894	C	A	missense_variant	agr-1	F41G3.12	c.2882G>T	p.Gly961Val
II	6793438	A	T	missense_variant	T13C2.2	T13C2.2	c.767A>T	p.Gln256Leu
II	6798897	G	A	missense_variant	T13C2.2	T13C2.2	c.1279G>A	p.Asp427Asn
II	6845832	C	T	missense_variant	egl-27	C04A2.3l.2	c.458C>T	p.Pro153Leu
II	6846666	C	T	missense_variant	egl-27	C04A2.3l.2	c.1249C>T	p.His417Tyr
II	6846741	CCAG	C	conservative_inframe_deletion	egl-27	C04A2.3l.2	c.1327_1329del GCA	p.Ala443del
II	6846742	CAG	C	frameshift_variant	egl-27	C04A2.3l.2	c.1326_1327del AG	p.Ala443fs
II	6846743	AGC	A	frameshift_variant	egl-27	C04A2.3l.2	c.1327_1328del GC	p.Ala443fs
II	6846744	GCAA	G	disruptive_inframe_deletion	egl-27	C04A2.3l.2	c.1328_1330del CAA	p.Ala443_Met444delins Val
II	6846777	A	G	missense_variant	egl-27	C04A2.3l.2	c.1360A>G	p.Ser454Gly
II	6846784	C	T	missense_variant	egl-27	C04A2.3l.2	c.1367C>T	p.Ala456Val
II	6846808	TAAATGGA	T	disruptive_inframe_deletion	egl-27	C04A2.3l.2	c.1392_1397del AATGGA	p.Met465_Glu466del
II	6846813	G	A	missense_variant	egl-27	C04A2.3l.2	c.1396G>A	p.Gln466Lys
II	6846814	A	C	missense_variant	egl-27	C04A2.3l.2	c.1397A>C	p.Gln466Ala
II	6846819	G	A	missense_variant	egl-27	C04A2.3l.2	c.1402G>A	p.Ala468Thr
II	6846820	C	A	missense_variant	egl-27	C04A2.3l.2	c.1403C>A	p.Ala468Asp
II	6846971	T	G	missense_variant	egl-27	C04A2.3l.2	c.1554T>G	p.Asp518Glu

APPENDIX B. ADDITIONAL TABLES

Table B.4: Genes with natural variation in the genomic interval chr. II (4794415 – 8173823)

CHR	POS	REF	ALT	effect	gene_name	feature_id	nt_change	aa_change
II	6847277	C	T	missense_variant	egl-27	C04A2.3l.2	c.1810C>T	p.His604Tyr
II	6847286	C	T	missense_variant	egl-27	C04A2.3l.2	c.1819C>T	p.Pro607Ser
II	6854080	C	T	missense_variant	dj1-5	C04A2.7c.1	c.2593G>A	p.Ala865Thr
II	6883191	T	C	missense_variant	pmp-1	C44B7.8	c.187A>G	p.Thr63Ala
II	6892938	G	T	missense_variant	C44B7.12	C44B7.12	c.514C>A	p.His172Asn
II	6900757	T	A	missense_variant	C44B7.2	C44B7.2a.2	c.1054T>A	p.Leu352Met
II	6902624	T	A	missense_variant	psmd-9	C44B7.1	c.421T>A	p.Phe141Ile
II	6902659	G	T	missense_variant	psmd-9	C44B7.1	c.456G>T	p.Lys152Asn
II	6930775	G	A	missense_variant	flec-1	F22D3.2g	c.193G>A	p.Val65Ile
II	6950485	T	C	missense_variant	F22D3.6	F22D3.6	c.514A>G	p.Ile172Val
II	6956515	G	A	missense_variant	tra-2	C15F1.3c.2	c.317C>T	p.Thr106Ile
II	6958136	T	G	stop_lost	tra-2	C15F1.3b	c.3443A>C	p.Ter1148Serext*?
II	6962570	C	G	missense_variant	tra-2	C15F1.3b	c.2307G>C	p.Lys769Asn
II	6976812	A	C	stop_lost	C15F1.8	C15F1.8	c.499T>G	p.Ter167Glyext*?
II	6976851	T	G	missense_variant	C15F1.8	C15F1.8	c.460A>C	p.Lys154Gln
II	6976879	T	A	missense_variant	C15F1.8	C15F1.8	c.432A>T	p.Glu144Asp
II	6976880	T	C	missense_variant	C15F1.8	C15F1.8	c.431A>G	p.Glu144Gly
II	6976898	T	C	missense_variant	C15F1.8	C15F1.8	c.413A>G	p.Lys138Arg
II	6976901	T	C	missense_variant	C15F1.8	C15F1.8	c.410A>G	p.Lys137Arg
II	6976950	T	G	missense_variant	C15F1.8	C15F1.8	c.361A>C	p.Lys121Gln
II	6976967	A	G	missense_variant	C15F1.8	C15F1.8	c.344T>C	p.Ile115Thr
II	6976993	A	C	missense_variant	C15F1.8	C15F1.8	c.318T>G	p.Asn106Lys
II	6976995	T	C	missense_variant	C15F1.8	C15F1.8	c.316A>G	p.Asn106Asp
II	6977001	C	A	missense_variant	C15F1.8	C15F1.8	c.310G>T	p.Ala104Ser
II	6977003	T	G	missense_variant	C15F1.8	C15F1.8	c.308A>C	p.Lys103Thr
II	6977151	T	C	missense_variant	C15F1.8	C15F1.8	c.160A>G	p.Lys54Glu
II	6977166	A	G	missense_variant	C15F1.8	C15F1.8	c.145T>C	p.Ser49Pro
II	6977273	G	A	missense_variant	C15F1.8	C15F1.8	c.95C>T	p.Ser32Phe
II	7003307	T	C	missense_variant	C52E12.t2	C52E12.t2	c.35T>C	p.Val12Ala
II	7018800	GCAT	G	conservative_inframe_deletion	lst-6	C52E12.4	c.1408_1410del ATG	p.Met470del
II	7018804	CATA	C	disruptive_inframe_deletion& splice_region_variant	lst-6	C52E12.4	c.1404_1406del TAT	p.Met469del
II	7040658	G	T	missense_variant	cin-4	ZK1127.7	c.2137G>T	p.Ala713Ser
II	7042913	CA	C	frameshift_variant	ZK1127.6	ZK1127.6.2	c.36delA	p.Lys12fs
II	7046256	C	CCTT	conservative_inframe_insertion	tcer-1	ZK1127.9c.4	c.1920_1921ins AAG	p.Met640_Asp641insLys
II	7047332	A	G	missense_variant	tcer-1	ZK1127.9e.1	c.947T>C	p.Val316Ala
II	7049498	G	C	missense_variant	tcer-1	ZK1127.9d.2	c.430C>G	p.Gln144Glu

Table B.4: Genes with natural variation in the genomic interval chr. II (4794415 – 8173823)

CHR	POS	REF	ALT	effect	gene_name	feature_id	nt_change	aa_change
II	7052564	ACCGATCT ACGGTAAC	ACTAGTCCGATCTACGGTAAC,A	conservative_inframe_deletion	ZK1127.12	ZK1127.12.4	c.139_153del CCGATCTA CGGTAAC	p.Pro47_Asn51del
II	7054385	A	T	missense_variant	ZK1127.4	ZK1127.4	c.389A>T	p.Tyr130Phe
II	7054461	A	C	missense_variant	ZK1127.4	ZK1127.4	c.465A>C	p.Leu153Phe
II	7065236	T	A	missense_variant& splice_region_variant	him-14	ZK1127.11	c.2411A>T	p.Asp804Val
II	7066534	G	T	missense_variant	nos-2	ZK1127.12	c.669G>T	p.Glu223Asp
II	7066882	T	G	missense_variant	him-14	ZK1127.11	c.2178A>C	p.Glu726Asp
II	7067400	T	C	missense_variant	him-14	ZK1127.11	c.1801A>G	p.Thr601Ala
II	7069699	A	G	missense_variant	him-14	ZK1127.11	c.226T>C	p.Cys76Arg
II	7092495	A	G	missense_variant	T02G5.14	T02G5.14	c.425A>G	p.Tyr142Cys
II	7094780	G	A	missense_variant	T02G5.3	T02G5.3	c.299G>A	p.Arg100Lys
II	7097353	C	T	missense_variant	mct-5	T02G5.12	c.1093G>A	p.Gly365Arg
II	7121661	ATTAC	A	splice_donor_variant& splice_region_variant& intron_variant	F10E7.11	F10E7.11	c.-70+1_70+4delGTAA	NA
II	7122252	C	A	missense_variant	F10E7.2	F10E7.2	c.11C>A	p.Thr4Lys
II	7123663	G	A	missense_variant	F10E7.2	F10E7.2	c.125G>A	p.Cys42Tyr
II	7123668	CAA	C	frameshift_variant	F10E7.2	F10E7.2	c.132_133delAA	p.Gln46fs
II	7123724	TCAACAACAAC AACAACAACA	TCAACAACAACAAC, TCAACAACAACAACAACA, TCAACAACAACAACAACA ACAACA, TCAACAACAACAACAACA, TCAACAGCAACAACAACA CAACAACAACAACAACA, TCAACAACAACAACAACA T, TCAACAGCAACAACAACAAC AACAACAACAACAACAACA, TCAACAACAACAACAACA, TCAACAACAACAACAACAAC AACAA, TCAACAACAACAACAACA TCAA	conservative_inframe_deletion	F10E7.2	F10E7.2	c.229_231delCAA	p.Gln77del
II	7123913	A	G	missense_variant	F10E7.2	F10E7.2	c.250A>G	p.Lys84Glu
II	7123915	A	T	missense_variant	F10E7.2	F10E7.2	c.252A>T	p.Lys84Asn
II	7123916	C	T	missense_variant	F10E7.2	F10E7.2	c.253C>T	p.Pro85Ser
II	7124047	C	CG	frameshift_variant	F10E7.2	F10E7.2	c.386dupG	p.Thr130fs

APPENDIX B. ADDITIONAL TABLES

Table B.4: Genes with natural variation in the genomic interval chr. II (4794415 – 8173823)

CHR	POS	REF	ALT	effect	gene_name	feature_id	nt_change	aa_change
II	7124146	G	GGACGGGGAT, GGACGGGGAC, GGACGGGGAA, GGACGAGGAT, GGACGGGGAC	disruptive_inframe_insertion	F10E7.2	F10E7.2	c.491_492ins TGACGGCGGA	p.Ala163_Glu164ins AspAspAla
II	7124147	G	GACGCGGAGA	conservative_inframe_insertion	F10E7.2	F10E7.2	c.492_493ins AACGGGGAG	p.Glu164_Val165ins AsnAlaGlu
II	7124148	A	ACGGGAGGC, ACGAGGAGGC, ACGGGAGGG, ACGGGAGGT	disruptive_inframe_insertion	F10E7.2	F10E7.2	c.486_494dup CGCGGAGGT	p.Val165_Thr166ins AlaGluVal
II	7124149	C	CGCGGAGGAT, CGCGGAGGAA, CGCGGAGGG, CTCGGAGGAA, CGCGGAGGAG, CGAGGAGGAG, CGCGGAGGCA	disruptive_inframe_insertion	F10E7.2	F10E7.2	c.493_494ins GGCGGGAGG	p.Glu164_Val165ins GlyAlaGlu
II	7124150	G	CGCGGAGGACT, GCGGAGGAGC, GCGGAGGACA, GCGGAGGTCC, GCGGAGGACT	disruptive_inframe_insertion	F10E7.2	F10E7.2	c.494_495ins CCCCGGAGGT	p.Val165_Thr166ins ProGluVal
II	7124151	C	CGGAGGACTA, CGGAGGACGT, CGGATGACGA, CGGAGGACCG, CGGAGGACGA, CGGAGGGGG	disruptive_inframe_insertion	F10E7.2	F10E7.2	c.493_494ins GGCGGGAGG	p.Glu164_Val165ins GlyGlyGlu
II	7124152	G	GGAGGACGCA, GGAGGACTCC, GGAGGACGCT, GGAGGACTCT, GGAGGTGCT, GGATGACGCT	disruptive_inframe_insertion	F10E7.2	F10E7.2	c.493_495ins CGCTGAGGT	p.Val165_Thr166ins AlaGluVal
II	7124153	G	GAGGACCGGA, GAGGACCGGT, GAGGACGGC, GAGGACCGGT	disruptive_inframe_insertion	F10E7.2	F10E7.2	c.493_494ins ACGGCGCAGG	p.Glu164_Val165ins AspAlaGln

APPENDIX B. ADDITIONAL TABLES

Table B.4: Genes with natural variation in the genomic interval chr. II (4794415 – 8173823)

CHR	POS	REF	ALT	effect	gene_name	feature_id	nt_change	aa_change
II	7124154	A	AGGACGGGGG, AGGACGGGGC, AGGGCGGGG, AGGACGGGT	disruptive_inframe_insertion	F10E7.2	F10E7.2	c.493_494ins GGCGGGGGG	p.Glu164_Val165ins GlyAlaGly
II	7124155	G	GGACGGGAA, GGACGGGGAT, GGACTCGGAT, GGACCCGGAC, GGACGGGCT, GGACGGGGAC	disruptive_inframe_insertion	F10E7.2	F10E7.2	c.493_494ins ACTCGGATG	p.Glu164_Val165ins AspSerAsp
II	7124156	G	GACGGGGAGT, GACGAGGAGT, GACGGGGAGA, GACGGGGAGT, GACGTGGAGT, GACGGGGAGC	disruptive_inframe_insertion	F10E7.2	F10E7.2	c.493_494ins ACGTGGAGT	p.Val165delins AspValGluPhe
II	7124157	T	TCCGGGAGGG	disruptive_inframe_insertion	F10E7.2	F10E7.2	c.494_495ins CGCGGAGGG	p.Val165_Thr166ins AlaGluGly
II	7124365	T	A	missense_variant	F10E7.2	F10E7.2	c.702T>A	p.Phe234Leu
II	7124371	AT	A	frameshift_variant	F10E7.2	F10E7.2	c.714delT	p.Phe238fs
II	7126757	A	T	start_lost	F10E7.1	F10E7.1	c.1A>T	p.Met1?
II	7126803	A	T	missense_variant	F10E7.1	F10E7.1	c.47A>T	p.Asn161le
II	7126920	T	C	missense_variant	F10E7.1	F10E7.1	c.74T>C	p.Leu25Pro
II	7126981	C	A	missense_variant& splice_region_variant	F10E7.1	F10E7.1	c.135C>A	p.Asn45Lys
II	7127228	A	G	missense_variant	F10E7.1	F10E7.1	c.269A>G	p.His90Arg
II	7127248	C	A	missense_variant	F10E7.1	F10E7.1	c.289C>A	p.Arg97Ser
II	7127492	C	T	missense_variant	F10E7.1	F10E7.1	c.488C>T	p.Ala163Val
II	7127495	G	A	missense_variant	F10E7.1	F10E7.1	c.491G>A	p.Arg164His
II	7127570	G	A	missense_variant	F10E7.1	F10E7.1	c.566G>A	p.Gly189Glu
II	7127627	A	T	missense_variant	F10E7.1	F10E7.1	c.579A>T	p.Lys193Asn
II	7127715	G	T	missense_variant	F10E7.1	F10E7.1	c.667G>T	p.Gly223Cys
II	7127719	C	A	missense_variant	F10E7.1	F10E7.1	c.671C>A	p.Pro224Gln
II	7128040	A	G	missense_variant	F10E7.1	F10E7.1	c.767A>G	p.Asn256Ser
II	7128121	T	C	missense_variant	F10E7.1	F10E7.1	c.848T>C	p.Leu283Ser
II	7128313	ATCCAAT	A	disruptive_inframe_deletion	F10E7.1	F10E7.1	c.1041_1046del TCCAAT	p.Pro348_Ile349del
II	7128314	TCCAATC	T	conservative_inframe_deletion	F10E7.1	F10E7.1	c.1042_1047del CCAATC	p.Pro348_Ile349del

APPENDIX B. ADDITIONAL TABLES

Table B.4: Genes with natural variation in the genomic interval chr. II (4794415 – 8173823)

CHR	POS	REF	ALT	effect	gene_name	feature_id	nt_change	aa_change
II	7128315	CCAAATCT	C	disruptive_inframe_deletion	F10E7.1	F10E7.1	c.1046_1051del TCTCAA	p.Ile349_Ser350del
II	7128317	AATCTC	A	frameshift_variant	F10E7.1	F10E7.1	c.1046_1050del TCTCA	p.Ile349fs
II	7128318	ATCTC	A,ACTC	frameshift_variant	F10E7.1	F10E7.1	c.1046_1049del TCTC	p.Ile349fs
II	7128319	TCTCAAC	T	disruptive_inframe_deletion	F10E7.1	F10E7.1	c.1047_1052del CTCAAC	p.Ser350_Thr351del
II	7128320	CT	C	frameshift_variant	F10E7.1	F10E7.1	c.1048delT	p.Ser350fs
II	7128409	G	T	missense_variant	F10E7.1	F10E7.1	c.1136G>T	p.Arg379Ile
II	7128464	T	TA	frameshift_variant	F10E7.1	F10E7.1	c.1191_1192insA	p.Pro398fs
II	7128589	T	C	missense_variant	F10E7.1	F10E7.1	c.1316T>C	p.Val439Ala
II	7128636	G	T	stop_gained	F10E7.1	F10E7.1	c.1363G>T	p.Gly455*
II	7128850	A	G	missense_variant	F10E7.1	F10E7.1	c.1577A>G	p.Asp526Gly
II	7149570	T	C	missense_variant	T07F8.1	T07F8.1	c.323T>C	p.Phe108Ser
II	7150150	T	TGGCATC,TGGC	conservative_inframe_insertion	T07F8.1	T07F8.1	c.903_904ins GGCATC	p.Ile301_Ile302ins GlyIle
II	7150986	A	T	missense_variant	T07F8.1	T07F8.1	c.1739A>T	p.Tyr580Phe
II	7151752	A	G	missense_variant	T07F8.1	T07F8.1	c.2461A>G	p.Arg821Gly
II	7158047	A	G	missense_variant	C27H5.2	C27H5.2c	c.256A>G	p.Thr86Ala
II	7158598	C	A	missense_variant	C27H5.2	C27H5.2c	c.650C>A	p.Thr217Lys
II	7161407	C	A	missense_variant	WBGene00271810	C27H5.12	c.85G>T	p.Asp29Tyr
II	7164027	C	T	missense_variant	C27H5.4	C27H5.4b	c.550G>A	p.Val184Ile
II	7174427	T	C	missense_variant	glc-4	C27H5.8	c.253A>G	p.Ile85Val
II	7189871	C	T	missense_variant	E04F6.4	E04F6.4	c.436C>T	p.Pro146Ser
II	7190505	A	G	missense_variant	E04F6.4	E04F6.4	c.731A>G	p.Lys244Arg
II	7193389	A	T	missense_variant	acdh-12	E04F6.5a.2	c.1187T>A	p.Leu396His
II	7194383	G	T	missense_variant	acdh-12	E04F6.5a.2	c.288C>A	p.Asp96Glu
II	7201531	T	A	missense_variant	pcrg-1	E04F6.2	c.608T>A	p.Ile203Asn
II	7202785	C	A	missense_variant	srd-56	E04F6.14	c.386G>T	p.Arg129Ile
II	7204457	C	T	missense_variant	srd-58	E04F6.13	c.482G>A	p.Arg161Gln
II	7207080	T	A	missense_variant	dhs-7	E04F6.7	c.936A>T	p.Lys312Asn
II	7207448	A	G	missense_variant	dhs-7	E04F6.7	c.568T>C	p.Ser190Pro
II	7218919	G	C	missense_variant	clh-3	E04F6.11g	c.2854C>G	p.Arg952Gly
II	7220464	C	T	missense_variant	clh-3	E04F6.11g	c.1916G>A	p.Gly639Asp
II	7249019	A	G	missense_variant	skpo-3	F32A5.2b.2	c.1729A>G	p.Met577Val
II	7250937	C	T	missense_variant	prx-13	F32A5.6	c.497G>A	p.Ser166Asn
II	7257589	G	C	missense_variant	F32A5.8	F32A5.8.2	c.300G>C	p.Gln100His
II	7276870	G	GC	frameshift_variant	C30G12.3	C30G12.3	c.156dupC	p.Ser53fs
II	7292298	G	T	missense_variant	C30G12.6	C30G12.6b	c.3271C>A	p.Leu109Ile

APPENDIX B. ADDITIONAL TABLES

Table B.4: Genes with natural variation in the genomic interval chr. II (4794415 – 8173823)

CHR	POS	REF	ALT	effect	gene_name	feature_id	nt_change	aa_change
II	7294386	T	C	missense_variant	C30G12.6	C30G12.6b	c.1438A>G	p.Ser480Cly
II	7294644	T	G	missense_variant	C30G12.6	C30G12.6b	c.1225A>C	p.Ile409Leu
II	7296342	T	A	missense_variant	puf-8	C30G12.7	c.1570A>T	p.Met524Leu
II	7302940	A	T	missense_variant	cul-4	F45E12.3	c.605A>T	p.Gln202Leu
II	7319134	C	A	missense_variant	F21D12.2	F21D12.2	c.260C>A	p.Thr87Lys
II	7351807	A	T	missense_variant	hsp-4	F43E2.8a.2	c.571T>A	p.Ser191Thr
II	7360151	G	A	missense_variant	insc-1	F43E2.3a	c.146G>A	p.Ser49Asn
II	7367472	A	G	stop_lost	F43E2.6	F43E2.6b.1	c.289T>C	p.Ter97Glnext*?
II	7367480	A	G	missense_variant	F43E2.6	F43E2.6b.1	c.281T>C	p.Ile94Thr
II	7367574	T	C	missense_variant	F43E2.6	F43E2.6b.1	c.187A>G	p.Asn63Asp
II	7367579	T	C	missense_variant	F43E2.6	F43E2.6b.1	c.182A>G	p.Asp61Gly
II	7399259	G	C	missense_variant	C07D10.1	C07D10.1	c.183G>C	p.Lys61Asn
II	7402593	T	C	missense_variant	nas-7	C07D10.4	c.115A>G	p.Asn39Asp
II	7411361	A	C	missense_variant	K02A2.5	K02A2.5	c.156T>G	p.His52Gln
II	7447967	T	A	missense_variant	Y9D1A.1	Y9D1A.1	c.93T>A	p.Ser313Thr
II	7447968	C	T	missense_variant	Y9D1A.1	Y9D1A.1	c.938C>T	p.Ser313Phe
II	7448088	G	A	splice_donor_variant& intron_variant	Y9D1A.1	Y9D1A.1	c.1057+1G>A	NA
II	7448383	T	G	missense_variant	Y9D1A.1	Y9D1A.1	c.1259T>G	p.Ile420Arg
II	7457079	G	T	missense_variant	D1022.4	D1022.4.2	c.163C>A	p.Gln55Lys
II	7473389	A	T	missense_variant	aka-1	D1022.7a.1	c.3647T>A	p.Ile1216Asn
II	7478202	C	T	missense_variant	aka-1	D1022.7c.1	c.280G>A	p.Asp94Asn
II	7489191	A	G	missense_variant	D1022.9	D1022.9a	c.281T>C	p.Val94Ala
II	7489309	G	A	missense_variant	D1022.9	D1022.9a	c.163C>T	p.Pro55Ser
II	7498377	A	G	missense_variant	srab-14	R10H1.2b	c.506T>C	p.Val169Ala
II	7520277	A	G	missense_variant	C28F5.4	C28F5.4	c.2465T>C	p.Ile822Thr
II	7521465	A	AT	frameshift_variant	C28F5.4	C28F5.4	c.1464dupA	p.Ser489fs
II	7526429	G	A	splice_acceptor_variant& intron_variant	rnh-1.1	ZK1290.6	c.371-1G>A	NA
II	7526557	TTCATCTTTGGGA	T	conservative_inframe_deletion	rnh-1.1	ZK1290.6	c.502_513del TCTTTGGGATCA	p.Ser168_Ser171del
II	7526561	TCTTTGGGATCAA	T	stop_gained& disruptive_inframe_deletion	rnh-1.1	ZK1290.6	c.503_514del CTTTGGGATCAA	p.Ser168_Lys172delins Ter
II	7539812	A	G	missense_variant	nfi-1	ZK1290.4b	c.802A>G	p.Ile268Val
II	7546978	T	A	missense_variant	fbxa-224	ZK1290.9	c.997A>T	p.Ile333Phe
II	7561981	G	A	missense_variant	ZK1290.1	ZK1290.1	c.172G>A	p.Val58Ile
II	7578857	TAATTTCAAG	T	splice_acceptor_variant& splice_region_variant& intron_variant	F35D2.6	F35D2.6	c.581-9_581-1del AATTTCAAG	NA
II	7579385	T	A	missense_variant	F35D2.6	F35D2.6	c.999T>A	p.Phe333Leu

APPENDIX B. ADDITIONAL TABLES

Table B.4: Genes with natural variation in the genomic interval chr. II (4794415 – 8173823)

CHR	POS	REF	ALT	effect	gene_name	feature_id	nt_change	aa_change
II	7602437	A	T	missense_variant	npp-21	R07G3.3a.2	c.2450A>T	p.Gln817Leu
II	7603513	G	T	missense_variant	npp-21	R07G3.3a.2	c.3387G>T	p.Lys1129Asn
II	7606346	T	C	missense_variant	npp-21	R07G3.3a.2	c.4694T>C	p.Val1565Ala
II	7614055	A	T	missense_variant	oig-4	R07G3.9	c.339A>T	p.Glu113Asp
II	7661676	C	A	missense_variant	alfa-1	F18A1.6b.3	c.340G>T	p.Ala14Ser
II	7661983	A	T	missense_variant	alfa-1	F18A1.6b.3	c.77T>A	p.Leu26Gln
II	7688276	T	C	splice_donor_variant&intron_variant	F18A1.1	F18A1.1	c.1976+2T>C	NA
II	7697381	T	A	missense_variant	B0495.5	B0495.5.2	c.912A>T	p.Glu304Asp
II	7697470	A	C	missense_variant	B0495.5	B0495.5.2	c.823T>G	p.Ser275Ala
II	7697472	C	T	missense_variant	B0495.5	B0495.5.2	c.821G>A	p.Gly274Glu
II	7702665	T	C	missense_variant	B0495.7	B0495.7.2	c.2377A>G	p.Lys793Glu
II	7729084	G	A	missense_variant	cpna-2	B0228.4a	c.2822G>A	p.Arg941Gln
II	7731709	T	C	missense_variant	cpna-2	B0228.4c	c.4322T>C	p.Leu1441Ser
II	7736909	A	C	missense_variant	cpna-2	B0228.4c	c.9522A>C	p.Glu3174Asp
II	7738506	A	T	missense_variant	cpna-2	B0228.4c	c.11119A>T	p.Thr3707Ser
II	7738990	A	G	missense_variant	cpna-2	B0228.4c	c.11603A>G	p.Asn3868Ser
II	7739781	G	A	missense_variant	cpna-2	B0228.4c	c.12394G>A	p.Ala4132Thr
II	7740331	T	A	missense_variant	cpna-2	B0228.4c	c.12944T>A	p.Met4315Lys
II	7740739	C	T	missense_variant	cpna-2	B0228.4c	c.13352C>T	p.Ser4451Phe
II	7741114	C	T	missense_variant	cpna-2	B0228.4c	c.13727C>T	p.Thr4576Ile
II	7743099	G	A	missense_variant	cpna-2	B0228.4c	c.703G>A	p.Asp235Asn
II	7743118	T	C	missense_variant	cpna-2	B0228.4c	c.722T>C	p.Leu241Pro
II	7746752	A	G	missense_variant	cpna-2	B0228.4c	c.3452A>G	p.Asn1151Ser
II	7746797	C	T	missense_variant	cpna-2	B0228.4c	c.3497C>T	p.Thr1166Ile
II	7747810	C	A	missense_variant	cpna-2	B0228.4c	c.4510C>A	p.Gln1504Lys
II	7748855	C	G	missense_variant	cpna-2	B0228.4c	c.5555C>G	p.Thr1852Arg
II	7750331	C	T	missense_variant	cpna-2	B0228.4c	c.7031C>T	p.Thr2344Met
II	7762498	A	G	missense_variant	B0228.9	B0228.9	c.940T>C	p.Phe314Leu
II	7764107	T	G	missense_variant	srh-39	C06A8.7	c.829A>C	p.Ile277Leu
II	7764831	G	A	missense_variant	srh-39	C06A8.7	c.388C>T	p.Arg130Cys
II	7769986	T	C	missense_variant	C06A8.8	C06A8.8a	c.451A>G	p.Ile151Val
II	7771681	C	T	missense_variant	C06A8.8	C06A8.8a	c.245G>A	p.Arg82Gln
II	7773606	A	T	missense_variant	C06A8.6	C06A8.6	c.841A>T	p.Ser281Cys
II	7784273	G	A	missense_variant	skr-17	C06A8.4	c.347G>A	p.Ser116Asn
II	7816705	CAA	C	frameshift_variant&start_lost	cki-2	T05A6.2c.2	c.-1_1delAA	p.Met1fs
II	7816707	AAT	A	frameshift_variant&start_lost	cki-2	T05A6.2c.2	c.1_2delAT	p.Met1fs

APPENDIX B. ADDITIONAL TABLES

Table B.4: Genes with natural variation in the genomic interval chr. II (4794415 – 8173823)

CHR	POS	REF	ALT	effect	gene_name	feature_id	nt_change	aa_change
II	7820642	T	C	missense_variant	irlid-47	T05A6.4	c.1303A>G	p.Ile435Val
II	7822230	AAACTCG	A	disruptive_inframe_deletion	irlid-47	T05A6.4	c.258_263delCGAGTT	p.Glu87_Phe88del
II	7825656	G	C	missense_variant	srw-62	T05A6.6	c.250G>C	p.Asp84His
II	7826270	A	G	missense_variant	srw-62	T05A6.6	c.757A>G	p.Thr253Ala
II	7844641	T	G	missense_variant	klp-3	T09A5.2a	c.1243T>G	p.Ser415Ala
II	7845370	C	G	missense_variant	klp-3	T09A5.2a	c.1784C>G	p.Ser595Cys
II	7845704	A	G	missense_variant	acr-7	T09A5.3	c.1577T>C	p.Phe526Ser
II	7847590	A	G	missense_variant	T09A5.4	T09A5.4	c.818A>G	p.Gln273Arg
II	7848235	A	C	missense_variant	acr-7	T09A5.3	c.1068T>G	p.Cys356Trp
II	7855662	A	T	missense_variant	T09A5.7	T09A5.7	c.45T>A	p.His15Gln
II	7856832	T	G	missense_variant	cec-3	T09A5.8	c.409A>C	p.Thr137Pro
II	7862255	T	C	missense_variant	lin-5	T09A5.10.2	c.1502T>C	p.Val501Ala
II	7864550	T	A	missense_variant	ostb-1	T09A5.11.2	c.43T>A	p.Phe15Ile
II	7865582	G	A	missense_variant	ostb-1	T09A5.11.2	c.962G>A	p.Gly321Glu
II	7872221	A	G	missense_variant	T01H3.3	T01H3.3	c.1073T>C	p.Ile358Thr
II	7876684	T	C	splice_acceptor_variant& intron_variant	T01H3.2	T01H3.2.2	c.1072-2A>G	NA
II	7896678	C	A	missense_variant	ptc-1	ZK675.1.2	c.1388C>A	p.Thr463Asn
II	7914328	T	G	missense_variant	spv-1	ZK669.1c	c.2519A>C	p.Gln840Pro
II	7935663	G	T	missense_variant	ZK669.2	ZK669.2	c.21C>A	p.Ser7Arg
II	7944817	A	T	missense_variant	dbt-1	ZK669.4	c.232T>A	p.Cys78Ser
II	7995716	T	A	missense_variant	pgp-11	DH11.3	c.3802A>T	p.Ser1268Cys
II	7996870	G	A	missense_variant	pgp-11	DH11.3	c.2905C>T	p.Pro969Ser
II	7997129	A	G	missense_variant	pgp-11	DH11.3	c.2696T>C	p.Val899Ala
II	7999355	C	T	missense_variant	pgp-11	DH11.3	c.868G>A	p.Gln290Lys
II	8008702	C	A	missense_variant	glna-2	DH11.1a.2	c.20C>A	p.Thr7Asn
II	8025320	A	T	missense_variant	swsn-7	C08B11.3	c.2252T>A	p.Ile751Asn
II	8048511	A	G	missense_variant	apn-1	T05H10.2.2	c.736A>G	p.Ile246Val
II	8080160	A	C	missense_variant	K02C4.5	K02C4.5	c.762A>C	p.Lys254Asn
II	8136516	C	T	missense_variant	ctns-1	C41C4.7b	c.301G>A	p.Asp101Asn
II	8139877	A	C	missense_variant	cdc-48.2	C41C4.8.2	c.1273A>C	p.Ile425Leu
II	8142704	C	A	missense_variant	F10B5.8	F10B5.8	c.501G>T	p.Met167Ile
II	8160992	C	T	missense_variant	emb-27	F10B5.6	c.1042G>A	p.Val348Ile
II	8169402	A	C	missense_variant	T05C12.1l	T05C12.1la	c.243T>G	p.Asn81Lys
II	8173823	T	A	missense_variant	T05C12.1	T05C12.1	c.74T>A	p.Leu25His

B.5 List of genes with natural variation on chromosome III

APPENDIX B. ADDITIONAL TABLES

Table B.5: List of genes with natural variation in the genomic interval chr. III (7631003 – 8917452)

CHR	POS	REF	ALT	effect	gene_name	feature_id	nt_change	aa_change
III	7631003	G	A	missense_variant&splice_region_variant	fbn-1	ZK783.li	c.3467G>A	p.Ser1156Asn
III	7634454	A	T	missense_variant	fbn-1	ZK783.lj	c.6436A>T	p.Ile2146Phe
III	7634539	T	C	missense_variant	fbn-1	ZK783.lj	c.6521T>C	p.Val2174Ala
III	7640351	A	G	missense_variant	fbn-1	ZK783.li	c.8099A>G	p.Asp2700Gly
III	7642190	G	A	missense_variant	ZK783.t1	ZK783.t1	c.40C>T	p.His147Tyr
III	7648798	C	A	missense_variant	ZK783.6	ZK783.6	c.11C>A	p.Ala4Glu
III	7648907	C	T	missense_variant	ZK783.6	ZK783.6	c.77C>T	p.Thr26Met
III	7649215	C	CAGCAGCAGGCCCCACAGATGT, CAGCAGCAGGCCCCCGATGT	frameshift_variant	ZK783.6	ZK783.6	c.401_402ins TGTAGCAGCAGGCCCCAGA	p.Glu134His
III	7663082	A	G	missense_variant	C18H2.1	C18H2.1	c.151A>G	p.Ile51Val
III	7663508	T	C	missense_variant	C18H2.1	C18H2.1	c.577T>C	p.Phe193Leu
III	7663991	A	G	missense_variant	C18H2.1	C18H2.1	c.1060A>G	p.Lys354Glu
III	7667860	T	C	missense_variant&splice_region_variant	C18H2.1	C18H2.1	c.3664T>C	p.Ser1222Pro
III	7669689	G	A	stop_gained	C18H2.1	C18H2.1	c.4182G>A	p.Trp1394*
III	7669697	A	C	missense_variant	C18H2.1	C18H2.1	c.4190A>C	p.Tyr1397Ser
III	7669844	G	T	missense_variant	C18H2.1	C18H2.1	c.4337G>T	p.Arg1446Met
III	7671312	A	C	missense_variant	C18H2.1	C18H2.1	c.4826A>C	p.Asn1609Thr
III	7671425	T	A	missense_variant	C18H2.1	C18H2.1	c.4864T>A	p.Ser1622Thr
III	7671443	T	G	missense_variant	C18H2.1	C18H2.1	c.4882T>G	p.Tyr1628Asp
III	7671503	A	T	missense_variant	C18H2.1	C18H2.1	c.4942A>T	p.Asn1648Tyr
III	7691282	T	C	missense_variant	C18H2.5	C18H2.5	c.2324A>G	p.Lys775Arg
III	7691334	T	A	missense_variant	C18H2.5	C18H2.5	c.2272A>T	p.Thr758Ser
III	7691398	C	A	missense_variant	C18H2.5	C18H2.5	c.2208G>T	p.Lys736Asn
III	7712835	G	A	missense_variant	C03B8.3	C03B8.3	c.298G>A	p.Asp100Asn
III	7735036	G	A	missense_variant	ncl-1	ZK112.2g	c.1344G>A	p.Met448Ile
III	7740073	A	G	missense_variant	ZK112.6	ZK112.6	c.278T>C	p.Val93Ala
III	7742102	T	G	missense_variant	cdh-3	ZK112.7	c.9441A>C	p.Gln3147His
III	7743740	T	G	missense_variant	cdh-3	ZK112.7	c.8093A>C	p.Asn2698Thr
III	7743882	G	T	missense_variant	cdh-3	ZK112.7	c.7951C>A	p.Pro2651Thr
III	7744166	G	A	missense_variant	cdh-3	ZK112.7	c.7739C>T	p.Thr2580Ile
III	7744971	G	A	missense_variant	cdh-3	ZK112.7	c.6980C>T	p.Ala2327Val
III	7746798	C	T	missense_variant	cdh-3	ZK112.7	c.5380G>A	p.Glu1794Lys
III	7759822	A	C	missense_variant	pcp-1	ZK112.1.2	c.622A>C	p.Asn208His
III	7760236	A	C	missense_variant	pcp-1	ZK112.1.2	c.989A>C	p.Lys330Thr
III	7762255	C	T	missense_variant	mtx-2	ZC97.la.2	c.608G>A	p.Ser203Asn
III	6980820	C	A	missense_variant	ZK686.3	ZK686.3	c.816G>T	p.Gln272His
III	7771655	C	A	missense_variant	ZK686.6	ZK686.6	c.857G>T	p.Arg286Leu

Table B.5: Genes with natural variation in the genomic interval chr. III (7631003 – 8917452)

CHR	POS	REF	ALT	effect	gene_name	feature_id	nt_change	aa_change
III	7771668	T	C	missense_variant	ZK686.6	ZK686.6	c.844A>G	p.Thr282Ala
III	7771766	G	A	missense_variant	ZK686.6	ZK686.6	c.746C>T	p.Thr249Ile
III	7771854	G	T	missense_variant	ZK686.6	ZK686.6	c.658C>A	p.Leu220Ile
III	7771893	C	T	missense_variant	ZK686.6	ZK686.6	c.619G>A	p.Glu207Lys
III	7771904	A	T	missense_variant	ZK686.6	ZK686.6	c.608T>A	p.Val203Asp
III	7772000	T	G	missense_variant	ZK686.6	ZK686.6	c.512A>C	p.Asp171Ala
III	7772022	C	A	missense_variant	ZK686.6	ZK686.6	c.490G>T	p.Ala164Ser
III	7772230	A	C	missense_variant	ZK686.6	ZK686.6	c.330T>G	p.Asn110Lys
III	7772231	T	G	missense_variant	ZK686.6	ZK686.6	c.329A>C	p.Asn110Thr
III	7772319	T	C	missense_variant	ZK686.6	ZK686.6	c.241A>G	p.Ile81Val
III	7772355	CCTGCG	C	frameshift_variant	ZK686.6	ZK686.6	c.200_204delCGCAG	p.Ser67fs
III	7772403	C	A	missense_variant	ZK686.6	ZK686.6	c.157G>T	p.Val53Leu
III	7772491	C	G	missense_variant	ZK686.6	ZK686.6	c.69G>C	p.Gln23His
III	7772501	G	A	missense_variant	ZK686.6	ZK686.6	c.59C>T	p.Ser20Leu
III	7772529	C	A	stop_gained	ZK686.6	ZK686.6	c.31G>T	p.Glu11*
III	7772531	C	T	missense_variant	ZK686.6	ZK686.6	c.29C>A	p.Arg10Lys
III	7775158	A	T	missense_variant	bath-15	C08C3.2.2	c.713T>A	p.Phe238Tyr
III	7775479	A	C	missense_variant& splice_region_variant	bath-15	C08C3.2.2	c.488T>G	p.Leu163Arg
III	7878493	T	G	missense_variant& splice_region_variant	sup-18	C02C2.5	c.244A>C	p.Lys82Gln
III	7886114	A	G	missense_variant	ZK688.5	ZK688.5d	c.4429T>C	p.Tyr1477His
III	7895824	A	C	stop_lost	ZK688.10	ZK688.10	c.817T>G	p.Ter273Gluext*?
III	7919657	C	T	missense_variant	tag-250	C29E4.5a	c.1148G>A	p.Ser383Asn
III	7928330	T	A	missense_variant	C29E4.13	C29E4.13b	c.132A>T	p.Lys44Asn
III	7931505	G	A	missense_variant	npp-15	C29E4.4	c.2170G>A	p.Gly724Arg
III	7938412	A	G	missense_variant	C29E4.10	C29E4.10	c.407A>G	p.Asn136Ser
III	7942454	A	G	missense_variant	kle-2	C29E4.2	c.658A>G	p.Lys220Glu
III	7945744	C	T	missense_variant	gsto-1	C29E4.7	c.604G>A	p.Glu202Lys
III	7949451	C	T	missense_variant	col-90	C29E4.1	c.562C>T	p.Pro188Ser
III	7965156	C	T,A	stop_gained	F54H12.2	F54H12.2	c.576C>A	p.Tyr192*
III	7965177	C	G,T	missense_variant	F54H12.2	F54H12.2	c.597C>G	p.Phe199Leu
III	7965378	A	T	missense_variant	F54H12.2	F54H12.2	c.798A>T	p.Arg266Ser
III	7965435	AC	A	frameshift_variant	F54H12.2	F54H12.2	c.857delC	p.Pro286fs
III	7965466	G	A	missense_variant	F54H12.2	F54H12.2	c.886G>A	p.Asp296Asn
III	7968339	G	A	missense_variant	F54H12.8	F54H12.8	c.665C>T	p.Thr222Ile
III	7973042	G	T	missense_variant	eef-1B.1	F54H12.6	c.111C>A	p.Phe37Leu
III	7978537	T	C	missense_variant	C06G4.t1	C06G4.t1	c.7A>G	p.Ile3Val
III	7983254	G	T	missense_variant	clp-1	C06G4.2b.3	c.703G>T	p.Asp235Tyr

APPENDIX B. ADDITIONAL TABLES

Table B.5: Genes with natural variation in the genomic interval chr. III (7631003 – 8917452)

CHR	POS	REF	ALT	effect	gene_name	feature_id	nt_change	aa_change
III	7984833	A	G	missense_variant	clp-1	C06G4.2b.3	c.1736A>G	p.Lys579Arg
III	7992819	A	G	missense_variant	npr-17	C06G4.5	c.632T>C	p.Met211Thr
III	8017923	AAGTTTGG	A	frameshift_variant	lin-36	F44B9.6	c.2344_2350delC	p.Pro782fs
III	8018436	G	T	missense_variant	lin-36	F44B9.6	c.1904C>A	p.Thr635Lys
III	8028569	G	T	missense_variant	F44B9.8	F44B9.8	c.853C>A	p.Gln285Lys
III	8266098	A	G	missense_variant	mmps-9	F09G8.3	c.37A>G	p.Met13Val
III	8267177	G	A	missense_variant	mmps-9	F09G8.3	c.767G>A	p.Arg256Gln
III	8304308	G	A	missense_variant	mig-10	F10E9.6b.2	c.1193G>A	p.Cys398Tyr
III	8311083	A	G	missense_variant	F10E9.3	F10E9.3	c.76A>G	p.Lys266Glu
III	8365279	T	C	missense_variant	R05D3.3	R05D3.3.3	c.358T>C	p.Ser120Pro
III	8389578	T	A	missense_variant	ZK353.4	ZK353.4	c.62A>T	p.Gln21Val
III	8410036	G	A	missense_variant&splice_region_variant	ZK353.9	ZK353.9	c.109G>A	p.Val37Ile
III	8422218	G	A	missense_variant	sor-1	ZK1236.3b	c.2280G>A	p.Met760Ile
III	8432896	T	G	missense_variant	ZK1236.9a	ZK1236.9a	c.348A>C	p.Glu116Asp
III	8445215	CCAT	C	disruptive_inframe_deletion	hsp-110	C30C11.4.2	c.1421_1423delATG	p.Asp474del
III	8445217	ATCG	A	disruptive_inframe_deletion	hsp-110	C30C11.4.2	c.1419_1421delCGA	p.Asp474del
III	8477639	A	C	missense_variant	mig-39	F42H10.5b	c.2441A>C	p.Asp814Ala
III	8490822	C	T	missense_variant	F42H10.2	F42H10.2	c.178C>T	p.His607Tyr
III	8525746	T	C	missense_variant	ZC21.3	ZC21.3c	c.299T>C	p.Ile100Thr
III	8528069	G	A	missense_variant	ZC21.3	ZC21.3a	c.475G>A	p.Val159Ile
III	8536762	G	T	missense_variant	trp-1	ZC21.2b	c.158G>T	p.Ser53Ile
III	8542215	A	T	missense_variant	trp-1	ZC21.2a	c.2780A>T	p.Tyr927Phe
III	8552616	A	G	missense_variant	C02D5.4	C02D5.4	c.31A>G	p.Ile11Val
III	8555640	A	G	missense_variant	gst-2	C02D5.3.2	c.724A>G	p.Ser242Gly
III	8561483	A	T	missense_variant	acdh-6	C02D5.1	c.247A>T	p.Thr83Ser
III	8561600	C	G	missense_variant	acdh-6	C02D5.1	c.364C>G	p.Gln122Glu
III	8562346	G	A	missense_variant	acdh-6	C02D5.1	c.982G>A	p.Ala328Thr
III	8590978	A	T	missense_variant	fpr-16	C06E1.6	c.5T>A	p.Ile2Asn
III	8591928	T	G	missense_variant	C06E1.7	C06E1.7	c.745A>C	p.Ile249Leu
III	8592779	C	T	missense_variant	C06E1.7	C06E1.7	c.128G>A	p.Arg43Lys
III	8596023	T	A	missense_variant	ztf-30	C06E1.8	c.853A>T	p.Arg285Trp
III	8608646	C	T	missense_variant	C06E1.9	C06E1.9	c.590G>A	p.Ser197Asn
III	8608725	G	T	missense_variant	C06E1.9	C06E1.9	c.511C>A	p.Leu171Ile
III	8612723	A	G	missense_variant	rha-2	C06E1.10	c.1499T>C	p.Val500Ala
III	8613943	CCTTCCTT	CCTT,C,CCTTCCTTCTTCTT, CCTTCCTT	disruptive_inframe_deletion	rha-2	C06E1.10	c.455_460delAAGAAG	p.Glu152_Glu153del
III	8614818	GACATGTTTC	G	conservative_inframe_deletion	C06E1.11	C06E1.11	c.1051_1059del GAAACATGT	p.Glu351_Cys353del

APPENDIX B. ADDITIONAL TABLES

Table B.5: Genes with natural variation in the genomic interval chr. III (7631003 – 8917452)

CHR	POS	REF	ALT	effect	gene_name	feature_id	nt_change	aa_change
III	8614825	TTCACATGTA	T	disruptive_inframe_deletion	C06E1.11	C06E1.11	c.1044_1052del TACATGTGA	p.Thr349_Glu351del
III	8614826	TCACATGTAG	T	disruptive_inframe_deletion	C06E1.11	C06E1.11	c.1043_1051del CTACATGTG	p.Thr348_Glu351delins Lys
III	8614827	C	G	missense_variant	C06E1.11	C06E1.11	c.1051G>C	p.Glu351Gln
III	8616104	G	GTGGT	frameshift_variant	C06E1.11	C06E1.11	c.106_107insACCA	p.Thr36fs
III	8617138	T	A	missense_variant	C13G5.2	C13G5.2.2	c.859A>T	p.Ser287Cys
III	8617550	T	TAATGTGATGGA	frameshift_variant	C13G5.2	C13G5.2.2	c.485_495dup TCCATCACATT	p.Asn166fs
III	8617557	A	ATGAAAATGTGC, ATGAAAATGTGT, ATGAAAATGTGG	frameshift_variant	C13G5.2	C13G5.2.2	c.488_489ins ACACATTTCCA	p.His163fs
III	8617561	A	AAATGTGATGGT, AAATGTGATGGG, AAATGTGTTGGT, AAATGTGATGGC	frameshift_variant	C13G5.2	C13G5.2.2	c.484_485ins ACCAACACATT	p.Phe162fs
III	8626441	C	A	missense_variant	polk-1	F22B7.6b	c.1356G>T	p.Met452Ile
III	8629332	C	T	missense_variant	gpr-1	F22B7.13	c.325C>T	p.Pro109Ser
III	8630315	G	A	missense_variant	gpr-1	F22B7.13	c.1258G>A	p.Glu420Lys
III	8632695	C	T	missense_variant	dhj-10	F22B7.5b	c.1157C>T	p.Thr386Met
III	8649122	T	A	missense_variant	F22B7.3	F22B7.3	c.60T>A	p.His20Gln
III	8685893	G	A	missense_variant	annt-1	B0303.2.2	c.508G>A	p.Val170Ile
III	8686016	A	T	missense_variant	annt-1	B0303.2.2	c.631A>T	p.Ile211Phe
III	8693701	G	C	missense_variant	B0303.7	B0303.7a	c.47G>C	p.Cys16Ser
III	8698863	C	T	missense_variant	B0303.7	B0303.7b	c.1073C>T	p.Ala358Val
III	8754489	C	T	missense_variant	ZK370.8	ZK370.8	c.211G>A	p.Val71Ile
III	8754647	C	A	missense_variant	ZK370.8	ZK370.8	c.99C>T	p.Lys33Asn
III	8756959	C	G	missense_variant	sma-2	ZK370.2.2	c.1214C>G	p.Ser405Cys
III	8774991	T	A	missense_variant	K02D10.1	K02D10.1b.2	c.277T>A	p.Leu93Ile
III	8792317	C	T	missense_variant	ztf-1	F54F2.5a	c.22G>A	p.Gly8Ser
III	8793347	C	T	missense_variant	zfp-1	F54F2.2b	c.361C>T	p.Arg121Cys
III	8804664	C	T	missense_variant	zfp-1	F54F2.2c	c.1393C>T	p.Pro465Ser
III	8847398	C	T	missense_variant	F44E2.7	F44E2.7d	c.262G>A	p.Ala88Thr
III	8861460	T	G	splice_acceptor_variant& intron_variant	plg-1	F44E2.11	n.238-1T>G	NA
III	8891742	T	C	splice_donor_variant& intron_variant	lnkn-1	ZK637.3	c.690+2T>C	NA
III	8903582	C	T	missense_variant	lin-9	ZK637.7a.1	c.340G>A	p.Ala114Thr
III	8907770	T	G	missense_variant	unc-32	ZK637.8b	c.744T>G	p.His248Gln
III	8914282	T	A	missense_variant	trxr-2	ZK637.10	c.408T>A	p.Asn136Lys

Table B.5: Genes with natural variation in the genomic interval chr. III (7631003 – 8917452)

CHR	POS	REF	ALT	effect	gene_name	feature_id	nt_change	aa_change
III	8915145	G	A	missense_variant	trxt-2	ZK637.10	c.922G>A	p.Ala308Thr
III	8917452	G	C	missense_variant	cdc-25.3	ZK637.11	c.397C>G	p.Gln133Glu

Appendix C

Graphs

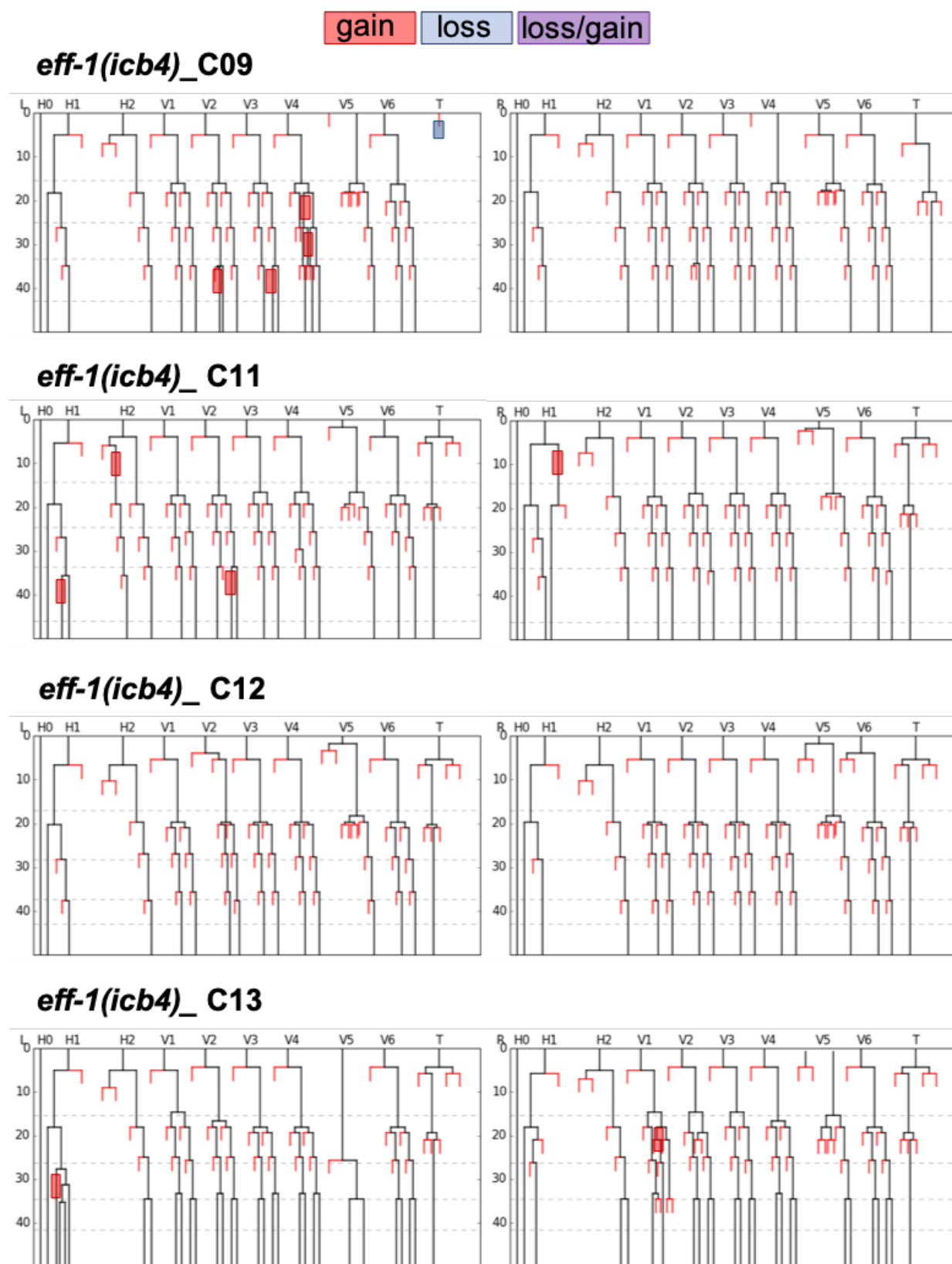


Figure C.1: Additional seam cell lineages of *eff-1(icb4)* animals. $n = 15$ lineages. The coloured boxes highlight a few developmental errors.

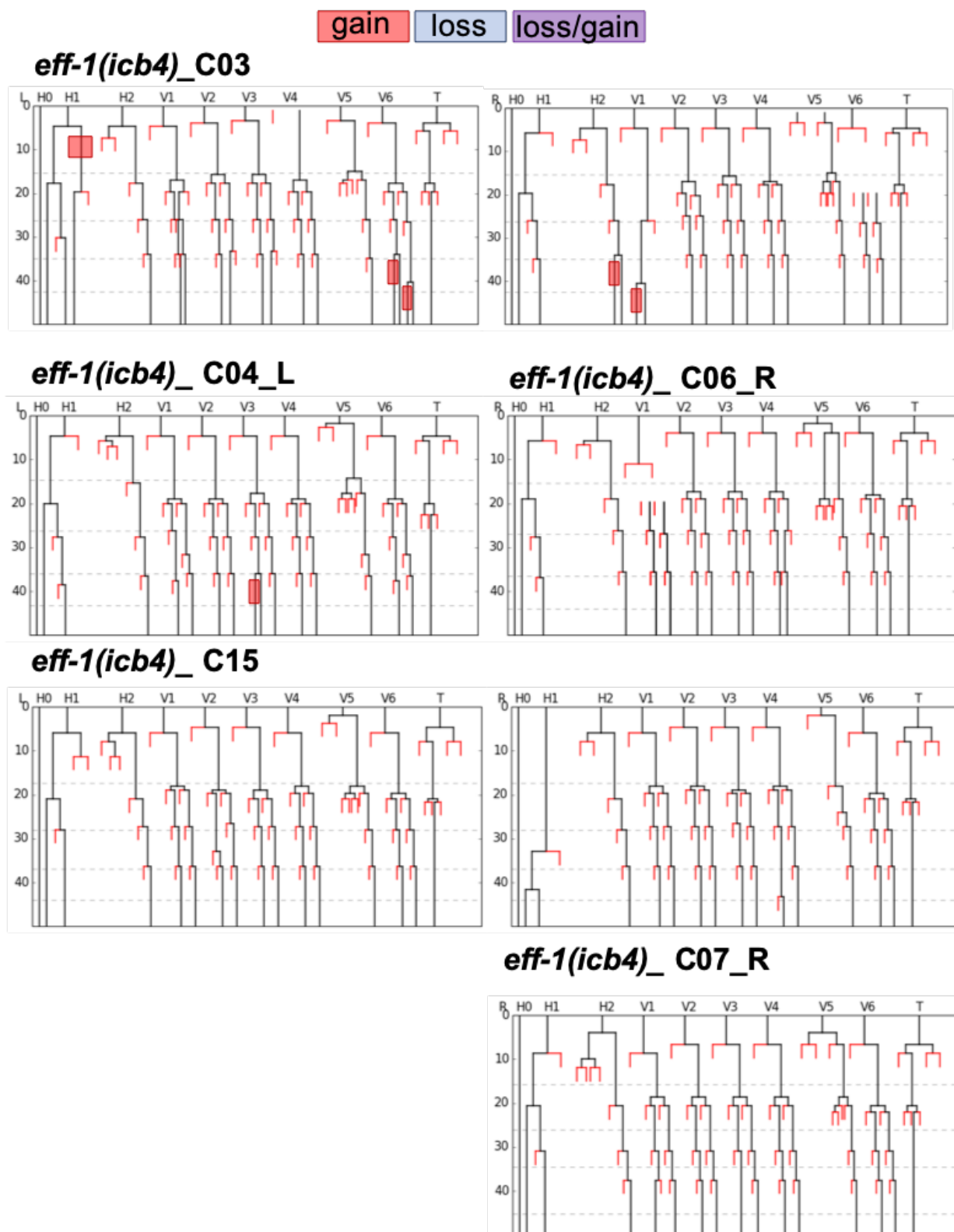


Figure C.2: Additional Seam cell lineages of *eff-1(icb4)* animals. $n = 15$ lineages. The coloured boxes highlight a few developmental errors.

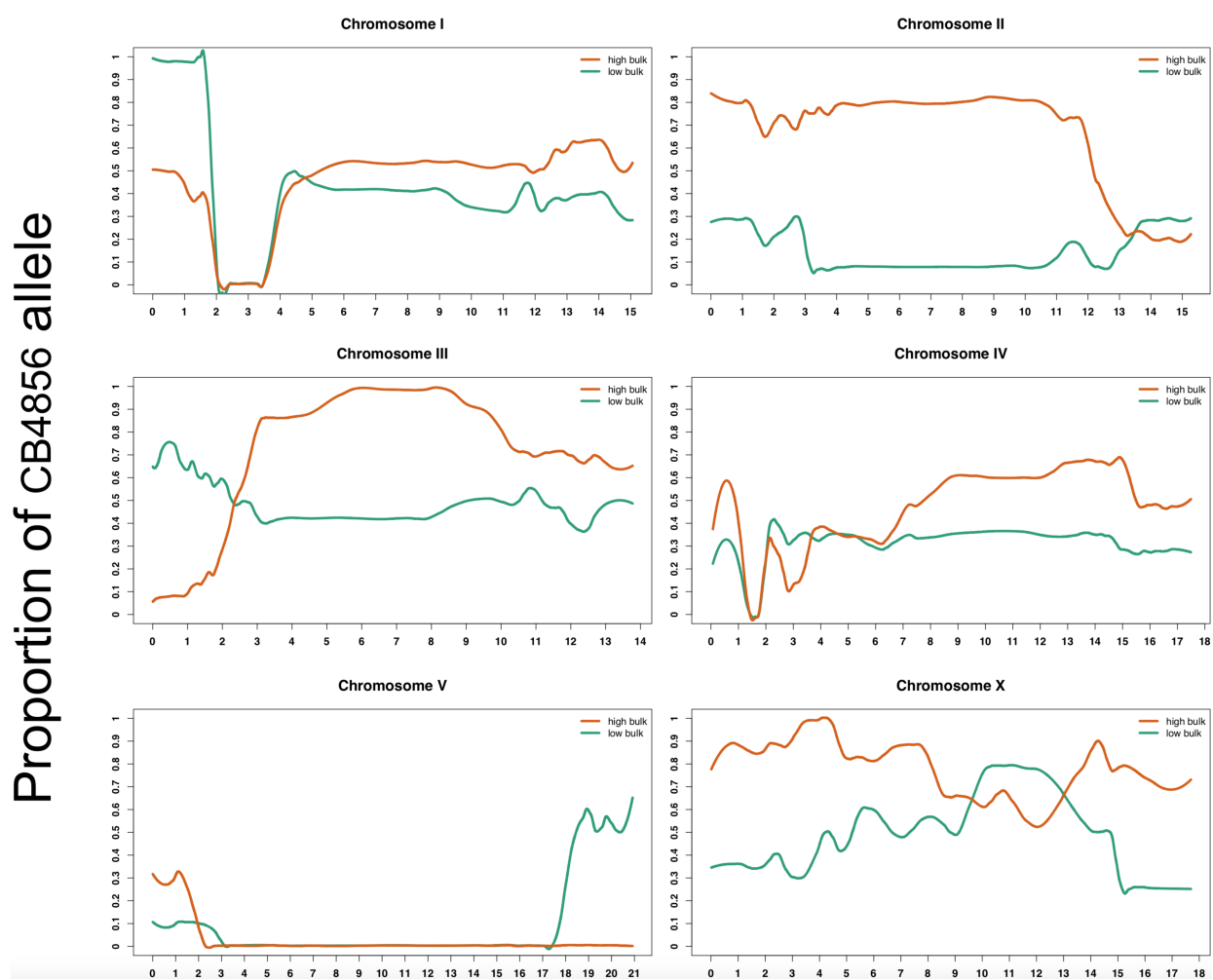


Figure C.3: **Proportion of CB4856 SNPs along the six chromosomes in stringent condition.** SNP frequencies in low-bulk and high-bulk are depicted by green and orange fitted curves, respectively. The curves represent locally weighted scatterplot smoothing (LOESS) regression lines from the SNP frequencies along the chromosomes with a span parameter of 0.1. x-axis and y-axis correspond to chromosomal position in Mb and proportion of CB4856 SNPs in the sequencing reads. 10 RILs were pooled in each group.

Appendix D

Permissions

**ELSEVIER LICENSE
TERMS AND CONDITIONS**

Apr 20, 2020

This Agreement between Ms. Sneha Koneru ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4812920849166
License date	Apr 20, 2020
Licensed Content Publisher	Elsevier
Licensed Content Publication	Current Opinion in Genetics & Development
Licensed Content Title	Genetic backgrounds and hidden trait complexity in natural populations
Licensed Content Author	Téo Fournier,Joseph Schacherer
Licensed Content Date	Dec 1, 2017
Licensed Content Volume	47
Licensed Content Issue	n/a
Licensed Content Pages	6
Start Page	48

End Page 53

Type of Use reuse in a thesis/dissertation

Portion figures/tables/illustrations

Number of figures/tables/illustrations 1

Format electronic

Are you the author of this Elsevier article? No

Will you be translating? No

Title Investigating the role of the fusogen eff-1 and natural genetic variation in *Caenorhabditis elegans* seam cell development

Institution name Imperial College London

Expected presentation date Apr 2020

Portions Figure 1

Requestor Location Ms. Sneha Koneru
Imperial College London
Sir Alexander Fleming building
London, SW7 2AZ
United Kingdom
Attn: Ms. Sneha Koneru

Publisher Tax ID GB 494 6272 12

**SPRINGER NATURE LICENSE
TERMS AND CONDITIONS**

Sep 23, 2019

This Agreement between Ms. Sneha Koneru ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4674940672737
License date	Sep 23, 2019
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Springer eBook
Licensed Content Title	Wnt Signaling Polarizes C. elegans Asymmetric Cell Divisions During Development
Licensed Content Author	Arielle Koonyee Lam, Bryan T. Phillips
Licensed Content Date	Jan 1, 2017
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Will you be translating?	no
Circulation/distribution	500 - 999
Author of this Springer Nature content	no
Title	Research Assistant/Ph.D student
Institution name	Imperial College London
Expected presentation date	Oct 2019
Portions	Results Probl Cell Differ. 2017; 61: 83–114. Figure 3 from the following chapter "Wnt signaling polarizes C. elegans asymmetric cell divisions during development"
Requestor Location	Ms. Sneha Koneru 6 Elsham Road Flat 1 London, United Kingdom W14 8HA United Kingdom Attn: Ms. Sneha Koneru

From: permissions permissions@biologists.com
Subject: RE: Permission to use a figure
Date: 3 October 2019 at 09:12
To: Koneru, Sneha L s.koneru@imperial.ac.uk
Cc: permissions permissions@biologists.com

Dear Sneha ,

Thank you for your request. Permission is granted for thesis use without charge.

The acknowledgement is very important and should state "reproduced/adapted with permission" and give the source journal name - the acknowledgement should either provide full citation details or refer to the relevant citation in the article reference list - the full citation details should include authors, journal, year, volume, issue and page citation.

Where appearing online or in other electronic media, a link should be provided to the original article (e.g. via DOI).

Journal of Cell Science: <http://www.biologists.com/journal-of-cell-science>

We wish you the best of luck with your thesis.

Kind regards

Richard

Richard Grove
Commercial Manager
The Company of Biologists Ltd
Bidder Building, Station Road, Histon, Cambridge, CB24 9LF, UK
T: +44 (0) 1223 632 850 | richard.grove@biologists.com | www.biologists.com

Registered office: The Company Of Biologists Ltd, Bidder Building, Station Road, Histon, Cambridge CB24 9LF, United Kingdom, Registered in England and Wales. Company Limited by Guarantee No 514735. Registered Charity No 277992 The information contained in this message and any attachment is confidential, legally privileged and is intended for the addressee only. Any dissemination, distribution, copying, disclosure or use of this message/attachment or its contents is strictly prohibited and may be unlawful. No contract is intended or implied, unless confirmed by hard copy. If you have received this message in error, please inform the sender and delete it from your mailbox or any other storage mechanism. The Company of Biologists Ltd cannot accept liability for any statements made which are clearly the senders' own and not expressly made on behalf of The Company of Biologists Ltd or one of their agents.

From: Koneru, Sneha L <s.koneru@imperial.ac.uk>
Sent: 27 September 2019 18:02
To: permissions <permissions@biologists.com>
Subject: Permission to use a figure