

**INTEGRATIVE ANALYSIS OF SPACE ENVIRONMENT RESPONSIVE  
GENE NETWORKS**

by

Irem Celen

A dissertation submitted to the Faculty of the University of Delaware in partial  
fulfillment of the requirements for the degree of Doctor of Philosophy in  
Bioinformatics and Systems Biology

Spring 2019

© 2019 Irem Celen  
All Rights Reserved

**INTEGRATIVE ANALYSIS OF SPACE ENVIRONMENT RESPONSIVE  
GENE NETWORKS**

by

Irem Celen

Approved: \_\_\_\_\_  
Cathy H. Wu, Ph.D.  
Chair of Bioinformatics and Computational Biology

Approved: \_\_\_\_\_  
Levi T. Thompson, Ph.D.  
Dean of the College of Engineer

Approved: \_\_\_\_\_  
Douglas J. Doren, Ph.D.  
Interim Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

---

Chandran R. Sabanayagam, Ph.D.  
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

---

Cathy H. Wu, Ph.D.  
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

---

Shawn Polson, Ph.D.  
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

---

Sevinc Ercan, Ph.D.  
Member of dissertation committee

## ACKNOWLEDGMENTS

I would like to thank my Ph.D. advisor, Dr. Chandran Sabanayagam, for giving me the intellectual freedom to become an independent scientist. His support, encouragement, and open-mindedness have greatly helped me step out of my comfort zone and push the boundaries in science.

My master's advisor and committee member Dr. Cathy Wu has always been an inspirational role model for me. Her invaluable leadership and professionalism have inspired me to achieve more both in academic and personal life. I appreciate her support over the years.

I thank my committee members Dr. Sevinc Ercan and Dr. Shawn Polson for their helpful feedback and guidance. I have learned a lot from their scientific expertise.

I am grateful to Republic of Turkey Ministry of National Education for enabling me to continue my graduate studies abroad, the University of Delaware for Graduate, Dissertation, and Summer Doctoral Fellowships, and Sigma Xi Scientific Honor Society for Grant-in-Aid of Research award.

I wish to thank my friends from CBCB. Especially, I thank Mengxi Lv and Maddy MacDonald (Stinner) for being great teammates. I had great time running the Bioinformatics Student Association together and enjoyed their friendship. Also, many thanks to Sushmita Paul who was an awesome lab-mate.

Special thanks to my dearest friends Huseyin Faik Sahin, Sultan Simsek Tanriseven, and Sinem Ersever Celik. I have always felt their support and love even if we are far apart.

I would like to thank Agni who has been a great support for me for many years.

Finally, I am grateful to my mother Berrin and grandmother Necla. I could not have achieved this goal without their support, understanding, and love.

## TABLE OF CONTENTS

LIST OF TABLES .....	x
LIST OF FIGURES .....	xii
ABSTRACT .....	xx

### Chapter

1	INTRODUCTION .....	1
1.1	Background.....	1
1.1.1	Effects of Microgravity on Health.....	4
1.1.2	Regulatory Gene Networks .....	5
1.1.3	Guilt-by-association .....	10
1.2	Related Work.....	10
2	EFFECTS OF LIQUID CULTIVATION ON GENE EXPRESSION AND PHENOTYPE OF <i>C. elegans</i> .....	18
2.1	Abstract.....	18
2.2	Background.....	19
2.3	Results .....	22
2.3.1	Distinct Gene Groups are Expressed Exclusively in Specific Environmental Conditions.....	25
2.3.2	Environment, Genotype, and Environment-Genotype Specific Expression of Genes .....	30
2.3.3	N2 Strain Presents Higher Differential Expression under Environment Changes .....	33
2.3.4	Functional Profiles of the Differentially Expressed Genes Show Correlation with the Observed Phenotype.....	48
2.4	Discussion.....	52
2.5	Conclusions .....	56
2.6	Methods .....	57
2.6.1	<i>C. elegans</i> strains and growth conditions .....	57

2.6.2	RNA Isolation, Illumina Sequencing .....	57
2.6.3	qRT-PCR .....	58
2.6.4	Bioinformatics Analyses .....	58
2.6.5	Microscopy .....	59
3	<b>INTERGENERATIONAL GENOME-WIDE CHROMATIN REMODELING IN <i>C. elegans</i> UNDER ENVIRONMENTAL CHANGES ..</b>	<b>61</b>
3.1	Abstract.....	61
3.2	Introduction .....	62
3.3	Results .....	64
3.3.1	Defining a core epigenetic memory and its role in adaptation ....	72
3.3.2	Epigenetic memory-forming abilities in response to environment change.....	77
3.3.3	Multivalence of histone marks changes during adaptation .....	78
3.3.4	Rapid histone modification dynamics during embryogenesis.....	83
3.3.5	Impact of histone modifications on gene expression .....	86
3.4	Discussion.....	88
3.5	Materials and methods.....	93
3.5.1	ChIP-seq Analysis .....	93
3.5.2	Immunocytochemistry and Microscopy .....	95
3.5.3	Statistical Analyses.....	96
3.5.4	RNA-seq Analysis .....	97
	Clustering and motif analysis.....	97
4	<b>COMPARATIVE TRANSCRIPTOMIC SIGNATURE OF THE SIMULATED MICROGRAVITY RESPONSE IN <i>Caenorhabditis elegans</i>..</b>	<b>98</b>
4.1	Abstract.....	98
4.2	Background.....	99
4.3	Results .....	102
4.3.1	Simulated microgravity triggers differential expression of hundreds of genes .....	104
4.3.2	Simulated microgravity-induced gene expression differences are highly maintained for eight days after return to ground conditions .....	106
4.3.3	Longevity regulating pathways are affected under simulated microgravity .....	111

4.3.4	Sphingolipid signaling pathway is suppressed in response to simulated microgravity .....	114
4.3.5	Identification of the “gravitome” through comparison of the microgravity-exposed worm gene expression profiles .....	117
4.4	Discussion.....	125
4.5	Conclusions .....	128
4.6	Materials and Methods .....	129
4.6.1	<i>C. elegans</i> strain and growth conditions.....	129
4.6.2	RNA isolation, Illumina sequencing .....	130
4.6.3	Gene expression analysis.....	130
4.6.4	Functional analysis of the genes.....	131
4.6.5	Comparison of DEGs to the previous studies in <i>C. elegans</i> .....	132
4.6.6	Mass spectrometry for ceramides .....	133
4.6.7	Microgravity simulation with clinorotation.....	134
5	INTEGRATIVE ANALYSIS OF ENVIRONMENTAL STRESS RESPONSIVE GENE NETWORKS AND THEIR DRIVERS IN <i>Caenorhabditis elegans</i> .....	136
5.1	Abstract.....	136
5.2	Background.....	138
5.3	Results .....	140
5.3.1	Co-expression network genes have biologically relevant functions to environmental stress conditions .....	151
5.4	Discussion.....	158
5.5	Methods .....	160
5.5.1	Data recruitment .....	160
5.5.2	Data preprocessing .....	161
5.5.3	Weighted gene co-expression analysis (WGCNA).....	161
5.5.4	Network Evaluation.....	163
5.5.5	Network Visualization, Hub Gene and Transcription Factor-Target Determination.....	164
6	CONCLUSION AND FUTURE DIRECTIONS .....	165
	REFERENCES .....	167

Appendix

A PERMISSIONS ..... 190

## LIST OF TABLES

Table 2.1: Sequencing read information for the RNA-seq libraries.....	23
Table 2.2: Gene expression comparison with a previous axenic medium study. Genes downregulated in axenic medium versus OP50 NGM (red downregulated; green upregulated gene expression; FPKM > 1 if bolded).....	38
Table 2.3: Gene expression comparison with a previous axenic medium study. Genes upregulated in axenic medium versus OP50 NGM (red downregulated; green upregulated gene expression; FPKM > 1 if bolded).....	39
Table 2.4: Differentially expressed long (purple), thin (pink), or long and thin (orange) phenotype genes under CeHR conditions in N2 and AB1 strains.....	41
Table 2.5: Enriched pathways in F1 generations in response to the liquid cultivations. The green and red arrows show upregulation and downregulation compared to the expression in the corresponding P0 generations. The P-values are Benjamini-Hochberg corrected .....	43
Table 3.1: Sequencing read information for the ChIP-seq libraries .....	66
Table 4.1: An overview of the <i>Caenorhabditis elegans</i> studies at the International Space Station. PRJNA146465 represents the study conducted by Universiti Kebangsaan Malaysia (UKM).....	102
Table 4.2: Previously reported (Higashibata et al., 2016) reproducible gene expression changes from previous spaceflight studies in <i>C. elegans</i> . The values represent the logFC of the gene expressions under microgravity versus ground control. NR: not reported, NA: not available, * (Higashibata et al., 2016), ** (Higashibata et al., 2006), *** (Then et al., 2016), # GEO2R, ## Cuffdiff2, ### DESeq2.....	120
Table 5.1: Number of modules positively correlated with the tested environmental conditions .....	142
Table 5.2: Gene modules and their comparison to previously known gene- and protein-networks show that the majority of the modules have significant overlaps	

with previously identified networks. Red represents the results that are not significant ..... 145

## LIST OF FIGURES

- Figure 2.1: Exposure of the domesticated and wild-isolate worms to environmental changes to induce a transcriptomic response. (a) Experimental design for multigenerational studies in N2 and AB1 strains of *C. elegans*. (b) Dendrogram and principal component analyses show a clear separation strains of *C. elegans*. (c) Dendrogram and principal component analyses show a clear separation strains of *C. elegans*. (d) Dendrogram and principal component analyses show a clear separation cultured on agar plates. .... 24
- Figure 2.2: Different strains and environmental conditions present variations in transcriptional responses. (a) Dendrogram created based on the gene expression (FPKM) values for the strains and the environmental conditions. (b) Principal component analysis plot of gene expression data. (c) Categorization of the expressed genes in the three experiments. Red labels represent the percentages of ncRNAs. (d) The barplots demonstrate ncRNA profiles in the three experiments. (e) Categorization of the previously unconfirmed genes that show expression in our experiments. .... 26
- Figure 2.3: The gene expression trends in response to environmental changes vary between the strains. N2 strain shows an overall decrease in transcripts when exposed to environmental changes. .... 27
- Figure 2.4: The ratio of the expressed mRNA and ncRNA molecules between the P0 and F1 generations. Both the molecules in the N2 strain show a decrease in response to the environmental changes. In comparison to the decrease in mRNA levels, the proportion of expressed ncRNA molecules are lower under exposure to environmental changes. .... 29

- Figure 2.5: Shown are the number of gene expressions exclusive to environment, genotype, or genotype-environment interactions. The number of the expressed ncRNAs are decreased under stress conditions. The table depicts the enriched gene ontology (GO) terms assigned to environment or genotype specifically expressed genes. Genotype-environment interaction genes did not demonstrate enrichment for GO terms. “None” represents the groups with no GO enrichment. Benjamini Hochberg corrected P-values for the GO terms: \* < 0.01; \*\* < 0.001. .... 31
- Figure 2.6: Quantitative Real Time PCR (qRT-PCR) Analysis of differentially expressed RNA. (a) qRT-PCR were conducted with total RNA extracted from N2 animals in grown on OP50 NGM, CeHR and CeHR Reversion (back onto OP50 NGM). All genes, with the exception of T16G1.13, analyzed via qRT-PCR were consistent with the RNA-seq data and analysis of the differential gene expression. (b) Pearson correlation of the qRT-PCR and RNA-seq data. .... 34
- Figure 2.7: The differential expression of the genes (FPKM > 1, FDR adj. P-value < 0.05, and  $\log_2(\text{fold change of the FPKM values}) \geq 2$ ) between the environment conditions and the strains. Red dashed lines represent the two-fold cutoff for differential gene expression. The corresponding genes for the enriched GO terms were represented with the colors. Benjamini Hochberg corrected and FDR adj. P-values < 0.05 for all the enriched GO terms. The GO terms were determined separately for downregulated and upregulated genes. .... 36
- Figure 2.8: Different strains and the environment conditions contribute to differences in the maternal gene expression profiles. (a) The DEGs in the F1, F2, and both (F1F2) generations were identified in comparison to the gene expressions in the P0 animals. Between 5 to 34% of the DEGs in the F1 generations were transmitted to the F2 generations (see the third category – F1F2) and these transmissions are higher than expected by chance (hypergeometric test, P-value < 0.001 for all the upregulated and downregulated genes individually). The gene expressions transmitted to the next generation after exposure to CeHR showed distinct enrichments for sequence motifs in their promoter regions for N2 and AB1 strains. The number of DEGs genes in F1, F2, and F1F2 were represented with blue, green, and purple, respectively. (b) Top five tissue enrichment of the DEGs under exposure to environmental changes in F1 (FDR-adj. P-value < 0.05). ... 45

Figure 2.9: Putative transcription factors functioning in response to environmental changes in N2 and AB1. (a) One motif from each strain showed enrichment as a recognition site for transcription factor MDL-1. The same sequence motifs demonstrated differences for the overrepresented biological functions in human. (b) Venn diagrams depicting the overlap of the differentially expressed transcription factors between environmental conditions or strains. The overlap between the F1 generations are statistically significant (Hypergeometric test, <i>P-value</i> < 0.0001) .....	47
Figure 2.10: Norm of reaction between the alae morphology variant annotated genes and the detected alae morphology variance between AB1 and N2 indicating a potential role for these genes in the observed phenotype. ....	50
Figure 2.11: Quantitative differences in the observed phenotypes in liquid cultivations. T-test was used to compare the conditions. Error bars represent the standard error. **: <i>P-value</i> < 0.01, ***: <i>P-value</i> < 0.0001. <i>P-value</i> > 0.05 is not reported. ....	51
Figure 3.1: Experimental design for transgenerational inheritance studies. ....	64
Figure 3.2: Distribution of the peaks on gene regions for each histone modification..	65
Figure 3.3: Static and elastic epigenetic memory of histone marks. (a) Legend for the Venn diagrams representing the potential roles of the histone modification associated genes in the conditions/generations. (b) Legend for the number of genes associated with the histone marks in generations. The ratio of the core epigenetic memory genes was used as a criterion to decide on whether histone modifications show static or elastic memory. (c) The numbers in the Venn diagrams denote the number of genes associated with histone marks individually or commonly in conditions, and the common genes found in both CeHR and S-medium experiments. Conserved sequence motif enrichments on the core epigenetic memory genes indicate that sequence recognition plays an essential role in transgenerational epigenetic memory. Abbreviations: RNA Pol II: RNA polymerase II.....	70
Figure 3.4: Number of genes associated with elastic histone marks individually or commonly in conditions, and the common genes found in both CeHR and S-medium experiments. Enriched sequence motifs for core epigenetic memory genes suggests that epigenetic memory occurs through recognition of sequence motifs. ....	73

Figure 3.5: Evaluation of the static/dynamic nature of histone marks. (a) Hierarchical clustering of the histone modifications based on the expressions of the genes that they associate with. Static H3K27me3 marked genes tend to cluster together. (b) Distribution of histone marks on chromosomes under distinct environmental conditions. Static histone modifications show a constant distribution on the chromosomes over generations while the elastic ones demonstrate variances. ....	76
Figure 3.6: Genome browser view of the three pentavalently marked gene promoters. The same data range (0 – 665) was used for all the presented data. ....	79
Figure 3.7: Multivalent nature of the individual histone marks. (a) Gene-histone mark networks for the three generations under different environmental conditions and the percentage of the gene promoters co-occupied by one to five histone marks (x-axis represents the number of marks). (b) Motif enrichment for the genes associated with multiple histone marks in P <sub>0</sub> generation indicates a sequence-specific recognition of the multivalently marked gene promoters. (c) Percentage of the multivalence or monovalence levels of the individual histone marks in P <sub>0</sub> generation. ....	80
Figure 3.8: Histone modification associated gene promoters showing multivalence and monovalence in core epigenetic memory and epigenetic memory forming. Multivalence for core epigenetic memory is 5.9% and 20% for the CeHR experiment and the S-Medium experiment, respectively. Multivalence in the epigenetic memory forming is 12.4% and 47.8% for the CeHR experiment and the S-Medium experiment, respectively. ....	82
Figure 3.9: Structured illumination microscopy of fluorescently-stained histone marks. The figure represents the short-term dynamics of the histone marks during embryogenesis and co-localized marks for P <sub>0</sub> animals. Scale bars: 3µm. ....	84
Figure 3.10: Histone modification multivalency causes changes in the gene expressions. (a) Tukey-Kramer test for the gene expressions among chromosomes for histone modification associated genes. The represented gene expressions are from both monovalently and multivalently marked genes. Means with the same letter are not significantly different. (b) K-means clustering for the histone modification associated gene expressions and the enriched gene ontology terms assigned to each cluster. Each color represents a separate cluster. ....	88

Figure 3.11: Pipeline development for ChIP-seq data analysis. (a) Workflow for the pipeline development. (b) Comparison of peak calling softwares for broad peaks called for H3K27me3..... 95

Figure 4.1: Transcriptomic response of *C. elegans* to simulated microgravity and return to ground conditions. (a) Clinostat used for simulating microgravity. (b) Experimental design for the effect of simulated microgravity on gene expressions during the exposure and after the return. (c) Dendrogram of the gene expression profiles for ground condition (GC), simulated microgravity (MG), four days after return to ground conditions (R1\_GC), eight days after return to ground conditions (R2\_GC), and twelve days after return to ground conditions (R3\_GC). (d) Log2 fold change of the gene expressions (FPKM) between the conditions..... 103

Figure 4.2: Differentially expressed gene profiles under simulated microgravity and after return to ground conditions. (a) Tissue enrichment of the upregulated (green) and downregulated (red) genes under simulated microgravity. (b) Pathway enrichment of the upregulated (green) and downregulated (red) genes during the exposure to simulated microgravity (MG), and four and eight days after return to ground conditions (R1\_GC and R2\_GC, respectively). (c) The number differentially expressed genes under simulated microgravity and the transmission of the differential expression after return to ground conditions. (d) Categorization of the upregulated genes in comparison to the ground control animals, and the enriched gene ontology terms assigned to them. (e) Categorization of the downregulated genes in comparison to the ground control animals, and the enriched gene ontology terms assigned to them. .... 105

Figure 4.3: Phenotype enrichment analysis of the DEG neuropeptide signaling genes indicate a movement variation under simulated microgravity.... 107

Figure 4.4: Simulated microgravity induced ncRNA molecules. (a) Percent of the expressed ncRNA molecules in ground control (GC), simulated microgravity (MG), four-, eight-, and twelve-days after return to ground conditions (R1\_GC, R2\_GC, R3\_GC). (b) The ncRNA molecules that are expressed during the exposure to simulated microgravity and twelve days after the return but not in ground conditions. (c) The ncRNA molecules that lost their expression and did not gain after the return. .... 109

- Figure 4.5: Putative transcription factor genes that are upregulated under simulated microgravity (MG), four days (R1 GC), eight days (R2 GC), and twelve days (R3 GC) after return to ground conditions. Y56A3A.28 and T24C4.2 are also induced in four-day CERISE. .... 111
- Figure 4.6: Longevity regulating pathway genes are differentially expressed under simulated microgravity. Adopted from the KEGG longevity regulating pathway – worm (cel04212)..... 113
- Figure 4.7: Sphingolipid signaling pathway is downregulated under simulated microgravity. (a) The sphingolipid signaling pathway genes *asah-1*, *asm-3*, and *gba-4* are downregulated under simulated microgravity. The downregulation pattern of *asm-3* and *gba-4* is maintained for eight days after return to ground conditions. (b) The levels of d18:1 sphingosine, total ceramide, hexosylceramide, and sphingoid base reduced while total lactosylceramide level increased under simulated microgravity. (c) The levels of different acyl-chain ceramides show alterations under simulated microgravity. The error bars represent the standard error of the mean (\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ) .... 115
- Figure 4.8: RNA-seq data analysis with DESeq2 (a) PCA analysis of the RNA-seq data from the ground control (GC), simulated microgravity (MG), and four-, eight-, and twelve-day after return to ground conditions. (b) The number differentially expressed genes under simulated microgravity and the transmission of the differential expression after return to ground conditions from the analysis with DESeq2. (c) Categorization of the upregulated (left) and downregulated (right) genes in comparison to the ground control animals, and the enriched gene ontology terms assigned to them from the analysis with DESeq2..... 121

Figure 4.9: Gravitome genes are consistently differentially expressed in the ISS and under simulated microgravity. (a) Shared DEGs among the space-flown worms and our simulated microgravity experiment. The highest number of common DEGs are between our experiment and four-day CERISE. These genes are named as the “gravitome”. (b) The hierarchical clustering of the overall gene expressions among the space-flown worms and our simulated microgravity experiment. Our results analyzed with two different data analysis pipelines demonstrated highly similar patterns. (c) The number common of DEGs were high for our results analyzed with two different pipelines. Both of the pipelines showed common number of DEGs are higher than the values expected by chance (hypergeometric test, $p < 0.05$ ). (d) Protein-protein interaction of the gravitome genes from STRING database. (e) Gene ontology enrichment of the gravitome genes. (f) Protein domain enrichment of the gravitome genes.....	123
Figure 4.10: Domain enrichment of the human orthologs of the gravitome.....	131
Figure 5.1: Experimental design for the gene co-expression network generation and downstream analyses.....	141
Figure 5.2: Module-trait relationship for the identified modules and the environmental conditions tested (red, blue, and numbers in parentheses represent strong positive and strong negative correlations and the p-values, respectively) .....	143
Figure 5.3: Module membership vs gene significance for the six mostly correlated modules with simulated microgravity (MG). .....	147
Figure 5.4: Module membership vs gene significance for the six mostly correlated modules with glucose supplemented diet.....	147
Figure 5.5: Module membership vs gene significance for the six mostly correlated modules with vitamin C supplemented diet .....	148
Figure 5.6: Module membership vs gene significance for the six mostly correlated modules with three-week CeHR cultivation .....	148
Figure 5.7: Module membership vs gene significance for the six mostly correlated modules with four-day CeHR cultivation .....	149
Figure 5.8: Module membership vs gene significance for the six mostly correlated modules with S-Medium cultivation .....	149

Figure 5.9: Module membership vs gene significance for the six mostly correlated modules with glucosamine supplemented diet.....	150
Figure 5.10: Module membership vs gene significance for the six mostly correlated modules with OP50 NGM growth condition .....	151
Figure 5.11: Top five GO biological process (BP) enrichments for the PPIs of the darkviolet module.....	152
Figure 5.12: Top five KEGG and Reactome pathway enrichments for the PPIs of the darkviolet module.....	153
Figure 5.13: Darkviolet module is correlated with the vitamin C supplement and four-day CeHR cultivation conditions. The transcription factors <i>daf-16</i> and <i>hlh-30</i> and their targets are co-expressed. ....	154
Figure 5.14: GO enrichment analysis for the known protein-protein interactions in the darkslateblue module.....	156
Figure 5.15: Pathway enrichment analysis for the known protein-protein interactions in the darkslateblue module .....	157
Figure 5.16: Darkslateblue module is correlated with simulated microgravity exposure and glucosamine supplemented diet conditions. The transcription factors <i>hmg-12</i> and <i>efl-1</i> and their targets are co-expressed in this module. ....	158
Figure 5.17: Scale free topology model fit and mean connectivity for soft threshold selection.....	162
Figure 5.18: Dendrogram of the co-expressed genes for the TOM-based dissimilarity. The colors represent the resulting modules after the dynamic tree cut algorithm and the merge of the highly correlated modules. ....	163

## ABSTRACT

Environmental changes are important factors that can cause crucial health and survival consequences for all domains of life. As such, twin studies indicated that environmental stressors account for more cancer cases than the inherited genetic code for particular cancer types. Space conditions present drastically different environmental conditions, such as different radiation levels and gravitational force. Given the elevated interest in human exploration of space, it is important to uncover the space-affected biological processes at systems level. Currently, the studies at the International Space Station (ISS) introduce a combination of environmental stressors including microgravity exposure and dietary changes. Separate studies are required to discover the individual impact of each environmental stressor on biological processes.

First, we investigated the intergenerational transcriptomic response to liquid cultivation and the response differences in a wild-isolate and the laboratory strain of *C. elegans*. The results revealed that the wild-isolate and the laboratory strain show vastly different transcriptomic responses to the same diet. Approximately 8% - 34% of liquid cultivation-induced transcriptomic responses are transmitted to the subsequent generation.

Second, we tested the intergenerational dynamics of five histone modifications (H3K4me3, H3K9me3, H4K20me1, H3K27me3, H3K36me1) under liquid cultivation

to better understand some of the underlying epigenetic mechanisms. Our results depicted that while H3K4me3 and H3K27me3 mostly maintained their distribution, the others demonstrated highly dynamic distributions on the chromosomes. These results indicate an adaptive function for H3K9me3, H4K20me1, and H3K36me1 in response to liquid cultivations.

Third, we evaluated the sole impact of simulated microgravity during and after exposure. Through the integration of our in-house data with the publicly available datasets from NASA's GeneLab, we determined a "gravitome" which we defined as the microgravity responsive transcriptomic signatures which are conserved between the worms and humans.

Finally, we identified the Weighted Gene Co-expression Networks (WGCNA) for space responsive gene networks by integrating our RNA-seq data with publicly available ones from Gene Expression Omnibus for dietary changes (vitamin C, glucose, and glucosamine supplement) in worms. We identified 50 co-expressed gene networks (modules) and their putative drivers. We believe our combinatory approach can be generally applicable to other environmental stressors to facilitate the exploration of disease-driving genetic responses and create potential therapeutic targets.

# Chapter 1

## INTRODUCTION

### 1.1 Background

Environmental changes have profound effects on well-being and survival of organisms. One striking example of this is that global warming has resulted in (>60%) mass coral reef bleaching and mortality in the Great Barrier Reef in 2016 (Hughes et al., 2017). This bleaching event triggers a cascade affecting the whole marine life starting from the algal symbionts of the coral reefs. The health of individuals is also greatly impacted by environmental changes. Exposure to toxins increases the risk of Parkinson's disease (Chinta et al., 2013), early life stress increases the future risk of cardiovascular disease (M. O. Murphy, Cohn, & Loria, 2017), and prenatal famine increases the risk of type 2 diabetes (Lumey, Khalangot, & Vaiserman, 2015), to name a few examples.

Even though the genetic factors (i.e., inherited genetic code) have been the main focus of interest for the disease studies, the correlation between genetics and disease is very minute. As a matter of fact, environmental changes had been reported to account for 70% - 90% of the chronic disease risks (Rappaport, 2016; Rappaport & Smith, 2010). Particularly, with twin studies, it became more prominent that the contribution of environmental factors to disease is much higher than of genetic factors

for many diseases. For example, in a study conducted on 44,788 pairs of twins, only 28% of the stomach cancer cases were found to be related to genetic factors and the rest was found to be associated with environmental factors (Lichtenstein et al., 2000). Given the crucial role of environmental factors on both human health and overall wellbeing of living systems, it is vital to identify the affected mechanisms (e.g., metabolic and transcriptomic networks) and their controlling mechanisms in systems level. Identification of these mechanisms can enable the development of personalized therapeutics to alleviate the impact of environmental stress on health.

How living organisms cope with abrupt environmental changes, and how the genetic and metabolic mechanisms react to these changes are not entirely understood. *Caenorhabditis elegans* (*C. elegans*) is an ideal model organism with completely documented cell lineage, high reproduction rate, short lifespan, and high similarity to the human genome and nutritional requirements (Çelen, Doh, & Sabanayagam, 2017; Félix & Braendle, 2010; Hunt-Newbury et al., 2007; Lehner, Crombie, Tischler, Fortunato, & Fraser, 2006; Watson & Walhout, 2014). *C. elegans* enables the usage of green fluorescent protein (GFP) in the detection of spatiotemporal gene expressions by being a small and transparent organism (Watson, MacNeil, Arda, Zhu, & Walhout, 2013). In addition, the worms allow studies on the whole organism rather than a cell line. This organism has been widely used in adaptation studies due to its ability to live in both terrestrial and aquatic conditions and to consume bacterial or axenic food source. Placing the worms from bacteria seeded agar plates (OP50 NGM) to an aquatic axenic media poses a drastic change for the animals. For example, the animals

display locomotive differences in the liquid by swimming via a rapid C-shape thrashing motion as opposed to crawling via a sinusoidal configuration on solid (Pierce-Shimomura et al., 2008). Behavior and physiological differences arise when the worms are grown in aquatic axenic media (such as CeMM; *C. elegans* Maintenance Medium or CeHR; *C. elegans* Habitation and Reproduction) CeMM (Nass & Hamza, 2007; Szewczyk, Kozak, & Conley, 2003). In axenic media, the worms exhibit phenotypical alterations such as delayed development, reduced fecundity, increased longevity, and altered body morphology (Doh, Moore, Çelen, Moore, & Sabanayagam, 2016; Houthoofd et al., 2002a; Szewczyk et al., 2006). Environment changes can trigger different gene expression profiles and these profiles can be transmitted to the subsequent generations (Daxinger & Whitelaw, 2010; Heard & Martienssen, 2014; Jaenisch & Bird, 2003). Changing the environment of the animals from agar to axenic media, therefore, can cause transcriptome-level alterations which may have a sustained effect on the progenies.

In its natural habitat, *C. elegans* occupies microbiota-rich decaying plant material and experiences a continuous environment shift between aquatic and terrestrial conditions along with the variations in its diet (Félix & Braendle, 2010). The commonly used wild-type laboratory strain of *C. elegans* (N2) has been cultured mainly on bacteria-seeded agar plates for over 40 years (Brenner, 1974). It has been reported that the domestication of the N2 strain maintained in the laboratory environment has created a selective pressure on the animals (Sterken, Snoek, Kammenga, & Andersen, 2015). Some of the essential questions that await answers

are whether: (i) the adaptive responses depend on the genotype or the environment change; (ii) the environment change-induced gene expression profiles are transmitted to the successive generation in equal levels for different genotypes; (iii) the environment change-induced gene expression profiles are transmitted to the next generation equally for different environmental conditions.

Environmental changes can cause differences in gene expression, and these changes can be transmitted to the next generations via an epigenetic memory (Daxinger & Whitelaw, 2010; Heard & Martienssen, 2014; Jaenisch & Bird, 2003). Histone modifications are involved in numerous biological processes, including longevity (Greer et al., 2011), X-chromosome repression (Gaydos, Wang, & Strome, 2014), and proper germline programming (Rechtsteiner et al., 2010), and this involvement can affect multiple generations. For example, previous studies with *C. elegans* revealed that H3K4me3 on the longevity-related genes are retained between generations (Greer et al., 2011). Understanding the transgenerational epigenetics heavily depends on understanding the weight of contribution each mark has on the epigenetic memory. However, it is largely unknown whether all histone modifications show an epigenetic memory, and if so, whether there are identifiable differences in the epigenetic memory levels among the histone modifications.

### **1.1.1 Effects of Microgravity on Health**

Significant efforts have been made to colonize Mars. NASA and SpaceX have planned round-trip landings to Mars by 2030s. SpaceX CEO Elon Musk has recently

explained their plan to place a million people to Mars in as early as 40 years. In 2013, there were already over 200,000 applications to live in the “Red Planet”. However, microgravity conditions in space has been shown to negatively affect biological processes by causing changes in the transcriptome leading to bone loss, muscle atrophy, and immune system impairment, to name a few examples. Hence, it is crucial to uncover the biological systems that are negatively affected by microgravity before colonization of Mars. One cost effective way to study the effect of microgravity is to simulated microgravity. For this purpose, various platforms including drop towers, parabolic flights and space flights are available. Because of their advantages to enable long term simulations in an effective and economical manner, clinostats are widely preferred to simulate microgravity (Wuest, Stern, Casartelli, & Egli, 2017).

### **1.1.2 Regulatory Gene Networks**

Biological processes are highly complex systems which require the involvement of numerous cellular molecules. Studies for determining the function of single genes or proteins are, therefore, not sufficient for understanding the systems level mechanisms. Efforts have been made to identify functional gene relationship networks *in silico* and *in vivo*. Supervised machine learning algorithms, such as support vector machines, and mutual information-based tools are among the most extensively used computational methods. One limitation of these methods is that their performance tends to show variations depending on the selected algorithm, parameters, and the requirement of a priori knowledge (Zhu, Panwar, & Guan, 2016).

The attempts including *in vivo* strategies have their advantages and limitations, as well. For instance, Yeast Two-Hybrid approach allows the detection of interacting proteins. However, this approach shows bias towards high-affinity interactions and has limitations in detecting self-activating proteins, proteins requiring posttranslational modifications, and membrane proteins (Brückner, Polge, Lentze, Auerbach, & Schlattner, 2009). The combination of forward and reverse genetic screens has also been applied for functional gene network identification, yet it can provide information about a limited number of genes (Watson et al., 2013). In addition, *in vivo* techniques are generally costlier and time consuming compared to the computational ones. Integration of *in silico* and *in vivo* techniques can be a promising approach to achieve more comprehensive and accurate functional gene networks.

A gene co-expression network represents a group of genes which show similar expression patterns across samples. Identification of the co-expressed genes is of great biological importance as these genes can share common biological functions or transcriptional regulation machinery. Thus, this approach can be used to uncover the involvement of genes in biological processes or disease states. However, it should be noted that gene co-expression networks do not indicate causality or provide a classification for regulatory or regulated genes as they consist of undirected edges. The advent of high throughput gene expression profiling technologies, such as microarray and RNA-sequencing, has enabled the genome-wide identification of gene expression dynamics under varying conditions. Data produced by these technologies are commonly used for creating gene co-expression networks. Indicating their

effectiveness, gene co-expression networks have been successfully used to analyze data from different species including yeast, worms, mice, humans, and plants (Carlson et al., 2006; T. F. Fuller et al., 2007; B. Kim, Suo, & Emmons, 2016; Wei et al., 2006; Xue et al., 2014).

Pearson correlation, Spearman correlation, partial correlation, and Bayesian approach are the most commonly used approaches in the construction of gene co-expression networks (Ala et al., 2008; Friedman, Linial, Nachman, & Pe'er, 2000; Langfelder & Horvath, 2008; Villa-Vialaneix et al., 2013). Since Bayesian approach does not perform well in large datasets and is more computationally expensive, the correlation-based methods are more widely preferred (Allen, Xie, Chen, Girard, & Xiao, 2012; Serin, Nijveen, Hilhorst, & Ligterink, 2016).

Weighted Gene Co-expression Network Analysis (WGCNA) is a systems biology method identifying correlation patterns among genes based on their expression patterns (Carlson et al., 2006; Langfelder & Horvath, 2008). This method enables the determination of the highly-correlated gene clusters (modules) and their association to sample traits. Because it is well-documented and efficient, WGCNA is a widely used method.

In WGCNA, genes are represented in nodes and labeled with indices  $i, j = 1, 2, \dots, n$ , and the pairwise correlation between expression of two genes is represented in edges. *Adjacency matrix*  $a_{ij}$ , a symmetric matrix with values zero or one, of a network is employed to show the connection strength between nodes  $i$  and  $j$ . The adjacency

matrix is calculated with co-expression similarity  $s_{ij}$  where  $s_{ij}$  is the absolute value of the correlation coefficient between nodes  $i$  and  $j$ :

$$s_{ij} = |\text{cor}(x_i, x_j)|$$

where  $x_i$  and  $x_j$  represent the expression of genes (nodes)  $i$  and  $j$ . The absolute value of the correlation coefficients is used to define unsigned networks. In correlation networks, the strongest negative correlation is symbolized with -1 and the strongest positive correlation is symbolized with 1. By using absolute correlation values, unsigned networks eliminate the negative values. Therefore, there exists a risk that highly negatively correlated and positively correlated genes are grouped in an unsigned network. Such gene groups mainly lose their ability to represent biologically meaningful networks. Signed networks overcome this issue by scaling the correlation values between zero and one where  $<0.5$  shows a negative and  $>0.5$  shows a positive correlation (van Dam, Vósa, van der Graaf, Franke, & de Magalhães, 2017) by using the following equation:

$$s_{ij} = 0.5 + 0.5\text{cor}(x_i, x_j)$$

In unweighted gene networks, the connections are represented with zero or one to show that genes are either connected or unconnected. In weighted networks, all genes are connected to each other. The strength of the connections is represented by continuous variables ranging between zero and one where one is assigned to the strongest co-expression. Previous studies have demonstrated that weighted gene networks provide more robust results compared to the unweighted ones by

overcoming the loss of information and sensitivity to the choice of the threshold problems of hard thresholding (B. Zhang & Horvath, 2005). Application of soft thresholding, introduced by Zhang and Horvath, enables the construction of weighted networks by employing the following function:

$$a_{ij} = (s_{ij})^\beta$$

where the soft thresholding parameter is represented with the power  $\beta$ .

After the construction of the network, WGCNA seeks to detect subsets of nodes (modules) which are clusters of genes with highly correlated expression profiles across samples. Towards that end, frequently used WGCNA methods is the employment of hierarchical clustering combined with topological overlap matrix (TOM) based dissimilarity which is followed by dynamic branch cutting methods (T. Fuller, Langfelder, Presson, & Horvath, 2011; Langfelder & Horvath, 2008; B. Zhang & Horvath, 2005). Gene ontology or pathway enrichment analyses are commonly used to examine whether the modules are biologically meaningful and not just noise. Moreover, the biologically meaningful modules can be investigated for their potential correlation with traits, such as cholesterol levels, body weight, or a disease (T. F. Fuller et al., 2007; Langfelder & Horvath, 2008; van Dam et al., 2017).

Highly connected genes in co-expression networks are called hub genes and these genes play essential roles in biological processes (Langfelder, Mischel, & Horvath, 2013; van Dam et al., 2017). The hub genes have been identified as the key drivers to signaling pathways by numerous studies. For example, Chou et al. detected

the hub genes in endometrial cancer (EC) and showed that these genes have high predictive power in the identification of the grade, type, and stage of EC (Chou, Cheng, Brotto, & Chuang, 2014). Knockout experiments revealed that hub proteins are required for the survival of lower organisms such as yeast, fly, and worm (Langfelder et al., 2013). In a study conducted by Lehner et al., it has been shown that hub genes work as genetic buffers by maintaining the overall biological outcomes and they are conserved across animal species (Lehner et al., 2006). Yet, it is important to note that hub genes do not always contribute to the observed phenotype nor they are necessarily essential genes. In fact, a study by Rosenkrantz et al. suggested that non-essential genes are predominantly observed as hub genes for the stress-adaptive gene networks in *Salmonella* (Rosenkrantz et al., 2013).

### **1.1.3 Guilt-by-association**

Guilt-by-association is a commonly used approach to identify the biological functions of less studied genes. This approach uses the assumption that highly correlated gene networks share similar biological functions. Guilt-by-association has been successfully used to determine the missing genes in biological pathways (B. Kim et al., 2016) and the regulatory disease genes (van Dam et al., 2017).

## **1.2 Related Work**

Several studies have been conducted in *C. elegans* to understand how changes in the environmental conditions affect the worms. For example, phenotypical

differences have been studied in response to dietary changes (Croll, Smith, & Zuckerman, 1977; Doh et al., 2016; Gems & Riddle, 2000; Houthoofd et al., 2002b, 2002a; Szewczyk et al., 2006; Watson et al., 2013; Watson & Walhout, 2014). Similar to this study, other groups have conducted experiments on liquid grown worms. Particularly, phenotypical changes caused by axenic liquid media has been the main study of interest. For instance, Szewczyk et al. evaluated the development, fecundity rates, lifespan, lipid and protein stores, and gene expression changes in CeMM (Szewczyk et al., 2006). While their study provides a comprehensive analysis for phenotypic traits, the gene expression analysis has several limitations. In their microarray analysis, Szewczyk et al. could manage to obtain reliable data for only 1,202 cDNAs on all three arrays consistently (Szewczyk et al., 2006). Hence, their results clearly miss many important adaptive genes. To our knowledge, there is no complete genome-wide gene expression analysis for finding the adaptive responses to axenic liquid cultures.

Others have reported that the domestication of the worms in the laboratory environment has created a selective pressure on the animals (Sterken et al., 2015). Impressive progress has been made for the quantitative trait loci analyses of wild-isolate and domesticated *C. elegans* strains (Denver, Morris, & Thomas, 2003; Haber et al., 2005; McGrath et al., 2009). However, how the laboratory domestication affects rapid adaptive gene expressions under environmental changes remains largely unknown.

To date, four studies from three missions to the International Space Station (ISS) have been conducted in *C. elegans*. However, the experimental design of these studies highly varies while usage of microarrays is common for all four. The first study was International *C. elegans* Experiment First (ICE-First) and it was conducted by Japan Aerospace Exploration Agency in 2004. They used mixed stage N2 strain *C. elegans* and maintained the worms on ground control and at the ISS for 10 days in axenic CeMM (Higashibata et al., 2006). This study reported only 24 differentially expressed genes. The usage of mixed stage animals high likely has caused the signal of stage-specific gene expressions to be diluted and resulted in detection of a limited number of differentially expressed genes. The second set of worms were sent to the ISS for six months by Universiti Kebangsaan Malaysia (UKM) in 2006. In this study, CC1 strain *C. elegans* was maintained in CeMM. Interestingly, however, their results did not show consistency with the other space studies and our results (explained later).

The third and fourth experiments (CERISE-4 and CERISE-8) were conducted by the Japan Aerospace Exploration Agency in 2009. Briefly, by using onboard 1G and ground 1G controls, the synchronized L1 worms were cultivated for four days and eight days, respectively. Differently from the ICE-First experiment, researchers cultured the worms in *E. coli* containing S-Medium in this experiment. Since the worms stay synchronized in four days, CERISE-4 experiment produced relatively robust results. However, in eight days, mixed stage embryos are observed as the worms start to produce progeny. Another issue with CERISE-8 is that the starting culture for this experiment was reported as 50 L1 larvae and this number is very

limited for obtaining statistically meaningful results. Potentially because of these artifacts in the design of CERISE-8, the results from CERISE-4 and CERISE-8 are drastically different from each other (explained later). Overall inconsistency among the space biology experiments highlights the necessity of a more organized and uniform experimental design for future studies. In addition, to detect low abundance transcripts and achieve less noise in the data, RNA-seq can be preferred over microarray technologies (Zhao, Fung-Leung, Bittner, Ngo, & Liu, 2014). The potential confounders at the ISS should be taken into account, as well. For instance, the ISS presents elevated radiation levels, microgravity, and the animals experience hypergravity during the launch. These mixed factors make it impossible to know what precisely caused the observed biological impact.

Model organism encyclopedia of DNA elements (modENCODE) project was initiated in 2007 to systematically annotate the functional genomic elements in *C. elegans* and *Drosophila melanogaster* (Gerstein et al., 2010). The project has provided a comprehensive analysis of numerous histone modifications across developmental stages of *C. elegans*. However, most of the data is generated by using ChIP-chip instead of ChIP-seq. Given that ChIP-seq performs better than ChIP-chip by providing higher sensitivity, specificity, spatial resolution, dynamic range, and genomic coverage, it is expected to observe inconsistencies between the results from these two technologies (Park, 2009). In fact, Ho et al. reported that the correlation between replicates of ChIP-chip and ChIP-seq experiments is only 0.41 (median  $r = 0.41$ ) (Ho et al., 2011). This inconsistency highlights the fact that there may exist noise in the

ChIP-chip generated data in modENCODE. In addition, the ChIP-seq data are either from embryonic or L3 larva stages and only two histone modifications are from the L4 larva stage in the ChIP-chip data in modENCODE. This fact limits researchers to compare their results from the L4 stage of *C. elegans*. Another limitation of modENCODE is that the data mainly were generated from domesticated N2 strains cultured on agar plates. Overall, modENCODE mostly lacks the information from wild-isolates and environmental stress-exposed worms as well as the trans- and inter-generational histone modification dynamics.

Both *in vivo* and *in silico* studies exist in literature to profile gene networks in *C. elegans* (Junyue Cao et al., 2017; B. Kim et al., 2016; Lehner et al., 2006; McKay et al., 2003; Norris, Gracida, & Calarco, 2017; Watson et al., 2013). In one of the earlier studies, McKay et al. constructed green fluorescent protein (GFP) fused genes to identify the expression patterns on developmental stages and tissues of human orthologs in *C. elegans* (McKay et al., 2003). Usage of promoter::GFP fusion allows single-cell resolution in the detection of gene expressions. Additionally, the same study utilized the serial analysis of gene expression (SAGE) technique to profile the temporal and tissue-specific gene expressions quantitatively. Extending this concept, Hunt-Newbury et al. used promoter::GFP fusion technique and produced expression profiles for 1,886 genes in 38 different cell and tissue types (Hunt-Newbury et al., 2007). Their approach enabled the detection of the co-expression in single cell level. Importantly, the expression data from these studies are publicly available at WormAtlas and WormBase databases. By using the interactive tool in WormAtlas

(<http://www.gfpworm.org/>), researchers can find the tissue-specific expressions for their gene of interest and order any of the studied strains from the Caenorhabditis Genetics Center for their study. However, as these two studies limited their research to predicted human orthologs and could only cover approximately 10% of the *C. elegans* genes.

In 2006, Lehner et al. systematically tested ~65,000 pairs of genes for genetic interactions by using RNA interference (RNAi) and found 349 genetic interactions among 162 genes (Lehner et al., 2006). Their approach could identify eight out of thirteen (61.5%) known EGF pathway genes. While this approach is robust, it has several limitations such as having high false negative rates (Kamath et al., 2003) and knocking down the gene expression instead of knocking-out. To overcome the limitations of the RNAi screen, Norris et al. have established a CRISPR/Cas9-based Synthetic Genetic Interaction approach for a systematic double-mutant generation (Norris et al., 2017). With their method, they evaluated 14 conserved RNA-binding protein genes for possible interactions and the phenotypical changes of the double mutants. However, their method only focused on a limited number of genes which are known to be expressed in the same tissue. In addition, CRISPR studies require more time and effort compared to RNAi via siRNAs (Boettcher & McManus, 2015).

Several studies have been conducted to characterize the environment-responsive gene networks. For instance, Watson et al. combined forward and reverse genetic screens to unravel the genetic networks affected under dietary changes (i.e. standard *E. coli* diet to *Comamonas* DA1877 diet) in *C. elegans* (Watson et al., 2013).

By performing forward genetics, Watson et al. identified genes causing activation of dietary sensor *Pacdh-1::GFP* when mutated. Combination of their forward genetics with reverse genetics via an RNAi screen enabled the identification of a genetic network of 184 genes affected by the dietary change. However, this study is limited to a single genetic network playing a role in a specific dietary change. Moreover, they mention that their genetic network is not complete as ORFeome RNAi library that they used has only half of all protein-coding genes of *C. elegans*.

Integration of computational and experimental methods have been employed for detection of more comprehensive gene networks. To generate a functional network for *C. elegans* early embryogenesis, Gunsalus et al. integrated protein interaction, gene co-expression, and phenotypic similarity data (Gunsalus et al., 2005). Their study has proven that applying computational methods for hypothesis development, and the investigation of these hypotheses with experimental methods enable the detection of larger networks. In a more recent study, Kim et al. conducted RNA-seq on male and hermaphrodite worms in five developmental stages at 6-hour intervals starting from the late L3 larval stage (B. Kim et al., 2016). They pooled their data from each time point and sex to create a gene co-expression network with WGCNA. By using a guilt-by-association approach they identified additional synaptic, ciliary, and semen genes. Briefly, they determined the unannotated genes whose expression show a high correlation to the expression of synapse, ciliary, and semen formation genes. They validated their approach by promoter::*GFP* tagging the correlated unannotated genes and examining the expression of these genes on the related tissues. However, this

study has several issues that can cause inaccuracies in the statistical results. First, pooling the data from different sexes and developmental stages would introduce batch effects during the creation of gene networks. Second, the study does not have replicates for the RNA-seq data which can cause the experiment to have lower statistical power. In addition, the developers of WGCNA recommends the minimum number of samples to be 15, but the study by Kim et al. used only 10 samples to build their network. A low number of samples can give rise to noisy networks and loss of biological meaningfulness. Even though this study provides a valuable way to determine and expand the biological networks, caution must be used in statistical analyses to avoid batch effects in such studies.

## Chapter 2

### EFFECTS OF LIQUID CULTIVATION ON GENE EXPRESSION AND PHENOTYPE OF *C. elegans*

#### 2.1 Abstract

##### Background

Liquid cultures have been commonly used in space, toxicology, and pharmacology studies of *Caenorhabditis elegans*. However, the knowledge about transcriptomic alterations caused by liquid cultivation remains limited. Moreover, the impact of different genotypes in rapid adaptive responses to environmental changes (e.g., liquid cultivation) is often overlooked. Here, we report the transcriptomic and phenotypic responses of laboratory N2 and the wild-isolate AB1 strains after culturing P<sub>0</sub> worms on agar plates, F<sub>1</sub> in liquid cultures, and F<sub>2</sub> back on agar plates.

##### Results

Significant variations were found in the gene expressions between the N2 and AB1 strains in response to liquid cultivation. The results demonstrated that 8% - 34% of the environmental change-induced transcriptional responses are transmitted to the subsequent generation. By categorizing the gene expressions for genotype, environment, and genotype-environment interactions, we identified that the genotype has a substantial impact on the adaptive responses. Functional analysis of the

transcriptome showed correlation with phenotypical changes. For example, the N2 strain exhibited alterations in both phenotype and gene expressions for germline and cuticle in axenic liquid cultivation. We found transcript evidence to approximately 21% of the computationally predicted genes in *C. elegans* by exposing the worms to environmental changes.

## **Conclusions**

The presented study reveals substantial differences between N2 and AB1 strains for transcriptomic and phenotypical responses to rapid environmental changes. Our data can provide standard controls for future studies for the liquid cultivation of *C. elegans* and enable the discovery of condition-specific genes.

## **2.2 Background**

All living systems possess the fundamental property to regulate physical or genetic states in response to environmental stimuli. *Caenorhabditis elegans* offers one of the unique models to study the effects of environmental changes as it can reproduce in large numbers and has a natural ability to live in both solid and liquid conditions. This eukaryotic organism provides the opportunity to study genetic mechanisms in the whole animal rather than a cell culture. In addition, it can consume a bacterial or an axenic food source thereby enabling the identification of diet-related biological processes. Placing the worms from bacteria seeded agar plates (OP50 NGM) to liquid axenic media poses a drastic change for the animals. For example, the animals display

locomotive differences in the liquid by swimming via a rapid C-shape thrashing motion as opposed to crawling via a sinusoidal configuration on solid (Pierce-Shimomura et al., 2008). Behavior and physiological differences arise when the worms are grown in liquid axenic media (such as CeMM; *C. elegans* Maintenance Medium or CeHR; *C. elegans* Habitation and Reproduction) CeMM (Nass & Hamza, 2007; Szewczyk et al., 2003). In axenic media, the worms exhibit phenotypical alterations such as delayed development, reduced fecundity, increased longevity, and altered body morphology (Doh et al., 2016; Houthoofd et al., 2002a; Szewczyk et al., 2006). Environment changes can trigger different gene expression profiles, and these profiles can be transmitted to the subsequent generations (Daxinger & Whitelaw, 2010; Heard & Martienssen, 2014; Jaenisch & Bird, 2003). Changing the environment of the animals from agar to axenic media, therefore, can cause transcriptome-level alterations which may have a sustained effect on the progenies. However, the knowledge about transcriptomic responses to axenic media is limited. For example, in their microarray analysis, Szewczyk et al. could manage to obtain reliable data for only 1,202 cDNAs on all three arrays consistently for CeMM-grown worms (Szewczyk et al., 2006). The common usage of axenic media in space, toxicology, and pharmacology studies highlights the necessity of a complete transcriptomic analysis, preferably with RNA-sequencing (RNA-seq) (Doh et al., 2016; Selch et al., 2008).

In its natural habitat, *C. elegans* occupies microbiota-rich decaying plant material and experiences a continuous environment shift between liquid and solid conditions along with the variations in its diet (Félix & Braendle, 2010). The wild-type

laboratory strain of *C. elegans* (N2) has been cultured in both liquid and solid conditions (Brenner, 1974), but bacteria-seeded agar plates are used more commonly (Stiernagle, 2006). It has been reported that the domestication of the N2 strain has created a selective pressure on the animals (Sterken et al., 2015). Some of the critical questions that await answers are whether: 1.) the adaptive responses depend on the genotype or the environment change; 2.) the environment change-induced gene expression profiles are transmitted to the successive generation in equal levels for different genotypes; 3.) the environment change-induced gene expression profiles are transmitted to the next generation equally for different environmental conditions.

To study the maternal effects on transcriptomic responses of *C. elegans* from solid to liquid environments, we maintained the P<sub>0</sub> worms on bacteria-seeded agar plates and F<sub>1</sub> in liquid cultures. By culturing F<sub>2</sub> generation back on agar, we identified the maintained maternal transcriptional responses. The N2 strain and a wild isolate AB1 strain were used to examine if a different genotype attributes to the transcriptional dynamics. It is important to note that even though the N2 and the AB1 strains are relatively close to each other in the phylogenetic tree (Barrière et al., 2005), AB1 is not the ancestral strain of N2. Therefore, the observed differences in norm of reaction can stem from the natural genetic differences between N2 and AB1 as much as the impact of laboratory domestication on adaptive responses. We found substantial variances in gene expression profiles resulting from different genotypes, environments, and genotype-environment interactions. Our results from RNA-seq data analysis and microscopy showed consistency for the predicted and observed

phenotype. Furthermore, we demonstrated that a large number of genes are expressed only in specific growth conditions, and many of these genes previously lacked experimental expressed sequence tag (EST) or cDNA evidence.

### **2.3 Results**

To induce an adaptive response and observe its sustained effects on the following generation, the P<sub>0</sub> generation *C. elegans* was grown on agar, the F<sub>1</sub> generation in liquid, and the F<sub>2</sub> generation back on agar (Figure 2.1a). We carried out three sets of experiments with different food sources and animal strains. For the first experiment, the F<sub>1</sub> generation N2 strain was grown in axenic CeHR Medium. To examine the effect of dietary changes in the liquid cultures, we cultured the F<sub>1</sub> generation N2 strain in bacterial S-Medium for the second experiment. Finally, the transcriptomic response differences were investigated by growing F<sub>1</sub> generation AB1 strain in CeHR Medium. For each generation, RNA-seq was performed on young adult animals with three replicates to determine the transcriptional responses (Table 2.1).

Table 2.1: Sequencing read information for the RNA-seq libraries

Strain	Condition	Replicate	Generation	Stage	Number of Reads	Mapped Reads
N2	Agar (OP50 NGM)	1	P0	Young adult	19,785,469	19,033,093 (96.2%)
N2	Agar (OP50 NGM)	2	P0	Young adult	17,804,139	17,078,192 (95.9%)
N2	Agar (OP50 NGM)	3	P0	Young adult	17,584,017	16,792,356 (95.5%)
N2	CeHR	1	F1	Young adult	29,616,506	27,904,230 (94.2%)
N2	CeHR	2	F1	Young adult	30,144,650	28,728,679 (95.3%)
N2	CeHR	3	F1	Young adult	29,749,460	27,582,792 (92.7%)
N2	CeHR Reversion	1	F2	Young adult	25,771,272	24,627,700 (95.6%)
N2	CeHR Reversion	2	F2	Young adult	49,360,215	46,951,509 (95.1%)
N2	CeHR Reversion	3	F2	Young adult	6,524,144	6,256,402 (95.9%)
N2	S-Medium	1	F1	Young adult	24,031,853	22,294,902 (92.8%)
N2	S-Medium	2	F1	Young adult	25,924,449	22,848,853 (88.1%)
N2	S-Medium	3	F1	Young adult	26,580,009	19,716,312 (74.2%)
N2	S-Medium Reversion	1	F2	Young adult	21,340,354	20,685,738 (96.9%)
N2	S-Medium Reversion	2	F2	Young adult	2,824,111	2,727,187 (96.6%)
N2	S-Medium Reversion	3	F2	Young adult	49,930,813	47,900,763 (95.9%)
AB1	Agar (OP50 NGM)	1	P0	Young adult	21,217,270	20,230,031 (95.3%)
AB1	Agar (OP50 NGM)	2	P0	Young adult	16,180,777	15,393,012 (95.1%)
AB1	Agar (OP50 NGM)	3	P0	Young adult	25,889,871	24,612,562 (95.1%)
AB1	CeHR	1	F1	Young adult	15,358,876	14,477,230 (94.3%)
AB1	CeHR	2	F1	Young adult	15,636,461	14,762,856 (94.4%)
AB1	CeHR	3	F1	Young adult	14,469,569	13,649,882 (94.3%)

AB1	CeHR Reversion	1	F2	Young adult	16,467,847	15,708,866 (95.4%)
AB1	CeHR Reversion	2	F2	Young adult	17,444,758	16,618,698 (95.3%)
AB1	CeHR Reversion	3	F2	Young adult	22,730,048	21,629,136 (95.2%)

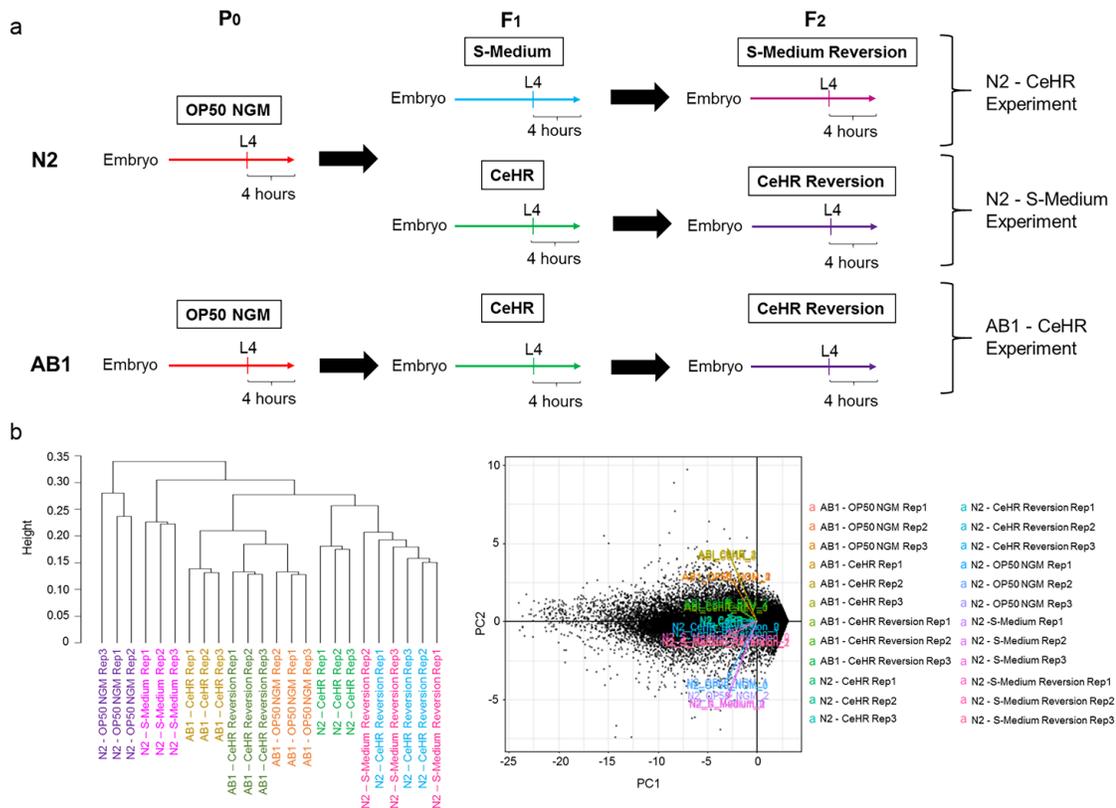


Figure 2.1: Exposure of the domesticated and wild-isolate worms to environmental changes to induce a transcriptomic response. (a) Experimental design for multigenerational studies in N2 and AB1 strains of *C. elegans*. (b) Dendrogram and principal component analyses show a clear separation strains of *C. elegans*. (c) Dendrogram and principal component analyses show a clear separation strains of *C. elegans*. (d) Dendrogram and principal component analyses show a clear separation cultured on agar plates.

### **2.3.1 Distinct Gene Groups are Expressed Exclusively in Specific Environmental Conditions**

The transcriptome-based clustering of the experiments in the dendrogram of Jensen-Shannon distances demonstrated that the genotypes are separated from each other (Figure 2.2a). Exposure to different environmental conditions causes the N2 transcriptome to be separated even further than the AB1 transcriptome as illustrated by the dendrogram and principal component analysis (Figure 2.2a, b , Figure 2.1b). These results suggest that the genotype plays a key role in adaptive response, and the different strains of the animals can show highly distinct transcriptome profiles under environmental changes.

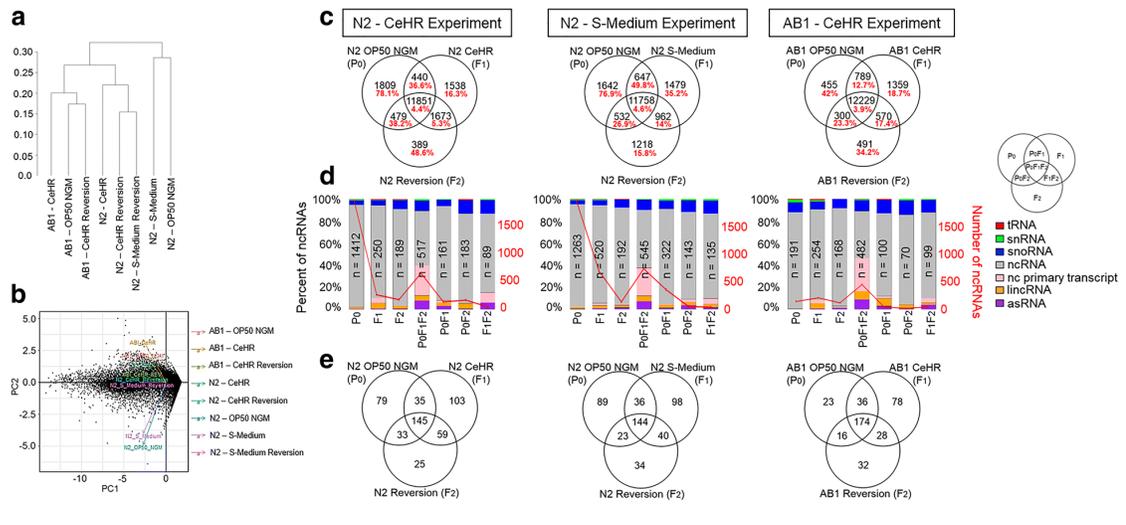


Figure 2.2: Different strains and environmental conditions present variations in transcriptional responses. (a) Dendrogram created based on the gene expression (FPKM) values for the strains and the environmental conditions. (b) Principal component analysis plot of gene expression data. (c) Categorization of the expressed genes in the three experiments. Red labels represent the percentages of ncRNAs. (d) The barplots demonstrate ncRNA profiles in the three experiments. (e) Categorization of the previously unconfirmed genes that show expression in our experiments.

We categorized the genes by using Venn diagrams to identify if their expressions are condition specific (Figure 2.2c). We considered the genes expressed if their Fragments Per Kilobase of transcript per Million mapped reads (FPKM) values are greater than one as this is a commonly used criterion for expression (Blazie et al., 2015; Duff et al., 2015; Kramer et al., 2015; Prabh & Rödelsperger, 2016). The majority of the expressed genes are conserved among the P<sub>0</sub>, F<sub>1</sub>, and F<sub>2</sub> generations. Nevertheless, hundreds of genes are expressed only under particular conditions. The number of the uniquely expressed genes showed a significant difference between

CeHR grown AB1 and N2 strains, but not within the N2 strain experiments (Chi-squared test with Yates correction;  $P$ -value = 0.01 and  $P$ -value = 0.9, respectively). In addition, the N2 strain showed an overall decrease in the transcription levels when exposed to the changes in growth conditions (Figure 2.3). AB1 strain, however, demonstrated an increase in the total FPKM but a decrease in the average FPKM under CeHR (Figure 2.3). Interestingly, the number of gene expression patterns transmitted to the next generation after exposure to an environment change ( $F_1F_2$ ) varies among the experiments (Chi-squared test with Yates correction;  $P$ -value < 0.001). Together, these results seem to suggest that the transmission of the gene expression is not exclusively dependent on the genotype and the environment separately—but dependent on their interaction, whereas the condition-specific expression of genes is affected by the genotype.

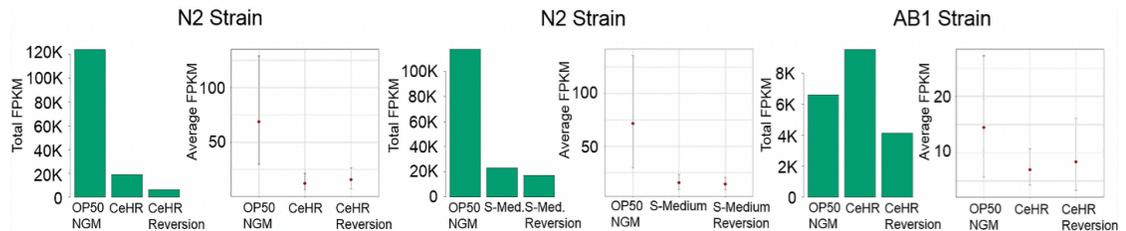


Figure 2.3: The gene expression trends in response to environmental changes vary between the strains. N2 strain shows an overall decrease in transcripts when exposed to environmental changes.

Further analyses revealed that non-coding RNA (ncRNA) molecules are enriched (16% - 78%) in our pool of condition-specific transcripts indicating their

function in rapid adaptive responses (Figure 2.2c). To note, only the polyadenylated ncRNA molecules were included in the analysis, and piRNA and miRNA molecules were excluded from the analysis due to their short lengths. The ncRNA molecules were predominantly observed in the P<sub>0</sub> generation of N2 strain compared to the P<sub>0</sub> AB1 strain. Moreover, fewer ncRNA molecules were expressed when the animals are exposed to an environment change (Figure 2.3c, d, and Figure 2.4). Our results indicate that ncRNA expression is mainly silenced when the worms are exposed to the tested environmental changes. The fact that a laboratory condition presents an environment change for the wild-isolate may explain why P<sub>0</sub> animals of AB1 showed less ncRNA expression compared to the N2 strain. One particular exception to this pattern is the ncRNA molecules that are commonly expressed in all three generations. We identified that the majority of these ncRNAs (59% - 66%, n = 322) are conserved among the three experiments. This conservation implies that this group of ncRNAs plays important regulatory functions in the cell and their expression is required regardless of the environmental changes. Overrepresented phenotype for these ncRNAs is dendrite development variant (FDR adj. *p-value* < 0.01, six-fold enrichment) suggesting a potential role for them in generation of neuronal extensions.

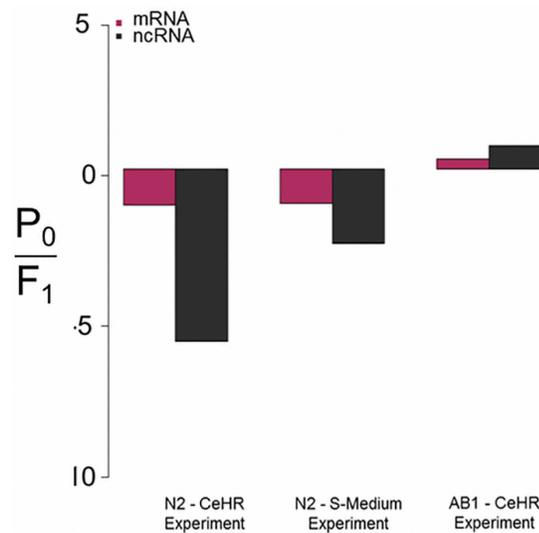


Figure 2.4: The ratio of the expressed mRNA and ncRNA molecules between the P<sub>0</sub> and F<sub>1</sub> generations. Both the molecules in the N2 strain show a decrease in response to the environmental changes. In comparison to the decrease in mRNA levels, the proportion of expressed ncRNA molecules are lower under exposure to environmental changes.

We reasoned that if a particular group of genes is only expressed under specific conditions, we might detect transcript evidence to computationally predicted but experimentally unconfirmed genes, especially considering that CeHR is not a common medium used in *C. elegans* research. Notably, we were able to identify transcript evidence to 21% of the list of unconfirmed genes from WormBase (Harris et al., 2014) by simply growing the animals in different environmental conditions (Figure 2.2e). Unexpectedly, a substantial amount of the transcript evidence was found with animals raised in commonly used OP50 NGM. We compared our list of expressed unconfirmed genes in OP50 NGM-grown N2 to those from a previous study (Dallaire, Proulx, Simard, & Lebel, 2014) and found that 54% of the unconfirmed genes from

our list (n = 292) have also shown expression in the study by Dallaire et al. but were not reported as transcript evidence in WormBase. The remaining 46% of the genes must have been induced due to the subtle differences introduced in laboratory environment (e.g., light and environment). Similar to the confirmed genes, gene expression transmission to the successive generation after environmental change demonstrated a difference between and within the strains (Chi-squared test with Yates correction;  $P$ -value < 0.001,  $P$ -value = 0.02, respectively). Given that it can robustly provide transcript evidence to unconfirmed genes, our approach and data can be a rich resource for discovery of novel genes.

### **2.3.2 Environment, Genotype, and Environment-Genotype Specific Expression of Genes**

We grouped the genes based on their genotype, environment, or genotype-environment interaction specific expressions (Figure 2.5). For this analysis, a more stringent criterion was used to eliminate any potential noise from the data. That is, we only considered the genes with expression levels higher than *pmp-3* of FPKM = 21, and excluded genes with FPKM < 21 from the analysis. This housekeeping gene was selected as a reference because of its stable expression levels found both in our analysis and previous studies (Hoogewijs, Houthoofd, Matthijssens, Vandesompele, & Vanfleteren, 2008; Yanqiong Zhang, Chen, Smith, Zhang, & Pan, 2012). Specifically for the ncRNA molecules, we reasoned that expression levels higher than that of a housekeeping gene should provide strong evidence for expression.

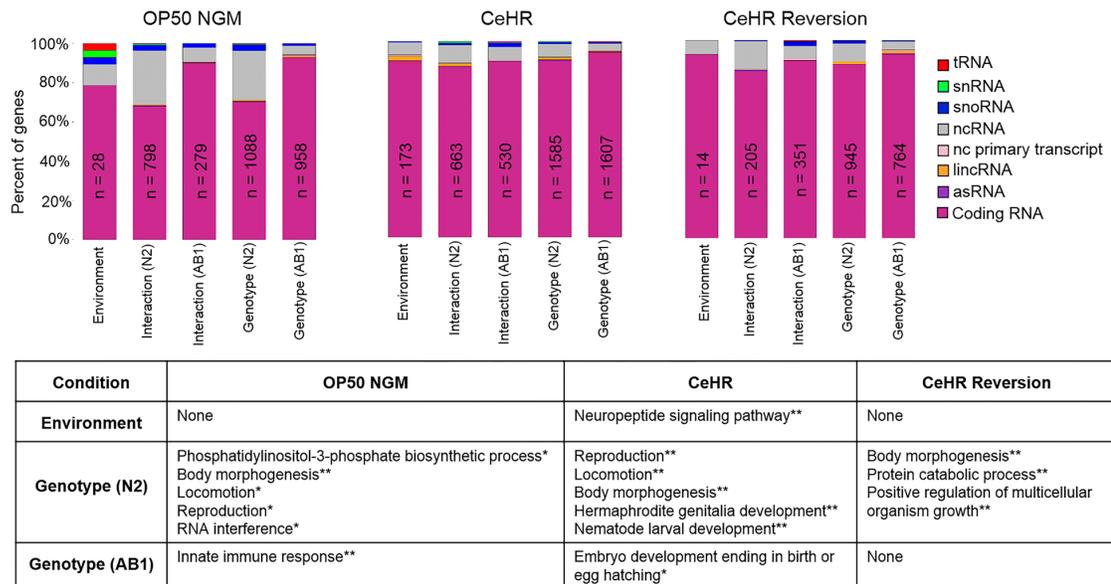


Figure 2.5: Shown are the number of gene expressions exclusive to environment, genotype, or genotype-environment interactions. The number of the expressed ncRNAs are decreased under stress conditions. The table depicts the enriched gene ontology (GO) terms assigned to environment or genotype specifically expressed genes. Genotype-environment interaction genes did not demonstrate enrichment for GO terms. “None” represents the groups with no GO enrichment. Benjamini Hochberg corrected P-values for the GO terms: \* < 0.01; \*\* < 0.001.

We defined the genes that are expressed in only one environmental condition for both the strains as “environment-specific”, and found that CeHR-specific gene expression displayed Gene Ontology (GO) enrichment for the neuropeptide signaling pathway. Given that this pathway functions in numerous behavioral activities such as reproduction, locomotion, mechanosensation and chemosensation, it is likely that the

neuropeptide signaling pathway plays a crucial role in rapid adaptation to CeHR and is evolutionarily conserved between the strains (C. Li & Kim, 2008).

The “genotype-specific” expression represents the genes found in only one of the strains under a single environmental condition. All the N2 specific genes showed GO enrichment for body morphogenesis in all the three generations while GO enrichment for reproduction and locomotion were observed in P<sub>0</sub> and F<sub>1</sub> generations (Figure 2.5). Previously, we reported that the N2 worms produce less progeny compared to the AB1 animals in CeHR (Doh et al., 2016). Our analysis showed that along with the reproduction genes, genes related to hermaphrodite genitalia development were expressed exclusively in the N2 strain in CeHR. This finding may suggest vulva aberrations in the reproductive system. In addition, the exclusive expression of the genes functioning in positive regulation of multicellular organism growth may help the animals readapt to the OP50 NGM condition in the F<sub>2</sub> generation of the N2 animals.

The genes expressed solely under a particular environment and strain were categorized as “genotype-environment interaction” genes (Figure 2.5). This group of genes did not demonstrate an enrichment of GO terms. The proportion of ncRNA molecules in our group of genes showed correlation with our hypothesis that the majority of the ncRNAs are silenced under the tested environmental changes. For example, the N2 worms showed relatively higher proportions of ncRNAs for interaction and genotype-related genes on OP50 NGM, but these proportions were much lower in more stressful CeHR and CeHR reversion (Figure 2.5).

### **2.3.3 N2 Strain Presents Higher Differential Expression under Environment Changes**

To investigate the differences in transcriptional responses between the strains and among the environmental conditions, we identified the differentially expressed genes (DEGs). We considered genes as differentially expressed if their FPKM is greater than one and their expression is altered over two-folds among the conditions with FDR-adjusted *p-values*  $\leq 0.05$ . The estimated fold change responses of twenty randomly selected transcripts (ten for OP50 NGM against CeHR and ten for CeHR against CeHR reversion) were confirmed by quantitative RT-PCR (Figure 2.6a, b).

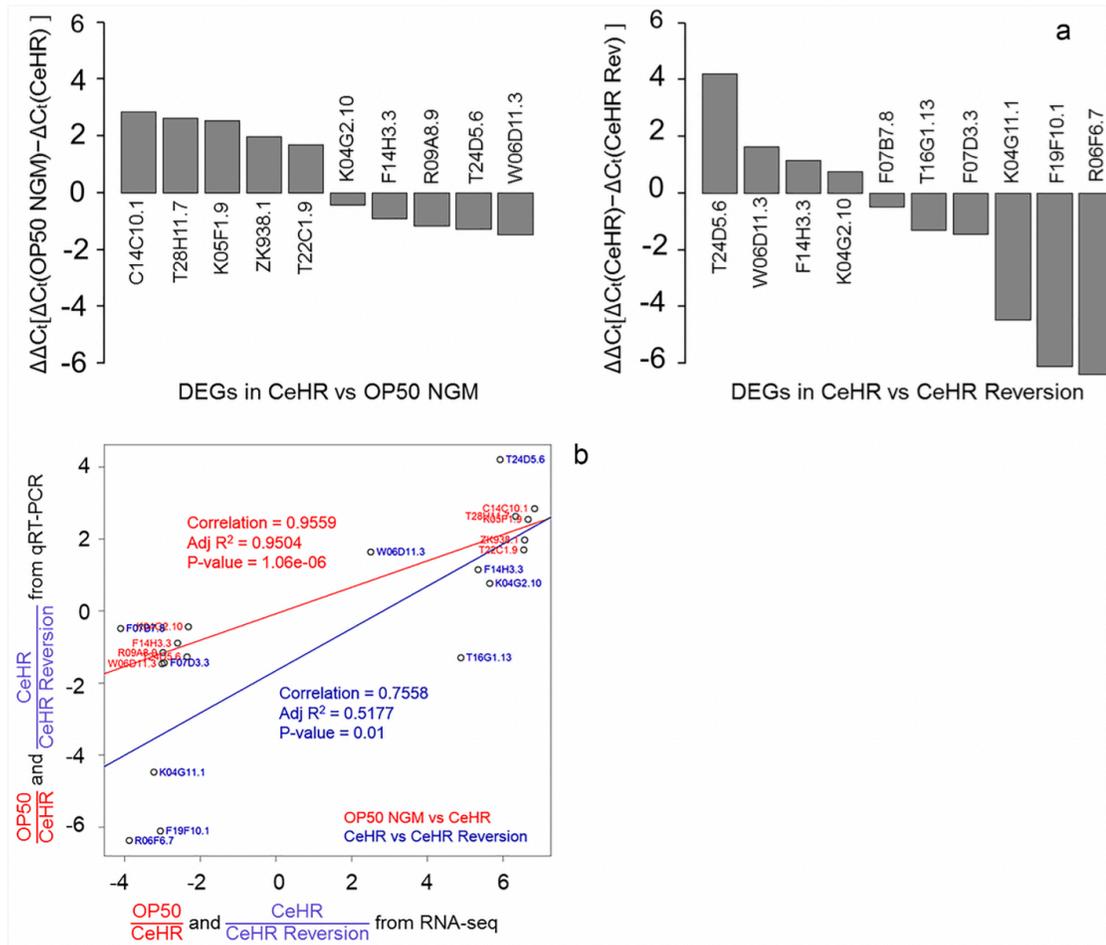


Figure 2.6: Quantitative Real Time PCR (qRT-PCR) Analysis of differentially expressed RNA. (a) qRT-PCR were conducted with total RNA extracted from N2 animals in grown on OP50 NGM, CeHR and CeHR Reversion (back onto OP50 NGM). All genes, with the exception of T16G1.13, analyzed via qRT-PCR were consistent with the RNA-seq data and analysis of the differential gene expression. (b) Pearson correlation of the qRT-PCR and RNA-seq data.

We found the highest number of DEGs between OP50 NGM and CeHR conditions for the N2 strain (Figure 2.7). This indicates that the change in diet and physical environment triggers more transcriptional responses for adaptation of the N2

strain. The number of DEGs were much lower for the F<sub>2</sub> generation of the same experiment, suggesting that the majority of the gene expressions started to resemble the P<sub>0</sub> animals. This finding was expected as most of the gene expression patterns return to the steady-state levels after a short period (López-Maury, Marguerat, & Bähler, 2008). The change in only the physical condition introduced by the S-Medium apparently did not cause as drastic transcriptional responses as the CeHR since the number of the DEGs were lower (38% and 45% lower number of upregulated and downregulated genes, respectively). However, S-Medium reversion conditions demonstrated higher DEGs compared to OP50 NGM. We note that we are able to maintain the worms in S-Medium for only one generation, potentially because of the toxins released by the bacteria in the liquid environment (Croll et al., 1977; Gems & Riddle, 2000). Therefore, additional genetic mechanisms may be required for re-adaptation of the animals from S-Medium. Transcriptional variations were more modest across the generations for AB1 animals than the N2 strain.

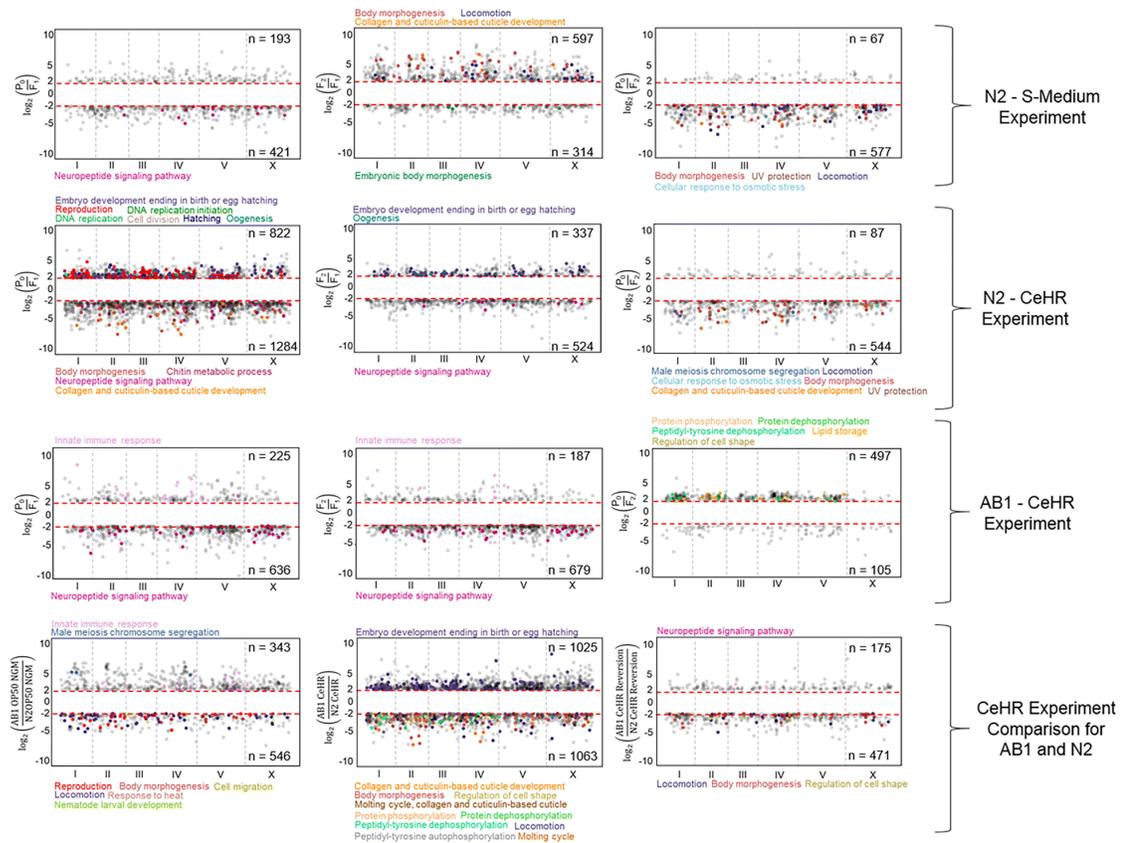


Figure 2.7: The differential expression of the genes ( $FPKM > 1$ , FDR adj. P-value  $< 0.05$ , and  $\log_2(\text{fold change of the FPKM values}) \geq 2$ ) between the environment conditions and the strains. Red dashed lines represent the two-fold cutoff for differential gene expression. The corresponding genes for the enriched GO terms were represented with the colors. Benjamini Hochberg corrected and FDR adj. P-values  $< 0.05$  for all the enriched GO terms. The GO terms were determined separately for downregulated and upregulated genes.

Substantial discrepancies have been reported for the *C. elegans* gene expressions among axenic media studies (Castelein, Hoogewijs, De Vreese, Braeckman, & Vanfleteren, 2008). We wished to compare our results to those from CeMM-grown worms as CeMM has a very similar base medium to CeHR. For the 21

upregulated and 26 downregulated genes reported in a previous study (Szewczyk et al., 2006), we could only identify one common downregulated gene in N2 and AB1 and four and seven common upregulated genes in N2 and AB1, respectively (Table 2.2 and Table 2.3, respectively). It is unclear why the DEGs among the axenic media studies mostly disagree, but some potential reasons can be the difference in the technology used and the modifications made in the media. For instance, microarray technology has been reported to introduce more background noise and cross-hybridization and detect a lesser number of DEGs compared to RNA-seq (Zhao et al., 2014). A more unified study conducted with RNA-seq may help resolve the discrepancy issues among the axenic media experiments.

Table 2.2: Gene expression comparison with a previous axenic medium study. Genes downregulated in axenic medium versus OP50 NGM (red downregulated; green upregulated gene expression; FPKM > 1 if bolded)

Transcript Name	Gene Name	Reported (Szewczyk et al., 2006)	RNA-seq (N2)	RNA-seq (AB1)
F59A1.10	dgat-2	2.12±0.16	1.6763	1.34785
C18C4.5	C18C4.5	2.09±0.17	1.23525	0.338265
T21B10.3	T21B10.3	2.06±0.07	1.21561	1.07386
F42H10.7	ess-2	2.23±0.16	0.982566	0.33833
K05C4.5	K05C4.5	2.31±0.22	1.08071	1.13907
ZK856.9	zhit-3	2.23±0.06	0.315686	0.887203
ZC395.8	ztf-8	2.58±0.15	0.639296	0.696436
T10F2.3	ulp-1	2.82±0.23	1.28647	0.771096
ZK593.8	fic-1	2.96±0.29	0.915745	0.0010156
F22B7.5	dnj-10	2.36±0.15	0.302129	0.281979
K12D12.1	top-2	2.07±0.12	1.54555	1.60623
T07A9.6	daf-18	3.21±0.32	1.61265	0.586463
Y39A1A.12	orc-1	2.70±0.15	<b>2.0381</b>	0.694941
F49E8.2	oef-1	2.27±0.11	1.79849	0.96101
ZC317.7	ZC317.7	2.39±0.16	0.367506	0.540345
T02B5.1	cest-1	2.78±0.17	1.67789	<b>2.24475</b>
ZK1320.9	ZK1320.9	2.44±0.22	0.231765	0.39998
F25D1.1	ppm-1	3.90±0.28	0.708939	0.52062
F56A6.1	sago-2	3.88±0.35	<b>2.45443</b>	1.43725
C55B7.1	glh-2	2.46±0.18	1.84883	1.80556
C10H11.1	viln-1	2.41±0.20	0.14354	0.179035
K07D8.1	mup-4	2.74±0.15	1.1408	1.43152
T22B7.7	T22B7.7	4.90±0.20	1.02483	<b>2.65249</b>
F59D8.C	vit-3	3.82±0.38	Low reads	Low reads
C47B2.6	gale-1	3.01±0.24	0.275766	1.08986
F02A9.4	F02A9.4	3.63±0.32	1.0187	1.01102

Table 2.3: Gene expression comparison with a previous axenic medium study. Genes upregulated in axenic medium versus OP50 NGM (red downregulated; green upregulated gene expression; FPKM > 1 if bolded)

Transcript Name	Gene Name	Reported (Szewczyk et al., 2006)	RNA-seq (N2)	RNA-seq (AB1)
R13H8.1a	daf-16	2.11±0.11	0.361921	0.22918
T20D3.6	T20D3.6	2.33±0.19	0.621571	0.299947
F31E3.6	F31E3.6	2.27±0.16	1.98507	0.189954
F35A5.2	F35A5.2	3.95±0.29	<b>2.55678</b>	<b>4.4825</b>
C18A11.2	C18A11.2	3.58±0.10	<b>2.81051</b>	<b>3.66397</b>
C39D10.2	C39D10.2	2.46±0.17	1.30746	<b>3.62688</b>
M03E7.4	M03E7.4	2.31±0.19	1.6438	<b>2.70823</b>
T08G5.10	mtl-2	3.92±0.36	0.243858	<b>6.35372</b>
T10B10.6	phat-6	3.80±0.21	<b>2.20061</b>	<b>5.70724</b>
F32H5.3	F32H5.3	3.15±0.09	<b>3.3075</b>	<b>2.74311</b>
C25E10.8	C25E10.8	2.15±0.14	1.09099	1.716
K11G9.6	mtl-1	2.85±0.27	1.52386	<b>6.75533</b>
F18E9.2	nlp-7	2.14±0.18	0.930323	<b>2.02463</b>
F36D3.9	cpr-2	7.64±0.18	<b>10.8343</b>	<b>10.2444</b>
F54E7.2	rps-12	2.37±0.20	1.25192	0.694212
F37C12.9	rps-14	2.28±0.16	0.976764	0.221806
F37C12.11	rps-21	2.60±0.20	1.49071	0.675332
F39B2.6	rps-26	2.10±0.15	0.912086	0.276755
B0412.4	rps-29	2.47±0.16	1.1284	0.760152
Y48B6A.2	rpl-43	2.49±0.12	0.284669	1.0072
F28F8.3	lsm-5	2.53±0.13	0.929627	0.287242
Y46G5.36	Not found	2.84±0.11	Not found	Not found

In our previous study, we have shown that CeHR-grown N2 and AB1 *C. elegans* display different phenotypical traits on their body morphology [6]. That is, adult animals of both strains demonstrated physical differences in CeHR compared to the OP50 NGM-grown ones, but the CeHR-grown AB1 strain was even longer and thinner than the CeHR-grown N2 strain in average (N2 length is  $825.2 \pm 74.3 \mu\text{m}$  and

width is  $45.9 \pm 10.1 \mu\text{m}$ , and AB1 length is  $876.8 \pm 67.0 \mu\text{m}$  and width is  $38.3 \pm 4.7 \mu\text{m}$ ). In accordance with this finding, the GO enrichment analysis results demonstrated enrichment for body morphogenesis for upregulated genes in CeHR compared to OP50 NGM in the N2 strain, but not the AB1 strain (Figure 2.7). Moreover, genes upregulated under the CeHR condition in the AB1 strain compared to the N2 strain were overrepresented for body morphogenesis. The fact that this GO term is not enriched for AB1 may imply that the phenotype is caused by a separate transcriptional mechanism. To test this, we acquired the genes associated with “long” and “thin” phenotypes from WormBase [20] and compared these genes to the DEGs in our list. These phenotypes are generally found by RNAi knockdown of genes. We found that six and nineteen differentially expressed genes functioning in long or thin phenotypes in N2 and AB1, respectively, but only one gene, *his-67*, was common between the strains (Table 2.4). Thus, it is possible that the different gene expression profiles contribute to the body morphology variations between the strains.

Table 2.4: Differentially expressed long (purple), thin (pink), or long and thin (orange) phenotype genes under CeHR conditions in N2 and AB1 strains

N2		AB1	
Upregulated in CeHR versus OP50 NGM	Downregulated in CeHR versus OP50 NGM	Upregulated in CeHR versus OP50 NGM	Downregulated in CeHR versus OP50 NGM
adt-2	egg-6	cki-1	acs-1
C15A11.2	F56F11.4		his-67
his-3	frh-1		snrp-200
jph-1	his-62		spg-7
lon-8	his-67		Y105E8A.25
	rpb-8		
	vps-28		
	W05F2.3		
	Y95D11A.3		
	cyb-2.1		
	cyb-2.2		
	K08F4.3		
	kca-1		
	Y54G9A.5		

Neuropeptide signaling pathway genes were enriched in response to environmental changes. The pathway genes were upregulated in the liquid environments compared to agar in N2 and AB1. The differential expression of this pathway genes in liquid conditions did not maintain their expression patterns in the next generation. This finding supports our hypothesis that neuropeptide signaling pathway can be a crucial mechanism in rapid adaptive responses and it is conserved in the different strains of *C. elegans*.

Given the component differences in OP50 NGM, CeHR, and S-Medium, it is expected to observe variations in the expressions of metabolic process genes among

these conditions. We performed pathway enrichment analysis on DEGs to determine whether dietary sensor pathways, such as insulin/insulin-like growth factor (IGF) and target of rapamycin (TOR), are affected in the tested liquid cultivations. We have not observed enrichment for the renowned dietary sensor pathways but found other metabolic pathways to be affected. For instance, in CeHR, metabolic process in N2 and lysosome and fatty acid metabolism in AB1 demonstrated overrepresentation (Table 2.5). CeMM-grown animals have been reported not to show pathologies related to starvation (Szewczyk et al., 2006). Similarly, we have not detected differential expression for dietary restriction linked insulin-like signaling, *pha-4*/FoxA, *skn-1*/Nrf, or *nhr-49*/Hnf4 (Hibshman, Hung, & Baugh, 2016; Narasimhan et al., 2015). Together, these results seem to suggest that CeHR does not trigger a starvation response in the worms.

Table 2.5: Enriched pathways in F1 generations in response to the liquid cultivations. The green and red arrows show upregulation and downregulation compared to the expression in the corresponding P0 generations. The P-values are Benjamini-Hochberg corrected

		Pathway	Fold Enrichment	<i>P-value</i>
N2 CeHR (F <sub>1</sub> )	↑	Metabolic process	2.2%	7.8E-5
N2 CeHR (F <sub>1</sub> )	↓	DNA replication	1.6%	5.8E-8
		Proteasome	1.1%	2.6E-3
		Pyrimidine metabolism	1.3%	4.2E-3
		RNA degradation	1.1%	3.9E-3
AB1 CeHR (F <sub>1</sub> )	↑	Lysosome	1.1%	1.2E-3
		Fatty acid metabolism	0.8%	9.4E-3
N2 S-Medium (F <sub>1</sub> )	↑	TGF-beta signaling pathway	1.2%	4.1E-4
		Wnt signaling	1.2%	1.6E-3
		Protein processing in ER	1.4%	1.9E-3
		Ubiquitin mediated proteolysis	1.2%	2.6E-3

Particular environmental change-triggered DEGs in F<sub>1</sub> maintained their upregulation or downregulation in the next generation. The transmission of the environment-induced differential gene expressions to the F<sub>2</sub> generation from F<sub>1</sub> was between 20% - 34% (Figure 2.8a). To assess whether the transmitted differential gene expressions are orchestrated by the same transcriptional regulators (i.e., transcription factors), we first evaluated the promoter regions of these genes for motif enrichments

as potential recognition sites. The S-Medium experiment did not present any motif enrichment while the CeHR did for multiple sequence motifs in both the strains. The enriched motifs from both the upregulated and downregulated genes demonstrated differences between the strains, indicating that distinct transcriptional regulators can potentially be in play. We further compared the motifs from the upregulated genes against known transcription factor recognition sites (Bailey et al., 2009) to find if common transcription factors induce gene expressions in CeHR for the strains. One such motif from both the strains is a recognition site for transcription factor MDL-1 (Figure 2.9a). This transcription factor has regulatory functions in the inhibition of germline growth and longevity (Riesen et al., 2014). Other motifs from upregulated genes did not exhibit enrichment for transcription factor recognition. To test the human biological processes related to the putative MDL-1 recognition motifs from both the strains, we performed GO analysis for the DNA motifs against human promoters (Bailey et al., 2009). Both the motifs were overrepresented with the sensory perception of smell GO term, presumably corresponding to chemosensation functioning in the detection of environmental changes in worms. However, other enriched terms were different for these two motifs. For instance, the motif from the N2 strain had an overrepresentation of DNA damage checkpoint and transcription initiation from RNA polymerase II promoter while the motif from the AB1 strain was enriched for G-protein coupled receptor protein signaling pathway and positive regulation of immune response (Figure 2.9a).

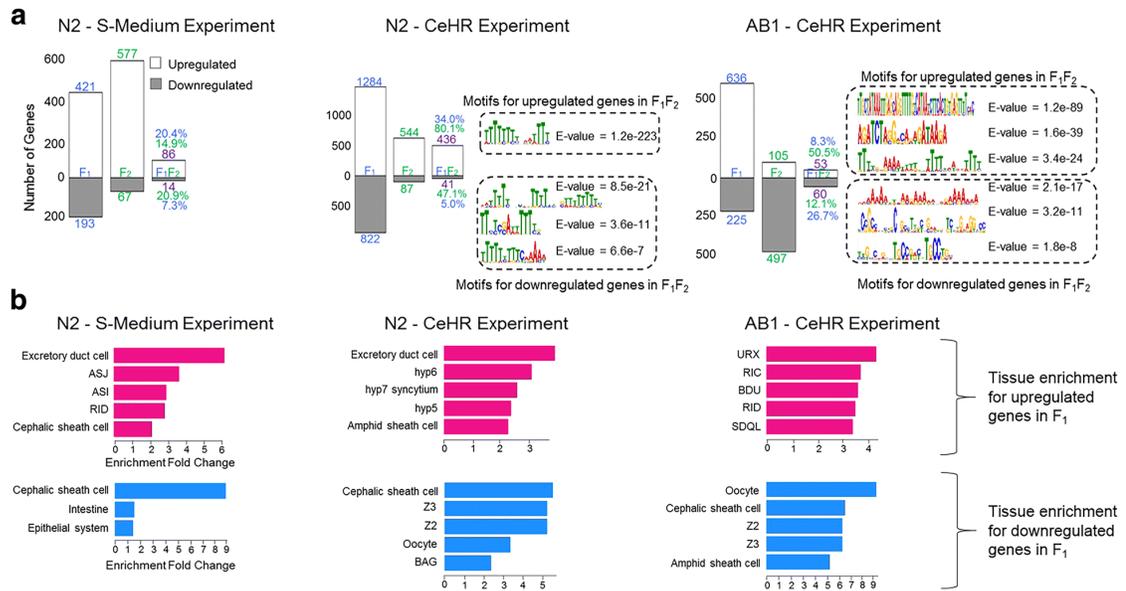


Figure 2.8: Different strains and the environment conditions contribute to differences in the maternal gene expression profiles. (a) The DEGs in the F<sub>1</sub>, F<sub>2</sub>, and both (F<sub>1</sub>F<sub>2</sub>) generations were identified in comparison to the gene expressions in the P<sub>0</sub> animals. Between 5 to 34% of the DEGs in the F<sub>1</sub> generations were transmitted to the F<sub>2</sub> generations (see the third category – F<sub>1</sub>F<sub>2</sub>) and these transmissions are higher than expected by chance (hypergeometric test, P-value < 0.001 for all the upregulated and downregulated genes individually). The gene expressions transmitted to the next generation after exposure to CeHR showed distinct enrichments for sequence motifs in their promoter regions for N2 and AB1 strains. The number of DEGs genes in F<sub>1</sub>, F<sub>2</sub>, and F<sub>1</sub>F<sub>2</sub> were represented with blue, green, and purple, respectively. (b) Top five tissue enrichment of the DEGs under exposure to environmental changes in F<sub>1</sub> (FDR-adj. P-value < 0.05).

We next sought to identify the transcriptional factor genes expressed at higher levels in the F<sub>1</sub> and F<sub>2</sub> generations. We considered the F<sub>1</sub>-induced transcription factor genes as the putative regulators of the environmental change responses and F<sub>2</sub>-induced ones as the putative regulators of re-adaptive responses. The comparison of our list of DEGs to the putative transcription factor genes (Reinke, Krause, & Okkema, 2013)

revealed that the majority of the differentially expressed transcription factors genes are strain and condition-specific (Figure 2.9b). None of the CeHR reversion-induced transcription factor genes were common between the strains indicating different transcriptional re-adaptive responses to agar plates for the N2 and AB1 strains. We identified only five shared CeHR-enriched transcription factor genes between the strains, and six shared genes between the CeHR and S-Medium conditions in N2 (hypergeometric test; *p-value* < 0.001). Neuron and muscle cell enriched R06C1.6 was commonly induced in CeHR for both the strains and in S-Medium for N2, suggesting a role for R06C1.6 in swimming behavior.

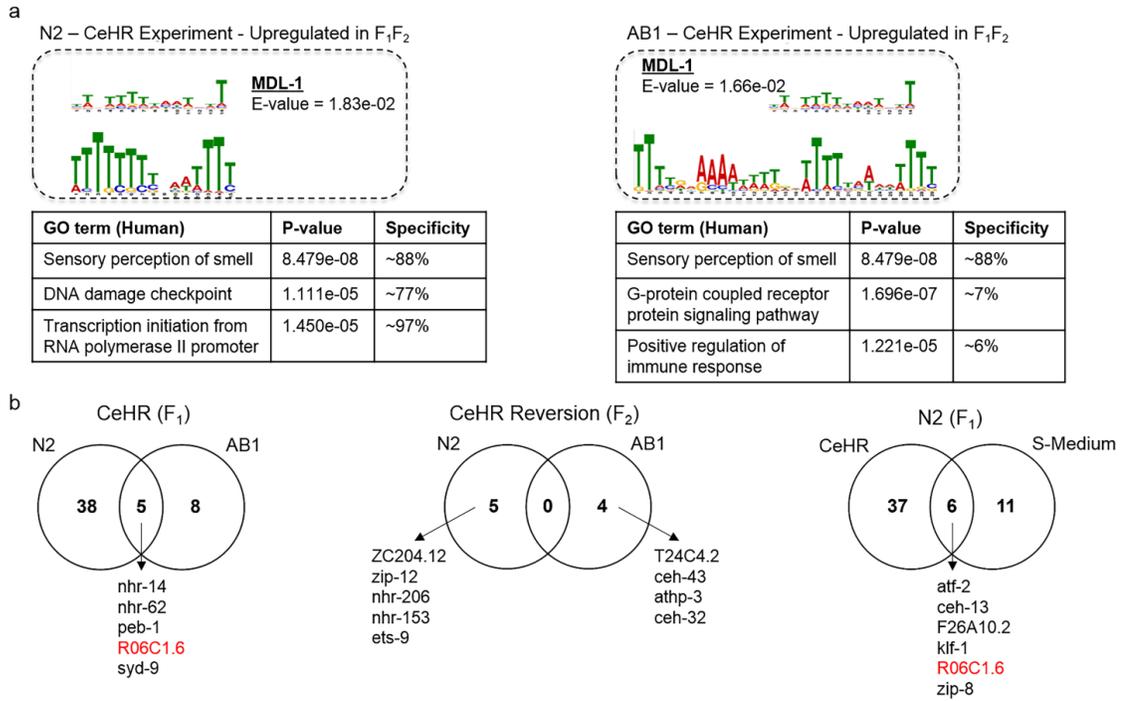


Figure 2.9: Putative transcription factors functioning in response to environmental changes in N2 and AB1. (a) One motif from each strain showed enrichment as a recognition site for transcription factor MDL-1. The same sequence motifs demonstrated differences for the overrepresented biological functions in human. (b) Venn diagrams depicting the overlap of the differentially expressed transcription factors between environmental conditions or strains. The overlap between the F<sub>1</sub> generations are statistically significant (Hypergeometric test,  $P$ -value < 0.0001)

To explore whether the DEGs among the F<sub>1</sub> generations show similar spatial expressions, we performed tissue enrichment analysis (Figure 2.8). We discarded any enrichments on embryonic cells as we conducted all our experiments on the L4 stage worms. The downregulated genes in CeHR demonstrated enrichment for reproductive system-related tissues such as the oocyte, Z2, and Z3 for both the strains. This

possibly reflects the changes in the reproduction of the worms in CeHR. However, upregulated genes in CeHR exhibited overrepresentation on different tissues between the strains: DEGs in N2 were mainly enriched for cylindrical hypodermal syncytium in head; DEGs in AB1 were enriched for neuronal tissues. The genes upregulated in response to S-Medium showed overrepresentation for neuronal tissues, but to different ones than the CeHR-exposed AB1 strain genes. Notably, downregulated genes in CeHR for both the strains and in S-Medium for N2 are enriched in cephalic sheath cell (Figure 2.8b). Altogether, these findings suggest that laboratory and wild-isolate strain worms show significant variations in the transcriptomic responses to the environmental alteration.

#### **2.3.4 Functional Profiles of the Differentially Expressed Genes Show Correlation with the Observed Phenotype**

We wondered if the functional data analysis results indeed point to phenotypical alterations in the animals. To test this, we employed microscopy techniques to detect the potential phenotypes. In the N2 strain, the genes functioning in reproduction and oogenesis have been downregulated in CeHR compared to agar and reversion. In AB1, the DEGs among the conditions did not show the same GO enrichment. This pattern is an indicator of phenotypical changes in the reproductive system that can be observed in CeHR but absent in reversion for N2. To test our prediction, we examined the germlines of N2 and AB1 adults in the different growth conditions via confocal microscopy (Figure 2.10 and Figure 2.11) (Çelen, Doh, &

Sabanayagam, 2018). CeHR grown N2 but not AB1 adults displayed aberrant germlines. Additionally, fewer oocytes were observed in CeHR-grown N2 strain compared to OP50 NGM-grown N2 strain. These findings suggest that the reduced fecundity in CeHR raised N2 animals may have resulted from the inability of animals to make gametes. In agreement with our hypothesis, the worms placed back to agar plates did not exhibit the similar phenotypical abnormalities with CeHR grown animals (Çelen et al., 2018).

Then, we asked if these germline abnormalities were caused by an increase in physiological germ cell apoptosis. We utilized a transgenic line of *C. elegans* that expresses functional CED-1::GFP fusion protein in sheath cells (Zhou, Hartwig, & Horvitz, 2001). CED-1 functions by initiating a signaling pathway in phagocytic cells that promotes cell corpse engulfment and is commonly used as a marker to detect cells undergoing the apoptotic pathway. In CeHR grown adults, CED-1::GFP was detected in a much higher proportion of germ cells (Figure 2.11). Why the number of apoptotic germ cells in CeHR grown N2 adults were increased and whether or not this was the sole cause of aberrant germlines will require further investigations.

For the N2 worms, another GO enrichment term was collagen and cuticulin-based cuticle development for upregulated genes in CeHR compared to agar and reversion generations. Hence, we reasoned that the cuticles might show morphological differences between agar and CeHR-grown N2 animals. In adult animals, cuticular structures were difficult to distinguish between N2 and AB1 via scanning electron

microscopy (SEM). More pronounced cuticular structures were detected in the earlier larval stages (L3) with SEM. We observed more protruding alae structures in CeHR-grown N2 larvae compared to AB1 larvae, but the annuli were indistinguishable between the two strains (Çelen et al., 2018). Comparing 42 genes annotated with alae morphology variant from WormBase to our data, we identified 15 of these genes (*dpy-2*, *dpy-3*, *dpy-5*, *dpy-7*, *dpy-8*, *dpy-9*, *dpy-10*, *dpy-18*, *lon-3*, *qua-1*, *rol-6*, *sqt-1*, *sqt-2*, *sqt-3*, and *mlt-10*) upregulated under CeHR conditions in N2 but not in AB1 (Figure 2.10 and Figure 2.11). Taken together, the correlation between the function of the DEGs and the observed phenotype suggests that these genes can potentially be a part of underlying genetic networks causing these phenotypes.

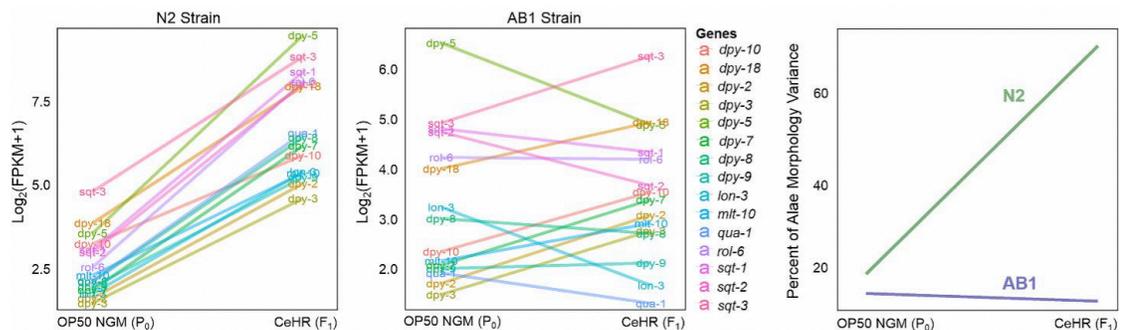


Figure 2.10: Norm of reaction between the alae morphology variant annotated genes and the detected alae morphology variance between AB1 and N2 indicating a potential role for these genes in the observed phenotype.

Previously, we demonstrated that the CeHR-grown animals were thinner and longer than the OP50 NGM raised animals (Doh et al., 2016). We asked if this change in body morphology was due to a reduction in fat stores that may have resulted from

the absence of a bacterial diet. N2 animals grown on OP50 NGM and CeHR were stained with Nile Red (Greenspan, Mayer, & Fowler, 1985) to visualize fat stores in all tissues of L3 and adult animals. We observed no significant difference in the number of fat stores between OP50 NGM and CeHR-grown N2 animals; however, larger fat stores were apparent in the abdominal and tail regions of OP50 NGM grown N2 adults (Figure 2.11 and (Çelen et al., 2018)). Thus, the lack of a bacterial diet apparently contributes to the morphological differences through the fat storage.

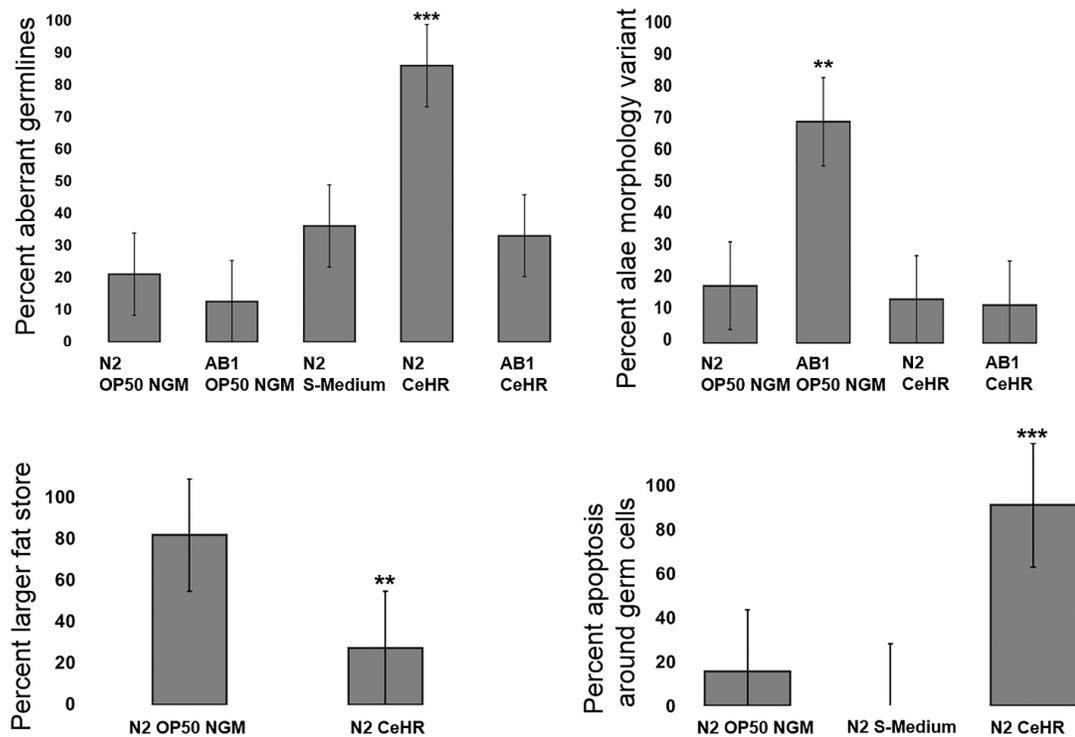


Figure 2.11: Quantitative differences in the observed phenotypes in liquid cultivations. T-test was used to compare the conditions. Error bars represent the standard error. \*\*: *P*-value < 0.01, \*\*\*: *P*-value < 0.0001. *P*-value > 0.05 is not reported.

## 2.4 Discussion

In this study, we demonstrated that environment changes cause significant alterations in the transcriptional responses in a *C. elegans* model. Different strains of the worms present unique transcriptional characteristics in the adaptive responses, and these characteristics show correlation with the observed phenotypes. For instance, we detected differential expression in the body morphology-related genes along with the phenotypical changes on the body morphology in CeHR-grown N2 strain, but not the AB1 strain. Furthermore, we found that different environments provoke the condition-specific expression of genes, and some of these genes lacked previous experimental EST or cDNA evidence.

Our results revealed that specific groups of genes are expressed only in particular environments. Some of these genes lacked transcript evidence previously (Figure 2.2e). Considerable effort has been devoted to annotating the genes in *C. elegans*, but the standard WormBase (Harris et al., 2014) annotation still contains thousands of “predicted genes” (about 8%) which lack direct experimental cDNA or EST evidence (Gerstein et al., 2010). The predicted genes in the *C. elegans* genome have been mainly identified by computational methods. Experimental detection of these genes was missing; mostly because they demonstrate poor expression, weak statistical signals or less conservation across species (Hillier et al., 2005). Others found that unused genetic information can remain in the genome for many generations (Collin & Cipriani, 2003). Reactivation of dormant genes has been used for the

treatment of particular disorders indicating that the silenced genes can have essential functions in the genome (H.-S. Huang et al., 2011). These studies provide evidence that dormant genes can be available on the genome throughout many generations and they can contain essential genetic information. Thus, we entertain the possibility that these predicted genes simply need a trigger to activate their expression, and this trigger may be elicited when animals are exposed to certain environmental stimuli.

A better understanding of the biological systems can be achieved through mining publicly available data sources (Çelen, Ross, Arighi, & Wu, 2015). We expanded our knowledge in the affected genetic systems by acquiring the ncRNA profiles from WormBase (Harris et al., 2014). Condition-specific expression of ncRNAs implies important regulatory functions for these molecules in adaptive responses. Earlier studies have identified various classes of ncRNAs and their roles in biological systems including RNA splicing, DNA replication, epigenetic regulation of gene expression, and X-chromosome inactivation (Jinneng Cao, 2014; Carthew & Sontheimer, 2009; Kaikkonen, Lam, & Glass, 2011; Kishore, 2006; A. T. Zhang et al., 2011). However, the majority of predicted ncRNAs have properties or functions that have not been identified yet. The list of unclassified ncRNAs, or “ncRNAs,” employed in this study was derived from the ‘7k-set’ that was generated by the modENCODE Consortium (Gerstein et al., 2010). The ‘7k-set’ was assembled via predictions based on conservation and RNA secondary structure, and therefore, functional genomic

studies of these “ncRNAs” are also still lacking. The co-enrichment of the ncRNA expressions with the coding ones suggest an interplay between these two.

In our RNA-seq experiments, we used poly(A) selection method but still observed numerous ncRNA molecules potentially because many eukaryotic ncRNAs are polyadenylated. For example, a poly(A) tail is part of the mature RNA for many long ncRNAs, i.e., *Xist* that mediates X-chromosome inactivation (Amaral & Mattick, 2008). We cannot rule out that the ncRNAs sampled in our study resulted from technical artifacts in the RNA-seq. Nevertheless, many ncRNAs had expression levels higher than the housekeeping gene *pmp-3*. These findings would seem to suggest that the sampling of these RNAs was not an artifact of RNA-seq. The inclusion of non-polyadenylated ncRNA molecules can bring more insights for understanding the ncRNA functions in adaptive responses.

Along with the phenotypical differences, there were considerable differences in RNA expressions between the N2 and AB1 strains which were cultured in the same growth conditions. N2 had been reported to exhibit innate immune response against pathogenic bacteria (Alper, McBride, Lackford, Freedman, & Schwartz, 2007; Battisti et al., 2016). However, previous studies have determined that the N2 and wild-isolate CB4856 strains display a high variation in gene expressions and many of these strain-specific variations are related to innate immunity genes (Capra, Skrovanek, & Kruglyak, 2008). Similarly, our GO analysis results showed enrichment in innate immune response for the genes expressed exclusively in the AB1 strain on OP50

NGM (Figure 2.5). These genes do not correspond to the previously reported divergent regions between the strains which might have caused the differences in gene expressions (Thompson et al., 2015). The AB1 animals are challenged more with pathogens in their natural habitat. The AB1-specific expression of innate immune response genes may indicate that the expression of these genes are more readily available in the AB1 strain than N2 to prepare the animals for a potential pathogen encounter. The results seem to suggest that *C. elegans* in the wild have the natural aptitude to survive not only on different food sources but also many other environmental variables including pathogens or the shifts between the solid and liquid settings. However, years of laboratory domestication of the N2 strain may have given rise to laboratory selections or genetic bottlenecks that may have hindered the ability of the N2 animals to acclimate to changing environments (Sterken et al., 2015). Alternatively, a potential presence of cis- and trans-eQTL between N2 and AB1 can be the underlying reason for the differences in the enrichment of innate response genes. Follow-up experiments are needed to determine the underlying reason for the expression differences in the innate immunity genes between N2 and the wild-isolates AB1 and CB4856.

For the majority of the *C. elegans* studies, the worms are cultured on bacteria-seeded agar plates instead of an axenic medium [13, 52]. However, a bacterial food source can present confounding effects on certain studies. For example, bacterial by-products can create genetic responses in the animals, and these responses are often

overlooked. Axenic media is desirable in space, biochemical, and toxicology studies as it enables automated culturing and experimentation, and it eliminates the potential contamination risks due to a bacterial diet (Adenle, Johnsen, & Szewczyk, 2009; Doh et al., 2016; Samuel, Sinclair, Pinter, & Hamza, 2014). Placing the animals in axenic media, however, have profound consequences altering a variety of biological processes. To distinguish whether the genetic responses are from the case study or the axenic media conditions, a separate investigation should be conducted. In light of this, we revealed that *C. elegans* acquire large variations in gene expressions upon single generation exposure to two types of liquid media, and these variations are more prominent for the N2 strain.

## **2.5 Conclusions**

Culturing the worms in axenic medium seems to affect the worms more than only changing their physical environment to liquid in the S-Medium. Our results highlight that caution must be used while studying model organisms as the constant laboratory environments can cause substantial differences in the transcriptomic and phenotypic responses to abrupt environmental changes. We believe our data can provide standard controls for future studies that utilize liquid cultivation of *C. elegans* for experimentations and is a rich resource for the discovery of genes showing environment-specific expression.

## **2.6 Methods**

### **2.6.1 *C. elegans* strains and growth conditions**

Wild-type N2, wild-isolate AB1, and MD701 bcIs39 [lim-7p::ced-1::GFP + lin-15(+)] strains were obtained from the Caenorhabditis Genetics Center (CGC). All animals were grown at 21°C. Embryos were isolated via the bleaching protocol in each step (Kenyon et al., 1988) and placed in the following growth media: *E. coli* OP50 seeded on NGM agar plates (OP50 NGM), 1 mL of S-Medium in 20 mL scintillation vials inoculated with a concentrated *E. coli* OP50 pellet made in 6 mL of an overnight culture (OP50 S-Medium), and CeHR medium in 20 mL scintillation vials according to Nass and Hamza (Nass & Hamza, 2007) with minor modifications (we used 50  $\mu$ M hemin instead of 20  $\mu$ M and used 30 $\mu$ M HEPES instead of 10 $\mu$ M).

### **2.6.2 RNA Isolation, Illumina Sequencing**

Total RNA from approximately 100,000 synchronized young adult *C. elegans* (4 hours post L4) were isolated with a modified TRIzol protocol and recovered by alcohol precipitation. Total RNA was further purified by PureLink™ RNA Mini Kit (Life Technologies). Two micrograms of total RNA were used for library preparations using the TruSeq Stranded mRNA LT Sample Prep Kit (Illumina). Libraries were sequenced on an Illumina HiSeq 2500 instrument set to the rapid run mode using single-end, 1 × 51 cycle sequencing reads as per manufacturer's instructions.

### 2.6.3 qRT-PCR

Quantitative RT-PCR was conducted using 1st strand cDNA synthesized from total RNA and gene-specific primers. Each cDNA sample was amplified using the SYBR® Premix Ex Taq™ II (Takara Bio) on the ABI 7500 Fast Real-time PCR System (Applied Biosystem) according to the manufacturers. The experiment was performed by three independent experiments with biological triplicates.

### 2.6.4 Bioinformatics Analyses

The Tuxedo pipeline (Trapnell et al., 2012) was used to find DEGs with default parameters. The reference genome (WBCel235) was obtained from Ensembl (Cunningham et al., 2014). Data from the study conducted by Dallaire et al. were retrieved from Gene Expression Omnibus (accession number GSE54173)(Dallaire et al., 2014) to compare the expression of the unconfirmed genes in OP50 NGM-grown N2, and the same pipeline with the same parameters were used for the analysis of the data. The same pipeline with same parameters were used to analyze RNA-seq data. We considered genes with FPKM > 1, FDR adjusted *p-values* < 0.05, and log<sub>2</sub> fold change > 2 as differentially expressed. miRNA, piRNA, and rRNA molecules were discarded from the analysis. The ncRNA molecules (WS250) and unconfirmed genes were acquired from WormBase through the ftp site and personal communications, respectively (Harris et al., 2014). The motif enrichment analysis was performed with MEME Suite (v4.11.1) (Bailey et al., 2009). The promoter sequences for the motif

enrichment were retrieved from UCSC Genome Browser database for cell (Rosenbloom et al., 2015). Find Individual Motif Occurrences (FIMO) was used to find the enrichment of the detected sequence motifs for the known transcription factor recognition sites (Grant, Bailey, & Noble, 2011). We utilized GOMO (v4.12.0) software to identify the enriched human GO terms for the promoter sequence motifs (Buske, Bodén, Bauer, & Bailey, 2010). GO and pathway enrichment analyses was made with the Database for Annotation, Visualization and Integrated Discovery (DAVID) (v6.8) (D. W. Huang, Sherman, & Lempicki, 2008), and Benjamini Hochberg corrected and FDR adj. *p-values* < 0.05 considered significant. Tissue and phenotype enrichment analyses were made by using WormBase Gene Set Enrichment Analysis tool (Angeles-Albores, N. Lee, Chan, & Sternberg, 2016), FDR adj. *p-values* < 0.05 considered significant. Long, thin, and alae morphology variant phenotype genes were retrieved from WormBase (WS262). R software was used for all statistical analyses and the plots (R Development Core Team, 2017).

### **2.6.5 Microscopy**

Nematodes were fixed in cold (-20°C) methanol for 15 minutes and stained with SYBR Gold (1:500,000 dilution in PBS) or DAPI (100 ng/mL) for 15 minutes, or Nile Red (5 ng/mL) for 30 minutes, at 21°C with gentle rocking. Worms were destained in PBST, washed three times with M9 and mounted on a 2% agarose pad for microscopy. All micrographs were taken with the Zeiss 710 laser scanning confocal

microscope with a 40×/1.2 N.A. water immersion objective. Dissected gonads were stained with SYBR gold as mentioned above. Nematodes were placed onto 0.22  $\mu$ M filters where the excess liquid was removed to visualize cuticular structures via cryo-SEM. Upon attaching the filter paper onto the sample holder, samples were plunged into liquid nitrogen slush. There a vacuum was pulled allowing sample transfer to the Gatan Alto 2500 cryo chamber at a temperature of -125°C. Samples were sublimated for 10 minutes at 90°C followed by cooling to -125°C. A thin layer of Gold Palladium was sputtered onto the samples. The samples were then transferred into a Hitachi S-4700 Field-Emission Scanning Electron Microscope for imaging. At least 20 animals were screened for each phenotype, and the represented figures were used for the observed phenotypes.

## Chapter 3

### INTERGENERATIONAL GENOME-WIDE CHROMATIN REMODELING IN *C. elegans* UNDER ENVIRONMENTAL CHANGES

#### 3.1 Abstract

##### Background

Chemical modifications to the tails of histone proteins act as gene regulators that play a pivotal role in adaptive responses to environmental stress. Determining the short and long term kinetics of histone marks is essential for understanding their functions in adaptation.

##### Results

We used *Caenorhabditis elegans* as a model organism to study the histone modification kinetics in response to environmental stress, taking advantage of their ability to live in both terrestrial and aquatic environments. We investigated the multigenerational genome-wide dynamics of five histone marks (H3K4me3, H3K27me3, H4K20me1, H3K36me1, and H3K9me3) by maintaining P<sub>0</sub> animals on terrestrial (agar plates), F<sub>1</sub> in aquatic cultures, and F<sub>2</sub> back on terrestrial environments. We determined the distributions of histone marks in the gene promoter regions and found that H4K20me1, H3K36me1, and H3K9me3 showed up to eleven-fold differences in density, whereas H3K4me3 and H3K27me3 remained highly constant during adaptation from terrestrial to aquatic environments. Furthermore, we predicted

that up to five combinations of histone marks can co-occupy single gene promoters and confirmed the colocalization of these histone marks by structured illumination microscopy. The co-occupancy increases with environment changes and different co-occupancy patterns contribute to variances in gene expressions and thereby presents a supporting evidence for the histone code hypothesis.

### **Conclusion**

Our study reveals that histone modifications show distinct intergenerational dynamics under environmental stress conditions.

### **3.2 Introduction**

Adaptation is essential for the proliferation of living organisms, and it involves changes in the genotype and phenotype in response to environmental stress. *Caenorhabditis elegans* is uniquely suited for multigenerational adaptation studies with its short life cycle and high homology to *Homo sapiens*. Their ability to live and reproduce in both terrestrial and aquatic cultures enables unique adaptation studies in response to transferring the animals between these environments. Additionally, these nematodes can be fed either bacteria or an axenic medium (such as CeMM; *C. elegans* Maintenance Medium or CeHR; *C. elegans* Habitation and Reproduction) allowing control of their food sources (Houthoofd et al., 2002a; Szewczyk et al., 2006). The phenotypical variations reported in *C. elegans* aquatic cultures, namely a decreased brood size, increased lifespan, and reduced fecundity, indicate that the changes in growth environment are drastic for the animals.

Environmental changes can cause differences in gene expression, and these changes can be transmitted to the next generations via an epigenetic memory (Daxinger & Whitelaw, 2010; Heard & Martienssen, 2014; Jaenisch & Bird, 2003). Histone modifications are involved in numerous biological processes, including longevity (Greer et al., 2011), X-chromosome repression (Gaydos et al., 2014), and proper germline programming (Rechtsteiner et al., 2010), and this involvement can affect multiple generations. For example, previous studies with *C. elegans* revealed that H3K4me3 on the longevity-related genes are retained between generations (Greer et al., 2011). The histone code hypothesis suggests that combinations of histone modifications play distinctive functions on the genome (Jenuwein & Allis, 2001; Strahl & Allis, 2000; Turner, 2002). To date, several bivalent histone marks such as H3K4me3 and H3K27me3 and their role in poising the developmental regulatory gene expression have been identified (Bernstein et al., 2006; Kelly, 2014; Matsumura et al., 2015; Voigt, Tee, & Reinberg, 2013). A critical question to answer is whether multivalence is required for adaptation to an environment change. Additionally, understanding the transgenerational epigenetics heavily depends on understanding the weight of contribution each mark has on the epigenetic memory. However, it is largely unknown whether all histone modifications show an epigenetic memory, and if so, whether there are identifiable differences in the epigenetic memory levels among the histone modifications.

To address the genome-wide role of epigenetics in the adaptation of *C. elegans* from terrestrial to aquatic environments, we cultured the P<sub>0</sub> worms on bacteria-seeded

agar plates and F<sub>1</sub> in liquid cultures. Placing F<sub>2</sub> nematodes back to agar plates enabled us to understand the memory levels of the epigenetic histone marks, that is to say, which marks are static throughout the environmental changes, and which are dynamic. For our studies, we selected H3K9me3 and H3K27me3 for their known enrichments on heterochromatic regions, and H3K4me3, H4K20me1, and H3K36me1 on euchromatic regions (Harr, Gonzalez-Sandoval, & Gasser, 2016). We identified that multiple histone modifications can occupy single genes by using bioinformatics analysis and fluorescent microscopy; co-localization of histone marks on same gene promoters indicates a specific function for histone modification pairs and supports the histone code hypothesis (Jenuwein & Allis, 2001).

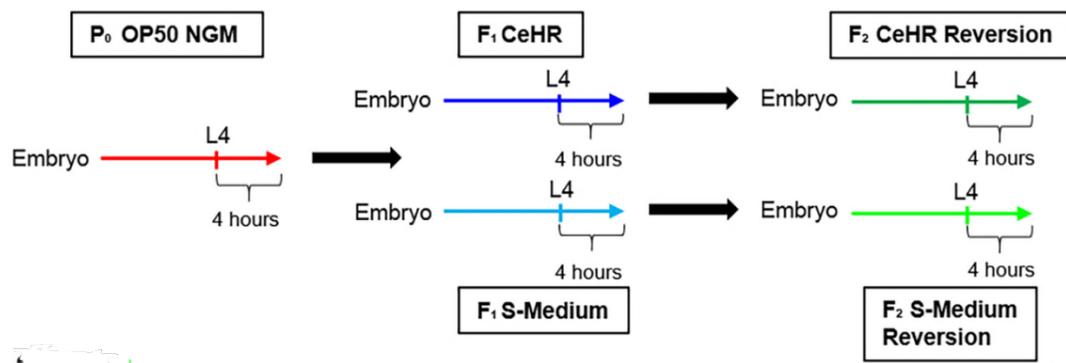


Figure 3.1: Experimental design for transgenerational inheritance studies.

### 3.3 Results

To elicit an adaptive response, the P<sub>0</sub> generation *C. elegans* was grown on bacteria-seeded agar plates (OP50 NGM), the F<sub>1</sub> generation was transferred to liquid,

and the F<sub>2</sub> generation reverted to agar (Figure 3.1). In addition to the changes in the physical environment, i.e., agar to liquid, we examined the effect of the dietary changes in the liquid cultures by using two different growth media: the commonly used bacterial S-Medium and more exotic axenic CeHR medium. For each generation, ChIP-seq was used to determine the distributions of the five histone modifications across the genome. In addition, RNA-seq was employed to identify the effect of histone modifications on gene expressions. The ChIP-seq and RNA-seq libraries were generated with two replicates and three replicates, respectively (Table 3.1 and (Çelen et al., 2017, 2018)).

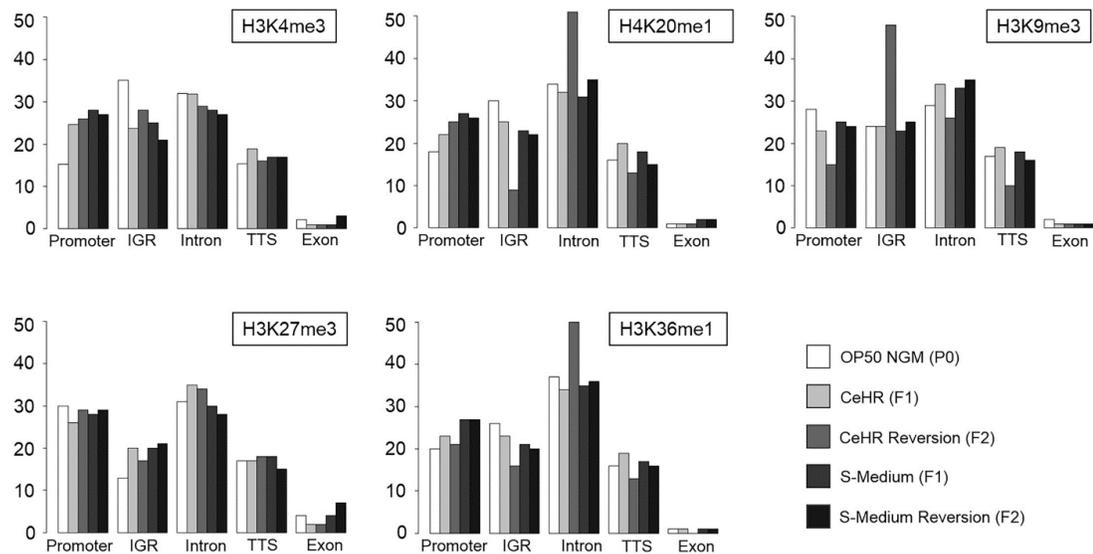


Figure 3.2: Distribution of the peaks on gene regions for each histone modification

Table 3.1: Sequencing read information for the ChIP-seq libraries

Strain	Condition	HM/RNAPII	Generation	Stage	Number of Reads	Mapped Reads
N2	OP50 NGM	RNAPII-1	P0	Young adult	33,242,761	32,432,359 (97.6%)
N2	OP50 NGM	RNAPII-2	P0	Young adult	21,405,422	21,061,305 (98.4%)
N2	OP50 NGM	H3K4me3-1	P0	Young adult	98,153,856	95,721,133 (97.5%)
N2	OP50 NGM	H3K4me3-2	P0	Young adult	61,233,901	60,687,627 (99.1%)
N2	OP50 NGM	H3K9me3-1	P0	Young adult	60,433,231	59,656,423 (98.7%)
N2	OP50 NGM	H3K9me3-2	P0	Young adult	59,589,393	59,002,510 (99.0%)
N2	OP50 NGM	H4K20me1-1	P0	Young adult	76,643,294	76,073,072 (99.3%)
N2	OP50 NGM	H4K20me1-2	P0	Young adult	79,606,462	78,883,730 (99.1%)
N2	OP50 NGM	H3K27me3-1	P0	Young adult	82,682,650	82,369,435 (99.6%)
N2	OP50 NGM	H3K27me3-2	P0	Young adult	110,403,541	109,947,892 (99.6%)
N2	OP50 NGM	H3K36me1-1	P0	Young adult	79,159,912	78,639,210 (99.3%)
N2	OP50 NGM	H3K36me1-2	P0	Young adult	58,089,120	57,636,022 (99.2%)
N2	CeHR	RNAPII-1	F1	Young adult	28,204,660	27,301,540 (96.8%)
N2	CeHR	RNAPII-2	F1	Young adult	8,996,507	8,427,692 (93.7%)
N2	CeHR	H3K4me3-1	F1	Young adult	48,199,119	47,292,679 (98.1%)
N2	CeHR	H3K4me3-2	F1	Young adult	30,961,169	30,458,115 (98.4%)
N2	CeHR	H3K9me3-1	F1	Young adult	63,115,408	62,722,736 (99.4%)
N2	CeHR	H3K9me3-2	F1	Young adult	49,227,188	48,810,858 (99.2%)
N2	CeHR	H4K20me1-1	F1	Young adult	53,778,753	53,301,487 (99.1%)
N2	CeHR	H4K20me1-2	F1	Young adult	54,430,956	53,501,313 (98.3%)
N2	CeHR	H3K27me3-1	F1	Young adult	50,461,677	50,226,861 (99.5%)

N2	CeHR	H3K27me3-2	F1	Young adult	66,905,312	66,629,883 (99.6%)
N2	CeHR	H3K36me1-1	F1	Young adult	47,012,677	46,524,531 (99.0%)
N2	CeHR	H3K36me1-2	F1	Young adult	37,267,937	36,899,320 (99.0%)
N2	CeHR Reversion	RNAPII-1	F2	Young adult	25,164,752	24,476,477 (97.3%)
N2	CeHR Reversion	RNAPII-2	F2	Young adult	29,371,272	27,691,788 (94.3%)
N2	CeHR Reversion	H3K4me3-1	F2	Young adult	29,496,065	27,475,896 (93.2%)
N2	CeHR Reversion	H3K4me3-2	F2	Young adult	61,076,111	59,272,158 (97.1%)
N2	CeHR Reversion	H3K9me3-1	F2	Young adult	51,003,367	50,320,889 (98.7%)
N2	CeHR Reversion	H3K9me3-2	F2	Young adult	74,611,326	73,214,653 (98.1%)
N2	CeHR Reversion	H4K20me1-1	F2	Young adult	82,043,157	80,944,392 (98.7%)
N2	CeHR Reversion	H4K20me1-2	F2	Young adult	65,356,382	64,371,586 (98.5%)
N2	CeHR Reversion	H3K27me3-1	F2	Young adult	71,104,419	70,411,001 (99.0%)
N2	CeHR Reversion	H3K27me3-2	F2	Young adult	93,138,324	92,326,314 (99.1%)
N2	CeHR Reversion	H3K36me1-1	F2	Young adult	66,677,646	65,206,682 (97.8%)
N2	CeHR Reversion	H3K36me1-2	F2	Young adult	63,557,710	62,459,790 (98.3%)
N2	S-Medium	RNAPII-1	F1	Young adult	32,655,058	20,891,346 (64.0%)
N2	S-Medium	RNAPII-2	F1	Young adult	87,157,853	79,027,131 (90.7%)
N2	S-Medium	H3K4me3-1	F1	Young adult	26,003,042	13,452,075 (51.7%)
N2	S-Medium	H3K4me3-2	F1	Young adult	20,867,794	11,337,002 (54.3%)
N2	S-Medium	H3K9me3-1	F1	Young adult	30,244,465	15,643,908 (51.7%)
N2	S-Medium	H3K9me3-2	F1	Young adult	33,614,653	21,946,615 (65.3%)
N2	S-Medium	H4K20me1-1	F1	Young adult	40,594,116	35,140,262 (86.6%)
N2	S-Medium	H4K20me1-2	F1	Young adult	85,775,129	83,623,413 (97.5%)

N2	S-Medium	H3K27me3-1	F1	Young adult	81,810,806	79,775,263 (97.5%)
N2	S-Medium	H3K27me3-2	F1	Young adult	99,513,914	98,025,137 (98.5%)
N2	S-Medium	H3K36me1-1	F1	Young adult	44,946,583	40,977,505 (91.2%)
N2	S-Medium	H3K36me1-2	F1	Young adult	22,664,268	17,987,641 (79.4%)
N2	S-Medium Reversion	RNAPII-1	F2	Young adult	44,242,786	34,090,863 (77.1%)
N2	S-Medium Reversion	RNAPII-2	F2	Young adult	103,442,596	92,892,350 (89.8%)
N2	S-Medium Reversion	H3K4me3-1	F2	Young adult	38,668,072	35,709,075 (92.4%)
N2	S-Medium Reversion	H3K4me3-2	F2	Young adult	47,545,322	44,428,038 (93.4%)
N2	S-Medium Reversion	H3K9me3-1	F2	Young adult	41,141,236	33,499,884 (81.4%)
N2	S-Medium Reversion	H3K9me3-2	F2	Young adult	27,341,693	13,722,282 (50.2%)
N2	S-Medium Reversion	H4K20me1-1	F2	Young adult	26,753,367	22,037,678 (82.4%)
N2	S-Medium Reversion	H4K20me1-2	F2	Young adult	28,267,956	26,322,836 (93.1%)
N2	S-Medium Reversion	H3K27me3-1	F2	Young adult	56,318,227	55,587,980 (98.7%)
N2	S-Medium Reversion	H3K27me3-2	F2	Young adult	90,015,297	88,741,917 (98.6%)
N2	S-Medium Reversion	H3K36me1-1	F2	Young adult	30,516,766	27,382,071 (89.7%)
N2	S-Medium Reversion	H3K36me1-2	F2	Young adult	34,696,645	30,412,566 (87.7%)

We found that all histone modifications studied displayed a higher density at the autosomal chromosome arms, as noted for H3K9Me3 in a previous study (Liu et al., 2011). A higher density of histone marks on the chromosomal arms can be the underlying reason for the greater variations in gene expression in these regions (Denver et al., 2005). The distribution of H3K4me3 is an exception as it is also enriched on the chromosome centers. Given that genes on the central regions show

higher expression compared to the ones on the chromosomal arms (Liu et al., 2011), this distribution pattern correlates with the notion that H3K4me3 is an active chromatin mark. The most notable difference between animals cultured in S-Medium and CeHR was the propensity for marks to remain constant over the environmental changes for the former whereas the latter showed loss of marks. This difference occurs presumably from the change in diet. Many H3K9me3, H4K20me1 and H3K36me1 marks were lost in the animals when reverted to OP50 NGM from CeHR. For all conditions, we found that the majority of five histone marks that were studied occur in the promoter, intergenic and intron genomic regions, while less than about 8% of the marks were associated with exons (Figure 3.2). Here, we present our findings on the role of histone marks in the gene promoter regions only in order to compare our epigenetic data with changes in the transcriptome. An analysis of histone marks associated with other genomic regions will be the subject of future publications.



the original environment, or it may be the case that there is a generation lag time before new marks are created due to the aquatic environment (i.e., the new marks are produced in the embryos of the  $F_1$  animals). We refer to the histone marks kept in all three generations as the “core epigenetic memory” because they remain unchanged during the environmental shifts. In contrast, new histone marks arise in aquatic cultures which presumably give the animals a competitive advantage in liquid; a subset of these aquatic marks disappear when the animals revert to terrestrial conditions, but some remain present even after returning to OP50 NGM forming an epigenetic memory of the aquatic environment. Common marks in  $P_0$  and  $F_2$  presumably regulate genes unique to terrestrial environments.

We considered the marks present only in S-Medium animals as being responsible for adaptation to aquatic conditions because they share the same bacterial food source as the  $P_0$  generation, whereas the marks solely observed in CeHR-grown worms function in adaptation to the axenic medium. We defined the histone marks with constant distributions as “epigenetically static”, and defined the histone marks with the dynamic changes in their distributions as “epigenetically elastic”. By distributions we mean the spatial placement of histone marks along the genome within each S-Medium and CeHR experiment (Figure 3.3b). Comparison of the core epigenetic memory marks between the S-Medium and CeHR experiments enabled the determination of the level of epigenetic memory and the general functions of distinct histone marks.

### 3.3.1 Defining a core epigenetic memory and its role in adaptation

We compared the number of genes in the core epigenetic memory of the CeHR and the S-Medium experiments to determine the static or elastic nature of each histone mark. For the CeHR and the S-Medium experiments, we grouped the genes associated with the five histone marks (H3K27me3, H3K4me3, H3K20me1, H3K36me1, and H3K9me3) and RNA Polymerase II (RNAPII) for a comparison (Figure 3.3c and Figure 3.4). We considered the similarity in the number of core epigenetic memory genes as an indicator of a static histone modification. To that end, we first used Chi-square with Yates correction to determine if the number of core epigenetic memory genes shows a significant difference in the two experiments. The results demonstrated that the number of these genes were not significantly different between the two experiments for H3K27me3 and RNAPII, attributed to their static nature. The other histone marks, however, showed a significant difference ( $P = 0.006$  for H3K4me3 and  $P < 0.001$  for the others) indicating their elasticity. Second, we examined the epigenetic memory levels by the ratio of the core epigenetic memory genes (RCEMG) between the CeHR and S-Medium experiments (Figure 3.3b). If RCEMG is less than two-fold, then the histone modification is considered to be epigenetically static because of their constant distributions. The histone modifications with RCEMG between two- to five-fold are moderately static, and the histone modifications with greater than five-fold RCEMG are defined as elastic. Using these criteria, we classify H3K27me3 and H3K4me3 as static (Figure 3.3c), H3K36me1 and H3K9me3 as moderately static, and H4K20me1 as elastic (Figure 3.4).

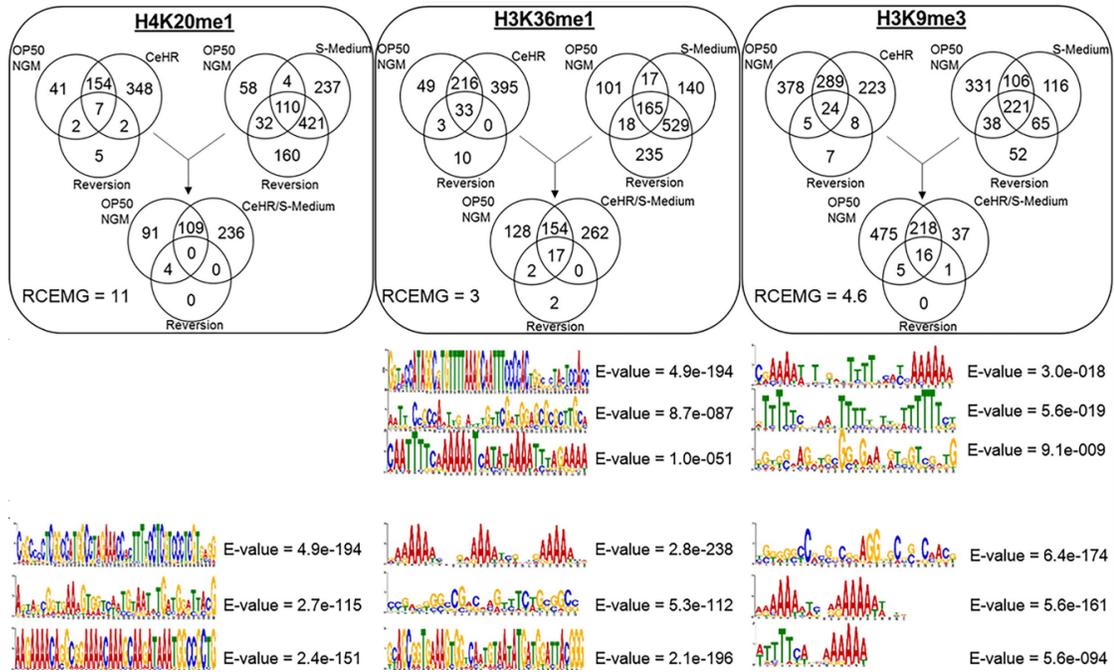


Figure 3.4: Number of genes associated with elastic histone marks individually or commonly in conditions, and the common genes found in both CeHR and S-medium experiments. Enriched sequence motifs for core epigenetic memory genes suggests that epigenetic memory occurs through recognition of sequence motifs.

We found that the number of mutual genes in the core memory of the two experiments was high for H3K27me3, H3K4me3, and RNAPII (42%, 68%, and 33%, respectively), indicating their preferential and consistent association with gene promoters. However, the number of mutual core memory genes between the CeHR and S-Medium experiments was much lower for H3K9me3 (7%), H3K36me1 (9%), and H4K20me1 (0%), suggesting a more dynamic nature in response to an environment change for these histone modifications. These results are congruent with

our hypothesis that H3K27me3, H3K4me3, and RNAPII are static whereas the others studied are elastic at the same developmental stage in the three generations.

Strikingly, RNAPII showed similar kinetics with H3K27me3 by having a static RCEMG. A rapid transcriptional reactivation of gene expressions can be promoted by transcriptional memory which is conserved in eukaryotes (D'Urso & Brickner, 2014; Ferraro et al., 2016). Epigenetic transcriptional memory generally requires a physical interaction with the Nuclear Pore Complex (NPC) and an altered chromatin structure of the promoter. These alterations enable the binding of poised RNAPII to the promoter and the maintenance of this binding for several generations (D'Urso & Brickner, 2014; Light & Brickner, 2013). Therefore, our finding supports the notion that RNAPII has a static transcriptional memory which enables a faster transcriptional reactivation.

All the histone modifications studied appear to preserve their distribution on particular hotspots along the chromosomes. For example, we and others observed that H3K9me3 is highly enriched on the arm of Chromosome V indicating specific functions on these hotspots (Liu et al., 2011). We found that the core epigenetic memory genes have significantly enriched sequence motifs for the histone modifications studied, and the conservation of these motifs between the experiments is an indicator of a static epigenetic memory. For the static mark H3K4me3, multiple sequence motifs were enriched, but only one of them was conserved between the experiments (Figure 3.3c). Predictably, static RNAPII did not demonstrate conservation for the sequence motifs between the experiments as it is a key catalyzer

for the whole transcription process, and gene expression profiles are different in the two experiments. The core epigenetic memory genes of H3K27me3 showed enrichment for only one A-block motif in both the S-Medium and CeHR experiments (Figure 3.3c). This enrichment suggests that the A-blocks are utilized to maintain the epigenetic memory in H3K27me3. The difference in the motif conservation between the static marks makes the unique static behavior of H3K27me3 more prominent. Moderately static and elastic histone modifications had distinct sequence motifs enriched between the experiments (Figure 3.4) implying higher dynamic and adaptive responses. Our results indicate that sequence motifs are essential for the maintenance of the core epigenetic memory, and the static histone modifications are guided by conserved sequence motifs to maintain a constant distribution.

To investigate the impact of the histone modifications on intergenerational gene expression, we created a hierarchical cluster in different conditions with the expression of 1014 genes associated with them. The results showed that the histone modifications and RNAPII are clustered with the other histone marks from the same environmental conditions instead of clustering together in different environments (Figure 3.5a). This pattern displays the highly dynamic nature of histone modifications under exposure to different conditions. One particular exception to this is H3K27me3 since it mainly clustered together even in distinct environments thereby confirming that H3K27me3 has a strong static epigenetic memory. Interestingly, H3K4me3 did not show the same pattern as H3K27me3, even though both were previously defined to be static marks. Thus, our results suggest additional factors, such as multivalence

(discussed later), affect the expression levels of H3K4me3-associated genes. We also identified that H3K9me3, H4K20me1, and H3K36me1 are generally grouped together indicating that these histone modifications function together or that they mark common genes.

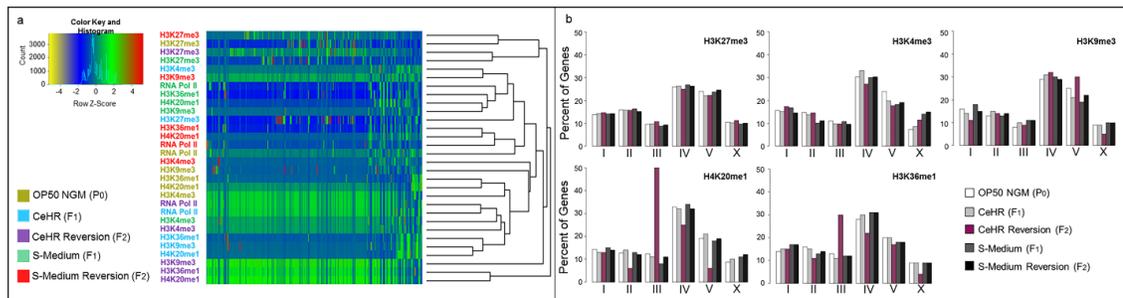


Figure 3.5. Evaluation of the static/dynamic nature of histone marks. (a) Hierarchical clustering of the histone modifications based on the expressions of the genes that they associate with. Static H3K27me3 marked genes tend to cluster together. (b) Distribution of histone marks on chromosomes under distinct environmental conditions. Static histone modifications show a constant distribution on the chromosomes over generations while the elastic ones demonstrate variances.

Next, we examined the distribution of histone modifications on the chromosomes under changing environmental conditions. While the distribution of the elastic marks showed variability, the static ones stayed constant on the chromosomes in response to environment changes (Figure 3.5b). For example, half of the Chromosome III genes were marked by H4K20me1 in CeHR reversion and only 8% in S-Medium. These findings support our hypothesis by suggesting that when the static histone modifications are highly stable and they potentially play a role in

housekeeping activities, elastic histone modifications are dynamic and they function in rapid adaptation.

### **3.3.2 Epigenetic memory-forming abilities in response to environment change**

In addition to the core epigenetic memory, histone modifications can mark a new set of genes in response to an environment change and maintain their association with those genes in subsequent generations. We define this type of epigenetic memory as “epigenetic memory forming”. We compared the number of genes associated with the histone modifications in both  $F_1$  and  $F_2$  generations in order to determine their memory forming abilities (Figure 3.3c) and found that under exposure to distinct conditions, individual histone marks were variable. The number of common genes in  $F_1$  and  $F_2$  showed a significant difference in H3K9me3, H3K36me1, H4K20me1, and RNAPII (Chi-squared test with Yates correction; P-value < 0.001). On the other hand, H3K4me3 and H3K27me3 had a constant epigenetic memory forming abilities for the two experiments (P-value > 0.01). H3K4me3 and H3K27me3 maintained a constant (23% - 18% and 5% - 6%), H3K9me3 showed a moderately variable (1% - 7%), and H3K36me1 and H4K20me1 presented highly variable (0% - 44% and 0% - 41%) epigenetic memory forming abilities.

Intergenerational epigenetic inheritance of several histone modifications, such as H3K27me3 (Gaydos et al., 2014) and H3K4me3 (Greer et al., 2011), have been revealed in earlier studies. Supporting these findings, we observed that H3K27me3 and H3K4me3 have a capability of forming and maintaining epigenetic memories.

Furthermore, we found that these histone modifications have a difference in the memory forming abilities when exposed to an environment change. For example, H3K4me3 forms higher levels of epigenetic memory than H3K27me3 in response to changes in the environment (Figure 3.3c). Nevertheless, H3K27me3 maintains its distribution on the chromosomes regardless of the environment. In S-Medium, elastic epigenetic memory forming marks H3K36me1 and H3K20me1 preserve their intergenerational distribution while in CeHR, they show either no or low mutual distribution between generations. These results would seem to suggest that H4K20me1 and H3K36me1 play a role in utilizing a bacterial food source. In addition, we note that worms in S-Medium can only be maintained for one generation, presumably due to the toxins released by the bacteria that remain in solution, and H4K20me1 and H3K36me1 could function in response to these toxins. Altogether, these findings indicate that all the histone modifications have the ability to form epigenetic memory in varying levels, but these levels depend upon the associated genes' role in the cell.

### **3.3.3 Multivalence of histone marks changes during adaptation**

Studies have revealed that bivalent histone marks, such as H3K4me3 and H3K27me3, play a role in poising the expression of developmental regulatory genes (Matsumura et al., 2015; Vastenhouw et al., 2010; Voigt et al., 2013). To investigate the potential co-regulation of the genes by distinct histone marks, we examined the occupancy of the histone modifications on single gene promoters for different

environment conditions (Figure 3.6). Intriguingly, we observed that many gene promoters are associated with multiple histone marks in varying environments (Figure 3.7a). Moreover, multivalence increases with an environment change. For example, in the  $F_1$  generations, about half of the gene promoters are occupied by multiple histone marks; however, single histone marks are predominantly observed in  $P_0$  animals. Likewise, others have found that bivalence between H3K4me3 and H3K27me3 increase as a stress response (Adriaens et al., 2016). For the  $F_2$  CeHR reversion animals, occupancy of histone marks resembles the  $P_0$  generation, but with S-Medium reversion animals, histone marks show high co-occupancy on single gene promoters. H4K20me1 and H3K36me1 represented different distribution dynamics in the two experiments. Their distributions are retained by  $F_2$  from S-Medium whereas these histone marks are rearranged by  $F_2$  from axenic CeHR cultures, implying a role for these two marks specifically in adaptation to S-Medium, presumably in combination with the other histone marks.

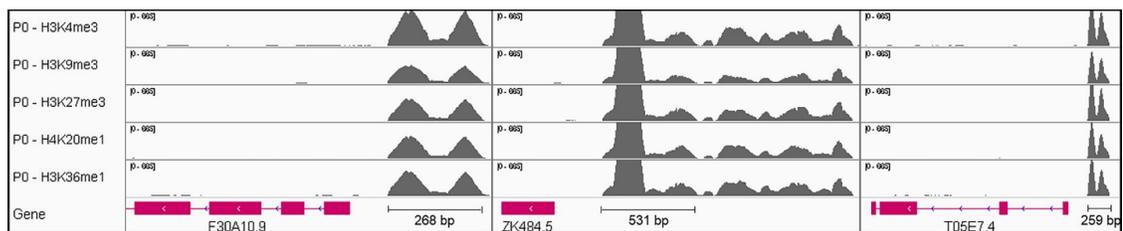


Figure 3.6: Genome browser view of the three pentavalently marked gene promoters. The same data range (0 – 665) was used for all the presented data.

Multivalence was lower in the core epigenetic memory compared to epigenetic memory-forming in both experiments. This result indicates that increased multivalence in response to environment change is maintained in the epigenetic memory (Figure 3.8). The S-Medium experiment showed higher multivalence in the epigenetic memory. The findings reported here support the assumption that S-Medium is a harsh condition for the worms and that cooperation of histone marks are vital for extreme stress conditions.

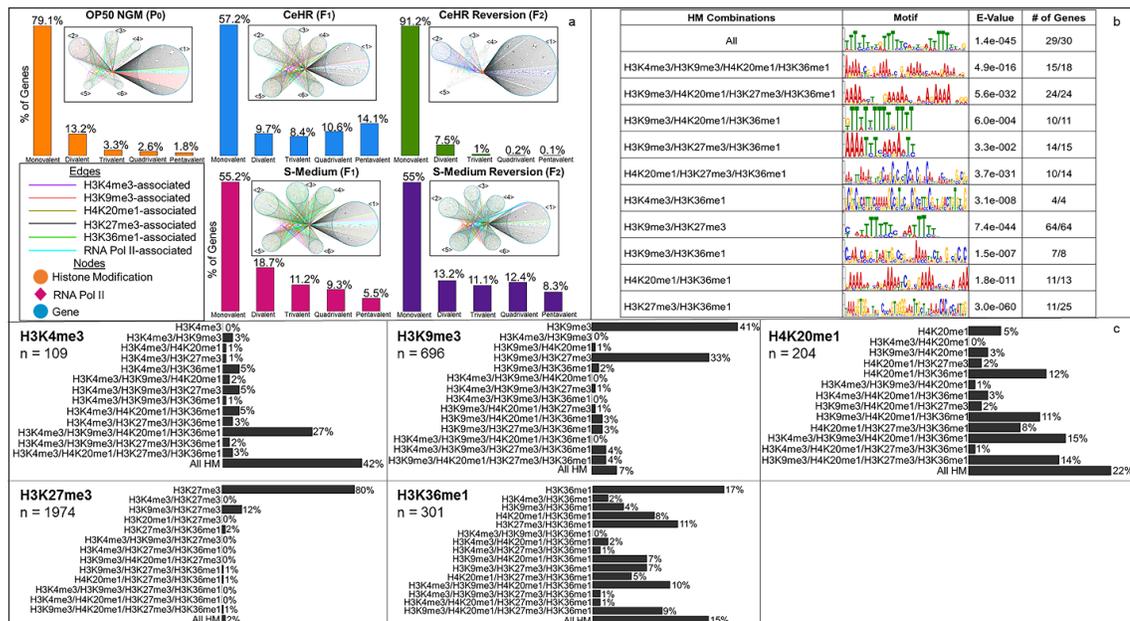


Figure 3.7: Multivalent nature of the individual histone marks. (a) Gene-histone mark networks for the three generations under different environmental conditions and the percentage of the gene promoters co-occupied by one to five histone marks (x-axis represents the number of marks). (b) Motif enrichment for the genes associated with multiple histone marks in P<sub>0</sub> generation indicates a sequence-specific recognition of the multivalently marked gene promoters. (c) Percentage of the multivalence or monovalence levels of the individual histone marks in P<sub>0</sub> generation.

To determine whether multivalent marks have a preference towards specific sequences, we analyzed the promoter sequences of the multivalently-marked genes for motif enrichment in P<sub>0</sub> animals. Our results showed that nine of the histone modification combinations have significantly enriched sequence motifs (Figure 3.7b). Taken together, these results indicate that for the proper expression of the genes, histone methylations can co-occupy promoters of single genes or that different cell types are regulated by distinct marks. RNAPII demonstrates highly enriched but distinct sequence motifs in the core epigenetic memory for both the experiments (Figure 3.3c). Such motifs indicate the existence of a histone code regulating the multigenerational gene expression. However, we have not observed these motifs enriched for the limited multivalent marks that we studied. Additional histone modifications are apparently involved in the multigenerational histone code to ensure the expression of the genes with the particular sequence enrichment on the promoter region.

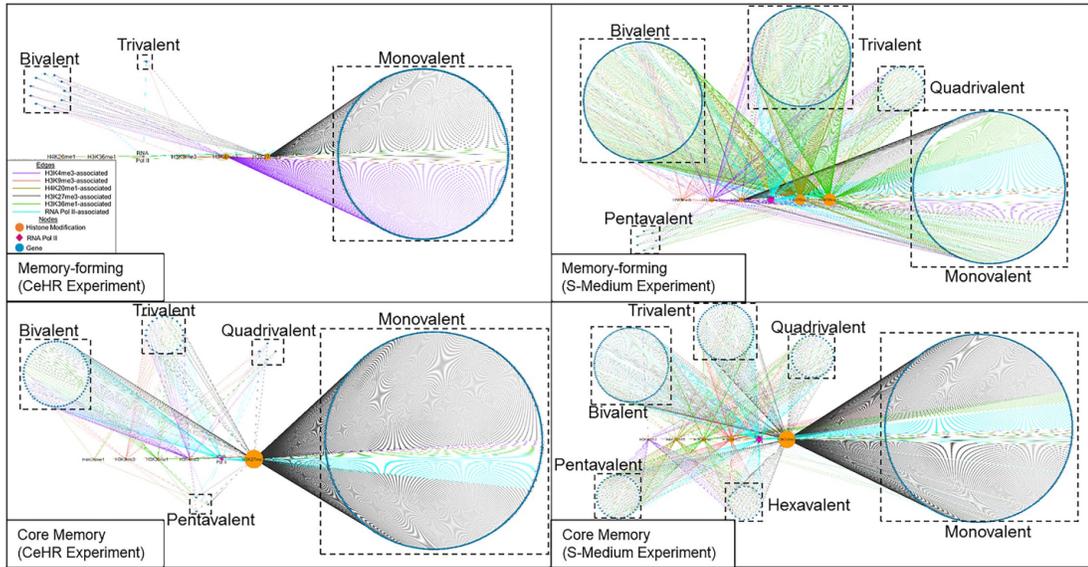


Figure 3.8: Histone modification associated gene promoters showing multivalence and monovalence in core epigenetic memory and epigenetic memory forming. Multivalence for core epigenetic memory is 5.9% and 20% for the CeHR experiment and the S-Medium experiment, respectively. Multivalence in the epigenetic memory forming is 12.4% and 47.8% for the CeHR experiment and the S-Medium experiment, respectively.

The hierarchical cluster of the histone-marked genes revealed that different histone modifications group with the others from the same environment (Figure 3.5a). One possible reason for this is the multivalence of the histone marks under different conditions. A high number of common genes associated with multiple histone marks can cause the histone marks to cluster together. For instance, H3K9me3, H4K20me1, and H3K36me1 were grouped together (Figure 3.5a) and were also found highly multivalent with each other (Figure 3.7c). Additionally, the existence of an enriched sequence motif for these three histone modifications strongly suggests that they form a distinct histone code. Another bivalent mark we found in high levels is H3K9me3 and

H3K27me3 as previously reported (Boros, Arnoult, Stroobant, Collet, & Decottignies, 2014). Even though the best studied bivalent histone modifications are H3K27me3 and H3K4me3, we found that H3K27me3 showed even more multivalence with H3K9me3. Besides having a strong static epigenetic memory, the dominance of monovalent behavior (Figure 3.7c) is potentially the main reason why H3K27me3 groups with itself under distinct environmental conditions. Similarly, due to its multivalent nature, static H3K4me3 marks do not cluster together in different conditions. Altogether, our results suggest that histone modifications can co-occupy single gene promoters through sequence motifs and thereby regulate gene expression in a specific fashion.

### **3.3.4 Rapid histone modification dynamics during embryogenesis**

Co-occupancy of the histone marks can be due to different spatial distributions of histone marks for different cell types since ChIP-seq analyses are performed on whole animals. To investigate that the histone marks indeed have a multivalent nature, we employed immunofluorescence followed by structured illumination microscopy (IF-SIM) for the high-resolution (~125 nm) localization in *P. C. elegans* embryos. Tracking the histone marks on different embryonic cell stages also enabled us to detect the short-term dynamics of the histone marks.

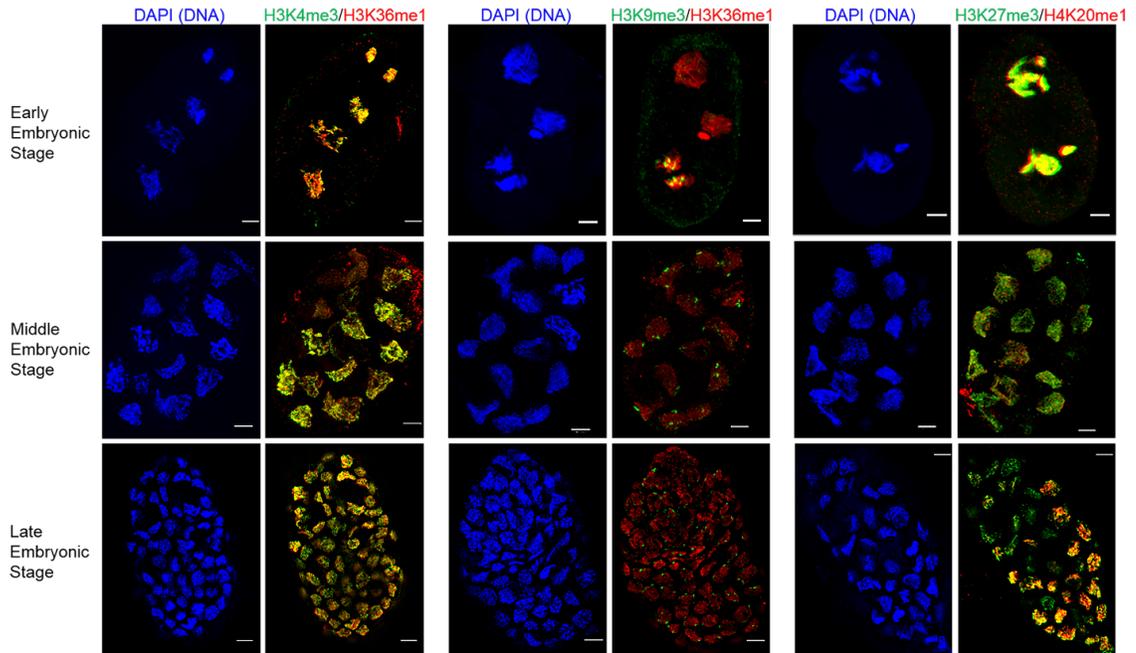


Figure 3.9: Structured illumination microscopy of fluorescently-stained histone marks. The figure represents the short-term dynamics of the histone marks during embryogenesis and co-localized marks for  $P_0$  animals. Scale bars:  $3\mu\text{m}$ .

In earlier stages of embryogenesis, the majority of the histone marks were present in all cells. By the later stages ( $> 40$ -cell), however, the concentration of particular histone marks varies across the cells. For example, H3K27me3, H3K4me3, and H3K36me1 have a broad distribution in all cells throughout embryogenesis. However, we and others (Vielle et al., 2012) observed that H4K20me1 levels are selectively higher in specific cells after 16-cell stage (Figure 3.9). H3K9me3 exhibits a unique distribution by marking mainly the nuclear periphery as shown in previous studies (Towbin et al., 2012). Moreover, unlike the other histone marks, H3K9me3 is observed only in specific cells during the early embryogenesis. These findings indicate

that the static histone marks are required for all cells, yet elastic histone marks can function in either all cells or selective cells.

We visualized H3K9me3/H3K36me1, H4K20me1/H3K27me3, and H3K4me3/H3K36me1 together to determine their co-occurrence experimentally. The signals from H3K4me3 and H3K36me1 were highly co-localized at the earlier stages of embryogenesis. The co-occupancy levels for these two histone marks are slightly decreased at the later stages of embryogenesis as the individual histone marks started to become more prominent in selective cells. Similarly, we observed that 88% of the H3K4me3 marks are co-located with H3K36me1 on promoter regions in young adult animals (Figure 3.7c). In earlier periods of embryogenesis, H4K20me1/H3K27me3 is highly co-localized. In later embryonic phases, however, H4K20me1 is observed only in specific cells, and H4K20me1/H3K27me3 co-localization is limited to those cells (Figure 3.9). Our bioinformatics analysis predicted that 49% of all the H4K20me1 marks coexist with H3K27me3 marks when around 6% of the H3K27me3 marks coexist with H4K20me1 in P<sub>0</sub> (Figure 3.7c). Correlated with this prediction, we have observed high levels of co-localization for H4K20me1 with H3K27me3 (Figure 3.9). On the promoter regions, around 23% of the H3K9me3 marks co-occupy same genes with H3K36me1. However, approximately 53% of the H3K36me1 marks on the promoters also have an H3K9me3 mark (Figure 3.7c). Based on the IF-SIM, H3K36me1 is broadly distributed in all cells (unlike H3K9me3), so it is expected to see a lower percentage of H3K36me1 and H3K9me3 co-localized marks. Our results for IF-SIM were consistent with the bioinformatics predictions, and they suggest that multiple

histone modification combinations potentially work together to accomplish a distinct function.

### **3.3.5 Impact of histone modifications on gene expression**

Histone modifications are known to function as transcriptional activators and repressors. We reasoned that if histone modifications play a particular role in transcriptional regulation, then histone-marked genes on different chromosomes should show similar expression profiles regardless of their genomic positions. We tested this hypothesis for OP50 NGM and CeHR conditions (Figure 3.10a). On OP50 NGM, except for H3K4me3 and H4K20me1, histone-marked genes showed differences in their expression levels, while in CeHR all histone-marked genes showed significant differential gene expression across chromosomes. As introducing a new environment increases histone multivalence, we investigated whether single histone-marked gene promoters have similar patterns. The singly-marked genes did not present significant differential gene expressions across chromosomes with the exception of H3K9me3 in CeHR, and H3K27me3 in both CeHR and OP50 NGM. These results indicate that the majority of individual histone modifications contribute to gene expressions in a specific fashion and when they work collaboratively they cause distinct patterns in gene expressions.

Genes associated with activator or repressor histone modifications did not always represent the assigned function as previously reported (Howe, Fischl, Murray, & Mellor, 2017; Pérez-Lluch et al., 2015). This observation can be due to the effect of

multivalent histone marks or can be the result of the distinct temporal behavior of histone modifications. To identify how histone modifications on promoter regions affect gene expression, we used k-means clustering (Figure 3.10b). We found that most of the histone modifications show two groups of expression patterns: one group is related to gene expression, while the other one is related to no or low gene expression, implying selective activator and repressor activities for histone modifications. In response to the dietary changes, the animals exhibit phenotypical changes, such as reduced number of progeny and altered body morphology (Doh et al., 2016). Such variations require changes in expression of the genes that can be passed to the subsequent generations. Therefore, we reasoned that the profiles of the histone modification marked genes may enable us to understand the functions of these individual epigenetic mechanisms during adaptation. Without the clustering, histone-marked genes did not have statistically significant gene ontology (GO) terms assigned. The histone-marked gene clusters, on the other hand, demonstrated enriched GO terms thereby indicating a potential role for histone modifications in function-specific gene regulation. Many of the identified functions for the low expressed gene clusters were in accordance with the observed phenotype as we reported in another paper (Çelen et al., 2017, 2018). For example, many of these gene clusters were enriched for body morphogenesis and gonad development, and correspondingly the worms demonstrated phenotypical alterations on the cuticle and the reproductive system.

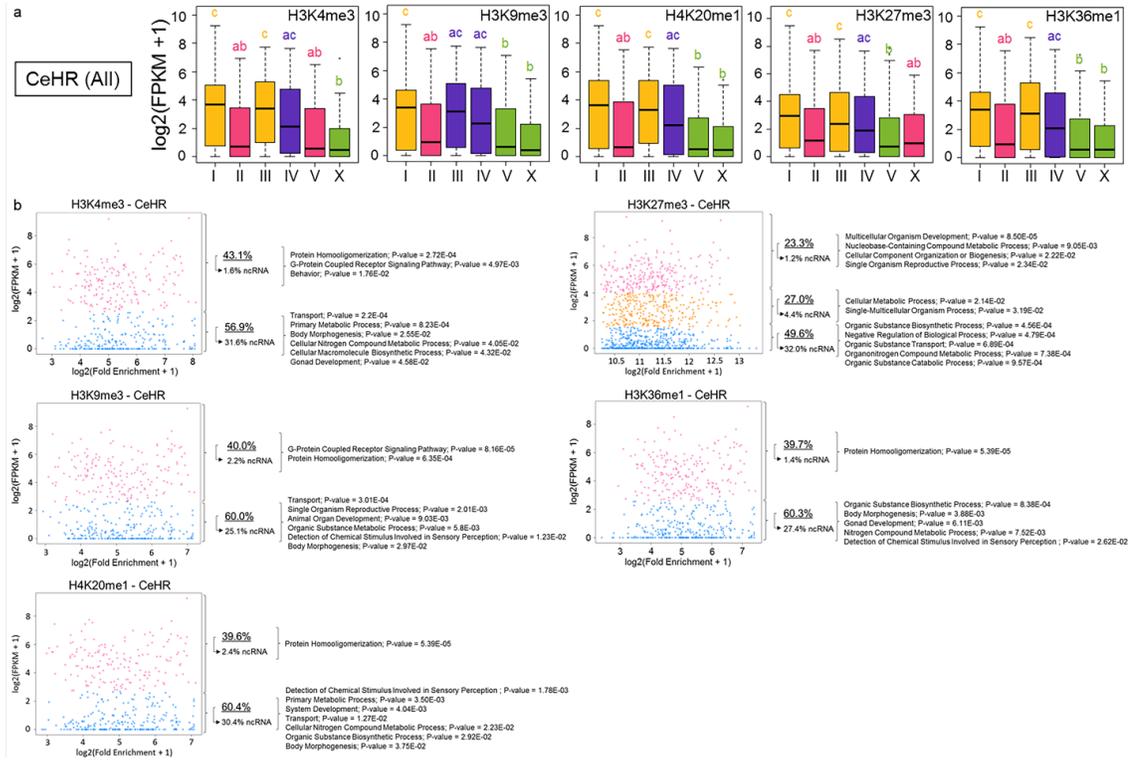


Figure 3.10: Histone modification multivalency causes changes in the gene expressions. (a) Tukey-Kramer test for the gene expressions among chromosomes for histone modification associated genes. The represented gene expressions are from both monovalently and multivalently marked genes. Means with the same letter are not significantly different. (b) K-means clustering for the histone modification associated gene expressions and the enriched gene ontology terms assigned to each cluster. Each color represents a separate cluster.

### 3.4 Discussion

In this study, we show that individual histone marks have static or elastic epigenetic memory which is directed by sequence motifs on the gene promoters. We have identified that five histone modifications demonstrate multivalency which increases with environmental changes and causes a difference in gene expression. Co-

localized histone marks and their static and elastic behaviors were also observed by using IF-SIM.

Histone modifications are typically categorized as transcription activators or repressors. In contrast, in our study single histone modifications did not demonstrate an absolute activator or repressor function (Figure 3.10b). Recent studies have shown that loss of H3K4me3 does not have a significant effect on transcriptional activity (Howe et al., 2017) and called the activator role of H3K4me3 into question. In another recent study, it was reported that “repressor” H3K9me3 plays a very limited role in transcriptional silencing and germline chromatin silencing and is not required for heritable transcriptional silencing in *C. elegans* (Kalinava, Ni, Peterman, Chen, & Gu, 2017). Similarly, others have reported repressor marks on active genes and activator marks on silenced genes. For example, H3K9me3 was also found to be enriched in active genes (Brinkman et al., 2006). One possible explanation for these expression patterns is that histone modifications do not work as simple on-off switches but rather dimmers that can regulate the amount of gene products. We observed that histone modifications can be associated with a continuum of low to highly expressed genes. Moreover, a functional enrichment could be achieved by clustering these genes based on the expression levels (Figure 3.10b). Distinct gene expression profiles observed on the multivalently marked promoters indicate that gene expressions are adjusted by the recruitment of specific histone modifications. Enrichment of sequence motifs on the multivalently marked promoters is also in agreement with the hypothesis that the

adjustment of gene expression requires the collaborative work of histone modifications.

The function of H3K36me1 and H4K20me1 have not been fully elucidated, yet it has been proposed that H4K20me1 plays a role in genome integrity since genomic instability is observed when its histone methyltransferase is disrupted (Jørgensen, Schotta, & Sørensen, 2013). Our analyses have revealed that H3K36me1 and H4K20me1 have moderately static and elastic epigenetic memory levels, respectively. These two histone marks had high epigenetic memory forming abilities in the S-Medium experiment, but not in the CeHR experiment. Presumably, these two marks have critical functions in the S-Medium experiment, but not in the CeHR experiments. We also observed that H4K20me1 can undergo a rapid redistribution in distinct cells during the embryogenesis. These results indicate that the two histone modifications are more likely to act as quick adaptive switches.

We found that sequence motifs in the core epigenetic memory play a role in maintaining the epigenetic memory presumably by working as a recognition sites for specific histone machineries. The histone modifications with a static epigenetic memory showed conservation for the sequence motifs. Specifically, H3K27me3 marked promoters were highly enriched for a single motif. Polycomb repressive complex-2 (PRC2) is responsible for the trimethylation of histone H3K27. It has been shown that only PRC2-generated H3K27 methylation provides a long-term epigenetic memory (Gaydos et al., 2014). The targets of PRC2 proteins were found to be enriched for CCG-repeats and GA-rich regions (Peng et al., 2009). However, the A-

blocks that we observed in the core epigenetic memory genes of H3K27me3 have not been suggested as a PRC2 recognition motif. Therefore, there is possibly an additional protein in PRC2 regulating the intergenerational epigenetic memory for H3K27me3. Recent studies found substantial evidence that transcription factors bind to specific DNA sequences and recruit the histone machineries to these regions (Benveniste, Sonntag, Sanguinetti, & Sproul, 2014; S. Lee et al., 2017). Further experiments would be highly valuable to test if transcription factors play a role in epigenetic memory through the recognition of sequences.

Integrating knowledge from publicly available databases enables a better understanding of complete picture of the biological systems (Çelen et al., 2015). By using WormBase (Yook et al., 2012), we identified that histone modifications marked many ncRNA promoters. For instance, in CeHR, 14% - 18% of the histone modifications were on piRNA promoters. piRNA can recruit histone methyltransferases and induce a multigenerational epigenetic memory (X. A. Huang et al., 2013; Shirayama et al., 2012). Based on these results it can be postulated that there is a feedback loop between piRNA and histone modifications where piRNA transcription is regulated by histone modifications and in return piRNA molecules recruit relevant enzymes for post-transcriptional modification of histones.

Only a small number of multivalent marks have been identified to date. For instance, H3K4me3/H3K27me3 and H3K4me3/H3K9me3 bivalent domains are found to be pivotal in maintaining this balance in embryonic stem cells and adipogenic master regulator cells (Matsumura et al., 2015; Vastenhouw et al., 2010). Strikingly,

we have observed that multiple histone marks are associated with single gene promoters. By using IF-SIM, we have visualized the existence of histone modification pairs in close proximities (within the footprint of Polycomb Complex) (Mohd-Sarip et al., 2006) and showed that the association of multiple histone marks on single gene promoters are not due to their enrichment variances on different cells. These observations imply several possibilities including 1.) histone marks occupy neighboring nucleosomes; 2.) histone marks coexist at the same nucleosome in an asymmetric conformation; and 3.) histone marks co-occupy the same histone tail (Voigt et al., 2013). Thus, our results indicate that the single gene promoters or other genomic regions can, in fact, be associated with multiple histone marks. Moreover, expression profiles of the histone-marked genes indicated that monovalent marks contribute to gene expression in a manner that is distinct from multivalent histone modifications. The only exception to this pattern was H3K27me3 for OP50 NGM and CeHR conditions, and H3K9me3 for OP50 NGM. These two marks have demonstrated the lowest multivalency with the other three marks. It is also possible that they show higher multivalency with other histone modifications that were not studied, causing different gene expression profiles on different chromosomes. Collectively, our results suggest that the histone modification combinations work jointly in the regulation of gene expression. Further studies can be valuable to examine the function of these potentially multivalent marks.

In this study, we employed a correlative approach to identify multigenerational histone modifications under environmental changes. The results reveal that individual histone modifications have variations in their epigenetic memory, and sequence motif recognition has a crucial function in the regulation of this variation. Moreover, we show up to five histone modifications can mark single gene promoters collaboratively which increases under environmental changes. The collaborative marking of the genes results in differences in the gene expressions.

### **3.5 Materials and methods**

#### **3.5.1 ChIP-seq Analysis**

ChIP-seq for the histone marks were performed as described (Doh et al., 2016) with additional antibodies. Five micrograms of each anti-H3K27me3 (MBL MABI0323), anti-H3K4me3 (MBL MABI0304), anti-H3K9me3 (MBL MABI0318), anti-RNA Polymerase II (MBL MABI0601), anti-H4K20me1 (abcam ab9051), and anti-H3K36me1 (abcam ab9048) antibodies and 300  $\mu$ g of extracts were used. Quality control was implemented on the raw sequence data with FASTQC software (version 0.11.2). The raw reads were then mapped to the reference genome (WBCel235), obtained from Ensembl (Cunningham et al., 2014) along with the annotation file, with Burrows-Wheel Aligner (BWA; version 0.7.4)(H. Li & Durbin, 2009). Different peak calling strategies were followed for broad (H3K27me3) and other peaks. For broad peak calling, we called broad peaks separately for replicates with MACS software

(version 2) (Yong Zhang et al., 2008), with parameters `--broad --nomodel -B -p 0.00001 -m 10 35`, and SICER software (version1\_1) (Zang et al., 2009). Then, we detected the overlapping peaks for the replicates in OP50 NGM, CeHR, and reversion for H3K27me3, and identified the level of common peaks from these two softwares by using DiffBind (Ross-Innes et al., 2012) R package. We found that MACS not only captures 52.3% - 97.8% of the peaks obtained from SICER, but also identifies more peaks. Therefore, we decided to use MACS for peak calling. In addition, we examined several strategies to call peaks from replicates. We have found that the level of replicates are inconsistent and highly reduced when we merge the replicates and call the peaks or call the peaks separately and find the common peaks in replicates by using DiffBind (Ross-Innes et al., 2012). However, when we called the peaks distinctly and compared the peak annotations for the replicates, we have observed a consistent level of common peaks; thus, we have used this method for the replicates. Next, we examined the distribution profiles of the histone modifications on genes. The histone modifications showed similar annotation profiles of the gene regions (Figure 3.11). Since our focus of study was the effect of histone modifications on gene expression, we concentrated our analysis mainly on to the ones associated on the promoter/TSS region. Integrative genomics viewer (IGV) (Robinson et al., 2011) was used to visualize the bam files presented in Figure 3.6.

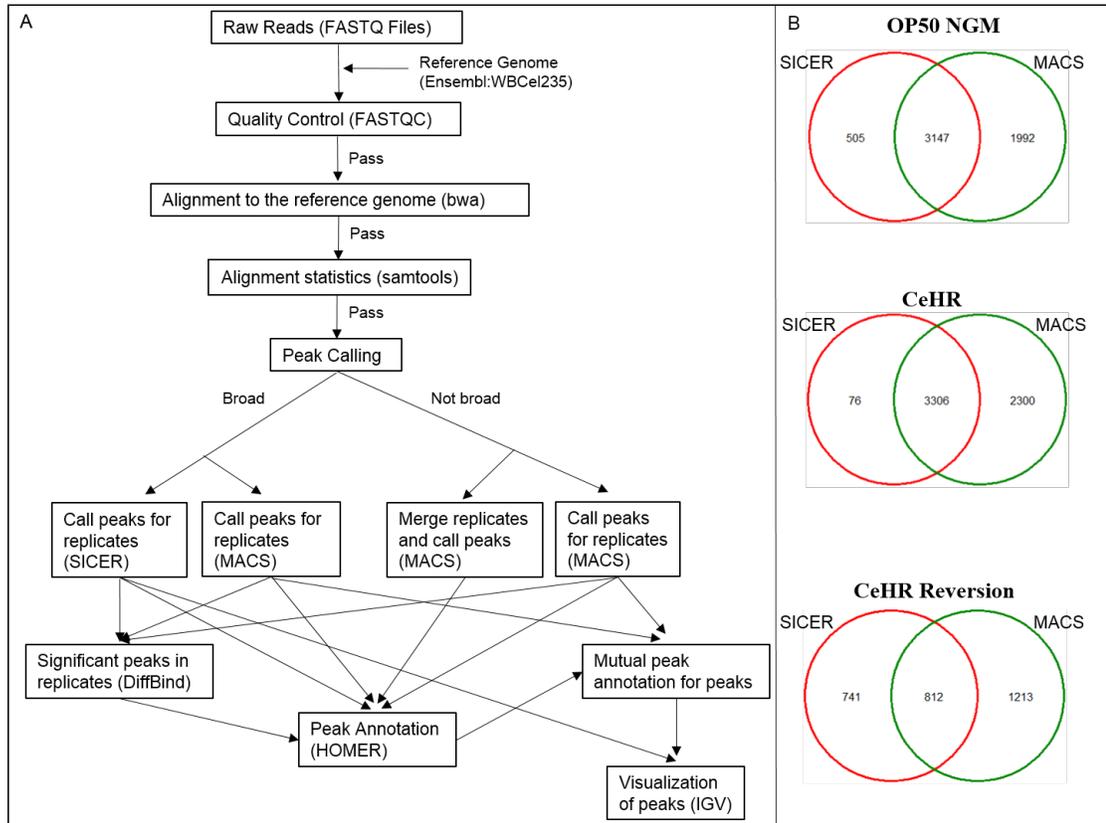


Figure 3.11: Pipeline development for ChIP-seq data analysis. (a) Workflow for the pipeline development. (b) Comparison of peak calling softwares for broad peaks called for H3K27me3.

### 3.5.2 Immunocytochemistry and Microscopy

Young adult worms were bleached with bleaching buffer (1:2:6 of 5N NaOH: 6% Bleach: 0.1M NaCl) for eight minutes, and rinsed with Mili-Q water three times. To fix the embryos, 2× Modified Ruvkun's Witches Brew buffer (MRWB; 160 mM KCl, 40 mM NaCl, 20 mM EGTA, 10 mM Spermidine, 30 mM PIPES at pH 7.4, 50% methanol) with 4% paraformaldehyde were used. The embryos were freeze cracked for one minute with liquid nitrogen, and thawed on ice for approximately 20 minutes,

and then washed with Tris Trinton buffer for two minutes. The samples were blocked with the blocking buffer (3% BSA and 3% Goat serum in Antibody buffer prepared as described previously (Duerr, 2006)) overnight at 4°C. For the immunostaining of the histone marks, the same antibodies from ChIP-seq experiment were used. After washing the embryos with Antibody buffer twice for 10 minutes each, primary antibodies (H3K9me3; 1:500 dilution, H3K4me3; 1:500 dilution, H3K36me1; 1:500 dilution, H3K27me3; 1:1000 dilution, H4K20me1; 1:250 dilution) were added, and kept at 4°C overnight. The samples were washed with the Antibody buffer four times and 1× PBS (with sodium azide) buffer once for 10 minutes each. The secondary antibodies (Goat anti-Rabbit AlexaFluor 568 and Goat anti-Mouse AlexaFluor 488) were added at 1:1000 dilution each, and the embryos were kept at 4°C overnight. We rinsed the samples with Antibody buffer for three times for 10 minutes, and with 1× PBS (with sodium azide) for at least five minutes. The samples were mounted with Prolong Diamond Antifade Mountant with DAPI (P36962). Super resolution – Structure Illumination Microscopy (SIM) images were acquired with a 63× Plan Apochromatic oil immersion objective on a Zeiss Elyra PS.I SIM, Carl Zeiss, Inc., Germany. The final z-stack images were created with ImageJ software.

### **3.5.3 Statistical Analyses.**

One-way ANOVA followed by Tukey-Kramer post-hoc test was used to test the expression difference among the chromosomes, and *p-value* < 0.05 was considered significant. Lloyd algorithm for k-means clustering was used to cluster the genes. Chi-

squared test with Yates correction was applied for detection of static or elastic epigenetic memory of histone modifications and RNAPII. The ncRNA molecules were obtained from WormBase (version WS252) (Yook et al., 2012). Genes annotated as piRNA, miRNA or rRNA were considered noise and discarded from the analysis as we did not perform total RNA-seq. All statistics analyses were performed by using the R software.

### **3.5.4 RNA-seq Analysis**

The cited protocol was followed for RNA-seq (Doh et al., 2016). The Tuxedo pipeline (Trapnell et al., 2012) was employed to find differential gene expressions with default parameters. The raw reads were mapped to the same reference genome (WBCel235) that was used in the ChIP-seq analysis.

Clustering and motif analysis. The heatmap was generated by using gplots R package (Warnes et al., 2015) with the gene expressions from the RNA-seq data. Only the genes with FPKM > 0 was included for the heatmap. Sequence motif enrichment analysis was performed by using MEME Suite (v.4.11.1) (Bailey et al., 2009). For the motif analysis of the promoter sequences, sequences 1,000 bases upstream of the annotated transcription start sites of the genes were acquired from UCSC Genome Browser database (Rosenbloom et al., 2015).

## Chapter 4

### COMPARATIVE TRANSCRIPTOMIC SIGNATURE OF THE SIMULATED MICROGRAVITY RESPONSE IN *Caenorhabditis elegans*

#### 4.1 Abstract

##### Background

Given the growing interest in human exploration of space, it is crucial to identify the effect of space conditions on biological processes. The International Space Station (ISS) greatly helps researchers determine these effects. However, the impact of the ISS-introduced potential confounders (e.g., the combination of radiation and microgravity exposures) on the biological processes are often neglected, and separate investigations are needed to uncover the impact of individual conditions.

##### Results

Here, we analyze the transcriptomic response of *Caenorhabditis elegans* to simulated microgravity and observe the maintained transcriptomic response after return to ground conditions for four, eight, and twelve days. Through the integration of our data with those in NASA GeneLab, we identify the gravitome, which we define as microgravity-responsive transcriptomic signatures. We show that 75% of the simulated microgravity-induced changes on gene expression persist after return to ground conditions for four days while most of these changes are reverted after twelve

days return to ground conditions. Our results from integrative RNA-seq and mass spectrometry analyses suggest that simulated microgravity affects longevity regulating insulin/IGF-1 and sphingolipid signaling pathways.

### **Conclusions**

Our results address the sole impact of simulated microgravity on transcriptome by controlling for the other space-introduced conditions and utilizing RNA-seq. Using an integrative approach, we identify a conserved transcriptomic signature to microgravity and its sustained impact after return to the ground. Moreover, we present the effect of simulated microgravity on distinct ceramide profiles. Overall, this work can provide insights into the sole effect of microgravity on biological systems.

## **4.2 Background**

Significant efforts have been made for the exploration of space. As such, NASA and SpaceX have planned round-trip landings by 2030s and placement of a million people in 40 years to Mars. However, space conditions present different types of environmental stress such as increased levels of radiation and reduced gravity (i.e., microgravity) that negatively affect biological processes by causing changes in the transcriptome leading to bone loss, muscle atrophy, and immune system impairment, to name a few examples. To overcome this problem, the affected biological mechanisms in space conditions should be uncovered to take precautions before the colonization of Mars.

The International Space Station (ISS) has provided a great opportunity to study the impact of space conditions on biological systems. However, potential confounders at the ISS are generally neglected. For example, experiments in the ISS introduces elevated radiation and reduced gravity along with the hypergravity during the launch. These mixed stress factors make it highly challenging to detect what precisely causes the observed phenotypes. To identify how the space conditions affect the biological systems, each of the factors require a separate investigation.

Microgravity is one of the major stress factors causing detrimental health effects on humans. Simulation of microgravity is a cost-effective way to study the impact of microgravity on biological systems. For this purpose, various platforms including drop towers, parabolic flights and space flights are available. Because of their advantages to enable long term simulations in an effective and economical manner, clinostats are widely preferred to study gravitational response in biological systems, mainly plants, for decades (Dedolph, Oemick, Wilson, & Smith, 1967; Gruener & Hoeger, 1991; Kessler, 1992; Sachs, 1887; Wuest et al., 2017). Clinostats rotate a sample around a horizontal axis, thereby exposing the sample to a rotating gravitational vector. This can reduce and even possibly remove gravitational bias in the development of an organism (Briegleb, 1992; Brown, Dahl, & Chapman, 1976; Dedolph et al., 1967; Häder, Hemmersbach, & Lebert, 2005; Hemmersbach, Volkmann, & Häder, 1999; Sievers & Hejnowicz, 1992).

*C. elegans* is an ideal model organism for space biology studies with its completely documented cell lineage, high reproduction rate, short lifespan, and high

similarity to the human genome (Çelen et al., 2017, 2018; Félix & Braendle, 2010; Hunt-Newbury et al., 2007; Lehner et al., 2006; Watson & Walhout, 2014). In the public bioinformatics repository of NASA (GeneLab; <https://genelab.nasa.gov>), four studies from three missions to the International Space Station (ISS) are available from *C. elegans* (Table 4.1). However, the experimental design of these studies highly varies. For example, these studies adopted different food source (e.g., axenic or monoxenic liquid media), duration of study, developmental stage, or the worm strain while the usage of microarray technology was common for all. We and others have reported substantial variations in the transcriptomic and phenotypical responses to different liquid cultivations of the worms (Çelen et al., 2017; Szewczyk et al., 2006). We additionally showed that two strains of *C. elegans* (N2 and AB1) exhibit highly different transcriptomic profiles under liquid cultivations (Çelen et al., 2017, 2018). Overall inconsistency among the space biology experiments on *C. elegans* highlights the necessity of a more organized and uniform experimental design for future studies where each stress factor is investigated individually.

This study investigates the sole effect of simulated microgravity on *C. elegans* transcriptome and the sustained impacts after return to ground control conditions. We report a group of putative microgravity-responsive genes, we define as the gravitome, through the integration of our in-house data with the re-analyzed *C. elegans* data from GeneLab. Through the RNA-sequencing (RNA-seq) and mass spectrometry analyses, we reveal the downregulation of the sphingolipid signaling pathway and its potential function in longevity under simulated microgravity.

Table 4.1: An overview of the *Caenorhabditis elegans* studies at the International Space Station. PRJNA146465 represents the study conducted by Universiti Kebangsaan Malaysia (UKM)

	<b>Food Source</b>	<b>Duration</b>	<b>Strain</b>	<b>Develop. Stage</b>	<b>Year</b>
<b>ICE-First</b>	CeMM	10 days	N2	Mixed	2004
<b>PRJNA146465</b>	CeMM	6 months	CC1	Mixed	2006
<b>CERISE-4</b>	S-Medium	4 days	N2	~L4	2009
<b>CERISE-8</b>	S-Medium	8 days	N2	Mixed	2009

### 4.3 Results

To dissect the biological processes affected under simulated microgravity and sustained after the exposure, we first cultured *C. elegans* in CeHR for three weeks on ground control condition. Then, we exposed the worms to clinostat-simulated microgravity for four days (Figure 4.1a) and observed the maintained impacts at four, eight, and twelve days after placing the worms back to ground conditions (Figure 4.1b). RNA-seq was performed for each condition with three replicates. Because RNA-seq can detect low abundance transcripts and achieve less noise in the data, unlike previous space biology studies on *C. elegans*, we preferred RNA-seq over microarray (Zhao et al., 2014).

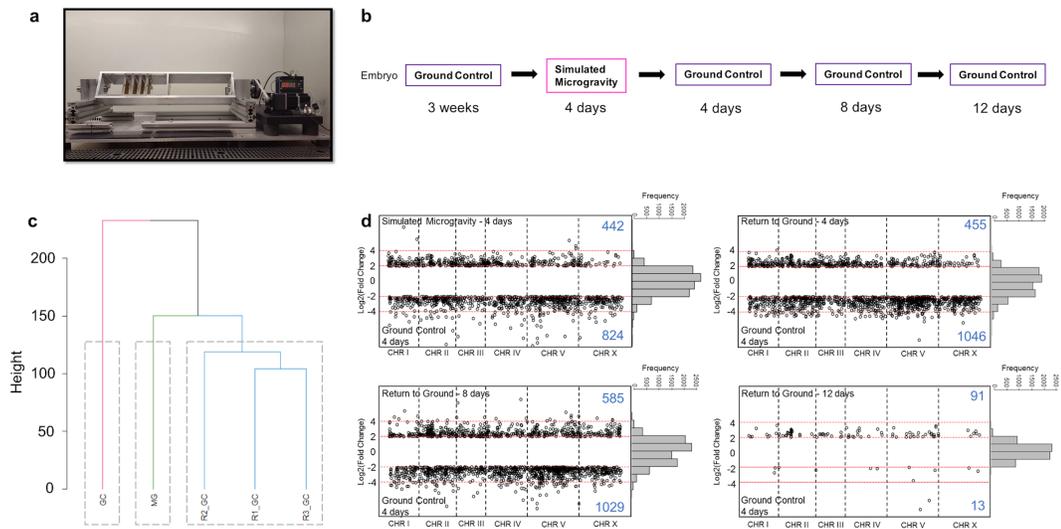


Figure 4.1: Transcriptomic response of *C. elegans* to simulated microgravity and return to ground conditions. (a) Clinostat used for simulating microgravity. (b) Experimental design for the effect of simulated microgravity on gene expressions during the exposure and after the return. (c) Dendrogram of the gene expression profiles for ground condition (GC), simulated microgravity (MG), four days after return to ground conditions (R1\_GC), eight days after return to ground conditions (R2\_GC), and twelve days after return to ground conditions (R3\_GC). (d) Log<sub>2</sub> fold change of the gene expressions (FPKM) between the conditions.

We generated a dendrogram for the transcriptomic profile in each condition tested and found significant differences during and after exposure to simulate microgravity compared to the ground control conditions (Figure 4.1c). Thus, this result suggests that exposure to simulated gravity induces highly distinct gene expression patterns which are maintained even after 12 days return to ground

conditions (approximately three generations of *C. elegans* in axenic medium (Doh et al., 2016)).

#### **4.3.1 Simulated microgravity triggers differential expression of hundreds of genes**

To identify the genes with the most distinctive expression levels after the exposure, we determined the differentially expressed genes (DEGs). The genes with over two-fold log<sub>2</sub> expression difference with the FDR-adjusted *p-value*  $\leq 0.05$  were considered as DEGs. Hundreds of genes demonstrated differential expression during exposure to simulated microgravity and up to eight days after return to ground conditions (Figure 4.1d). Twelve days after the return, the gene expression levels started to resemble the ones in the ground control with only 91 upregulated and 13 downregulated genes.

The determination of the spatial expression of the DEGs is of great importance to identify the potentially affected tissues. Thus, we performed tissue enrichment analysis for the DEGs in simulated microgravity (Figure 4.2a). The downregulated genes were overrepresented in neuronal and epithelial tissues, and intestine while the upregulated genes were enriched in the reproductive system-related tissues.

Next, we conducted pathway enrichment analysis for the DEGs in simulated microgravity and return to ground conditions to identify the altered pathways and whether they are remained to be altered after return to ground conditions (Figure 4.2b). Our results suggested an upregulation of dorso-ventral axis formation and

downregulation of lysosome during the exposure, and these expression patterns were maintained for four and eight days after the return, respectively.

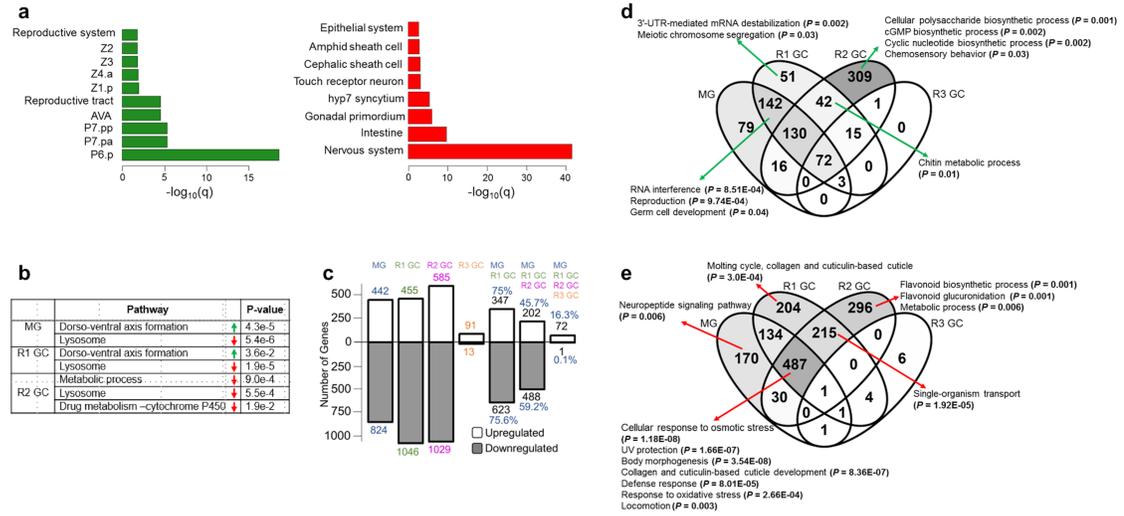


Figure 4.2: Differentially expressed gene profiles under simulated microgravity and after return to ground conditions. (a) Tissue enrichment of the upregulated (green) and downregulated (red) genes under simulated microgravity. (b) Pathway enrichment of the upregulated (green) and downregulated (red) genes during the exposure to simulated microgravity (MG), and four and eight days after return to ground conditions (R1 GC and R2 GC, respectively). (c) The number differentially expressed genes under simulated microgravity and the transmission of the differential expression after return to ground conditions. (d) Categorization of the upregulated genes in comparison to the ground control animals, and the enriched gene ontology terms assigned to them. (e) Categorization of the downregulated genes in comparison to the ground control animals, and the enriched gene ontology terms assigned to them.

### **4.3.2 Simulated microgravity-induced gene expression differences are highly maintained for eight days after return to ground conditions**

We then identified the number of simulated microgravity-induced DEGs that preserved their expression patterns after the return (Figure 4.2c). The majority of the DEGs (approximately 75%) from the exposed animals maintained their expression patterns after the return for four days. The shared number of DEGs decreased drastically at 12 days after the return to 16% for commonly upregulated and <1% downregulated genes with the exposed animals.

To elucidate the genes showing altered expression under simulated microgravity and maintaining these expression patterns after return to ground conditions for short (4 days) and long-term (12 days), we categorized the DEGs in Venn diagrams (Figure 4.2d, e). The genes solely upregulated in the exposed animals did not exhibit enrichment for any gene ontology (GO) term. The genes upregulated during the exposure and four days after the return, however, showed an overrepresentation for reproduction-related processes (Figure 4.2d). Chitin metabolic process-related genes were induced at four days after return to ground conditions, and the expression profiles were conserved for eight days after the return.

We found that the downregulated genes were enriched for the biological processes which are affected in space conditions. In particular, genes functioning in body morphology, collagen and cuticulin-based cuticle development, defense response, and locomotion were downregulated under simulated microgravity and up to eight days after return to ground conditions (Figure 4.2e). In simulated microgravity

and spaceflight studies, both small movement defects (Higashibata et al., 2006) and no difference in locomotory behavior (Qiao et al., 2013; Tee et al., 2017) have been noted. The reason behind these discrepancies is unclear, but our results indicated differences in the expression of the locomotion genes in close to weightless environment. Similarly, neuropeptide signaling pathway genes (*flp-22*, *flp-8*, *flp-26*, *nlp-10*, *nlp-12*, *nlp-17*, *nlp-20*, *nlp-24*, *nlp-25*, *nlp-26*, *nlp-39*, *nlp-33*, *nlp-28*, *nlp-29*, *nlp-30*, and *flp-24*) were downregulated during the exposure, and this expression profile was lost after return to ground conditions (Figure 4.2e). These neuropeptide signaling pathway genes also exhibit enrichment for movement variant phenotypes (Figure 4.3) indicating altered locomotion in response to simulated microgravity.

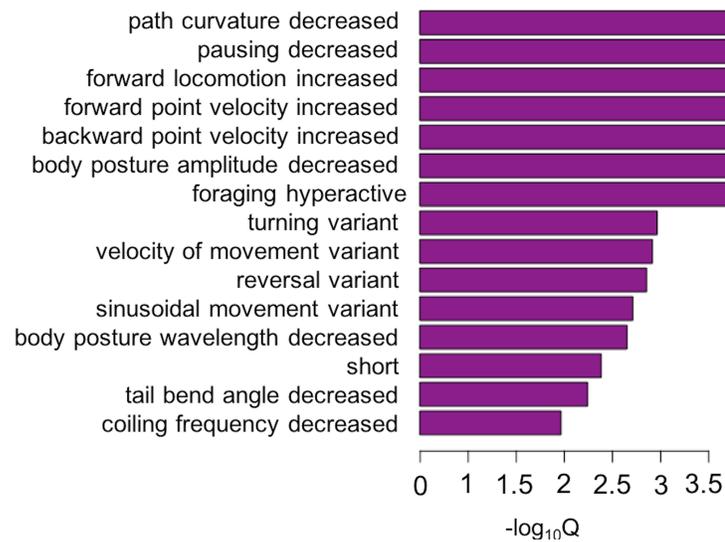


Figure 4.3: Phenotype enrichment analysis of the DEG neuropeptide signaling genes indicate a movement variation under simulated microgravity.

During the spaceflight, the neuromuscular system and collagen have been negatively affected (Higashibata et al., 2006). In agreement with these findings, collagen genes were downregulated in our experiment during the exposure, and the downregulation was sustained for eight days after return to ground conditions. Previously, Higashibata et al. (2006) reported decreased expression for the myogenic transcription factors and myosin heavy chains in space-flown worms (Higashibata et al., 2006). We did not observe differential expression for these muscle-related genes. Since their reported flight to ground gene expression ratio for the downregulated genes was less than one, and we only consider the genes with the log<sub>2</sub> ratio greater than two as DEG, it is expected not to observe those genes in our group of DEGs. To test whether other muscle-related genes are differentially expressed under simulated microgravity, we compared our DEGs to 287 genes reported in WormBase for involvement in muscle system morphology variant or any of its transitive descendant terms via RNAi or variation. Only T14A8.2, *col-103*, and *sqt-3* among the 287 genes showed downregulation under simulated microgravity (hypergeometric test, *p-value* = 0.99). Thus, our results did not suggest a significant change in the expression of the genes known to function in muscle morphology.

In our previous work, we revealed that the expression of many ncRNA molecules are triggered in response to environmental changes (Çelen et al., 2018). To identify whether a similar pattern occurs under simulated microgravity, we determined the expressed ncRNA molecules in ground control, simulated microgravity, and return. Since the housekeeping gene *pmp-3* presents consistent expression patterns

(Çelen et al., 2018), we considered its minimum expression level (FPKM = 24.7) in our experiment as the cutoff for the expression of ncRNA molecules. The number of expressed ncRNA molecules slightly decreased in response to simulated microgravity (Figure 3.4). Interestingly, this number increased drastically in the eight day of return. We found that the expression of 126 ncRNA molecules are induced in simulated microgravity and for 12 days after the return (Figure 4.4) while the expression of 16 ncRNA molecules are lost during and after simulated microgravity (Figure 4.4). Among the classified ncRNA molecules, mostly snoRNA and lincRNA molecules are induced while other set of snoRNA and asRNA molecules lose expression in simulated microgravity. For instance, asRNA molecules *anr-33*, K12G11.14, and ZK822.8 are induced whereas *anr-2*, *anr-9*, and Y49A3A.6 are silenced during and 12 days after the exposure.

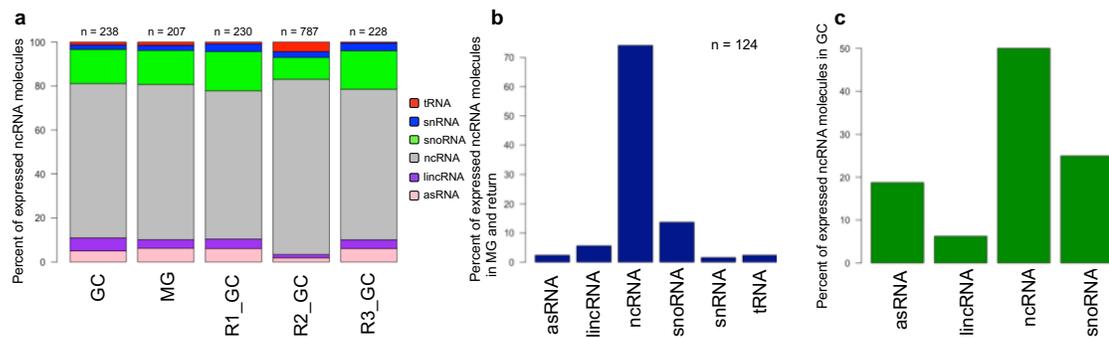


Figure 4.4: Simulated microgravity induced ncRNA molecules. (a) Percent of the expressed ncRNA molecules in ground control (GC), simulated microgravity (MG), four-, eight-, and twelve-days after return to ground conditions (R1\_GC, R2\_GC, R3\_GC). (b) The ncRNA molecules that are expressed during the exposure to simulated microgravity and twelve days after the return but not in ground conditions. (c) The ncRNA molecules that lost their expression and did not gain after the return.

We next sought to identify the putative transcriptional regulators (i.e., transcription factors) of the simulated microgravity-induced genes and whether these regulators maintain their impact after return to ground conditions. That is, we determined the putative transcription factor (TF) genes (Reinke et al., 2013) that are upregulated under simulated microgravity and after the return. Our results have revealed that 20 TF genes are upregulated during the exposure and 90%, 75%, and 15% of these genes maintained their upregulation after the return for four, eight, and twelve days, respectively (Figure 4.5). These TFs play a role in a variety of mechanisms such as double strand break repair or sex determination. For example, *tbx-43* is upregulated during the exposure and for at least 12 days after the return. The best human BLASTP-match of *tbx-43* in WormBase, TBR, functions in developmental process and is required for normal brain development (E-value =  $3e-33$ ; identity = 68.5%). However, many of the upregulated TFs do not have a known function. Hence, our list of simulated microgravity-induced TF gene expressions can be a rich source for the discovery of the microgravity-related transcriptional regulators.

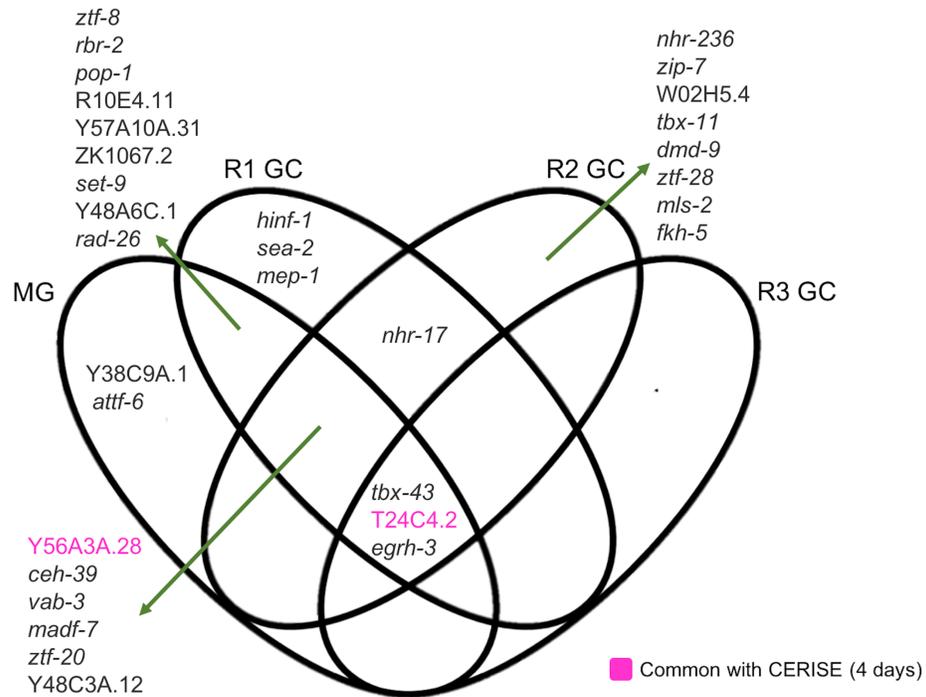


Figure 4.5: Putative transcription factor genes that are upregulated under simulated microgravity (MG), four days (R1 GC), eight days (R2 GC), and twelve days (R3 GC) after return to ground conditions. Y56A3A.28 and T24C4.2 are also induced in four-day CERISE.

### 4.3.3 Longevity regulating pathways are affected under simulated microgravity

Spaceflight affects the mechanisms involved in delayed aging in worms. For instance, Honda et al. found that age-dependent increase of 35-glutamine repeat aggregation is suppressed and seven longevity-controlling genes are differentially expressed in spaceflight (Honda et al., 2012; Honda, Honda, Narici, & Szewczyk, 2014b). To examine whether simulated microgravity induces such changes, we mapped the DEGs to longevity regulating pathway in the Kyoto Encyclopedia of

Genes and Genomes (KEGG). We used a less stringent criterion for differential expression by including the genes with  $\log_2(\text{FPKM}) > 1.5$  in this analysis. We determined 11 DEGs in the longevity regulating pathway genes most of which are involved in the insulin/insulin-like growth factor signaling pathway (insulin/IGF-1) (Figure 4.6). Interestingly, the transcription factor DAF-16 gene did not exhibit differential expression unlike its targets *sod-3*, *lips-17*, and *ctl-1*. Because the translocation of DAF-16 into the nucleus activates or represses target genes functioning in longevity, metabolism, and stress response, this finding was unexpected (S. S. Lee, Kennedy, Tolonen, & Ruvkun, 2003; K. Lin, Hsin, Libina, & Kenyon, 2001; C. T. Murphy et al., 2003; Seung et al., 2006). The nuclear localization of DAF-16 is antagonized by transcription factor PQM-1 (Tepper et al., 2013), and thus an upregulation in *pqm-1* may indicate inhibition of DAF-16 translocation. Our data, however, did not present a differential expression for *pqm-1* and thus it does not suggest inhibition of the DAF-16 translocation. It is possible that the DAF-16 targets are differentially expressed due to translocation state of DAF-16 or involvement of other factors (e.g., other transcriptional regulators).

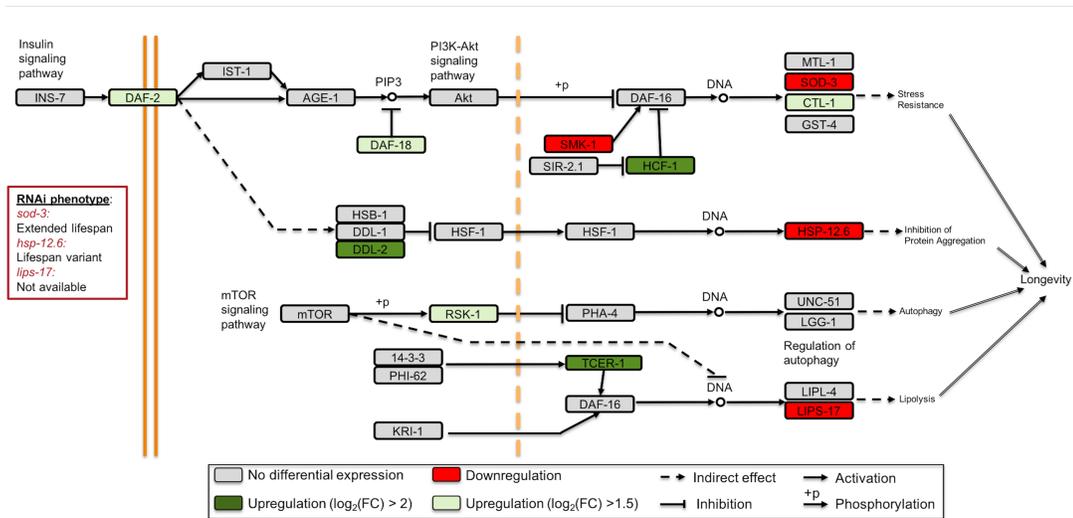


Figure 4.6: Longevity regulating pathway genes are differentially expressed under simulated microgravity. Adopted from the KEGG longevity regulating pathway – worm (cel04212).

The upregulation of the longevity regulating pathway transcriptional targets generally contributes to increased lifespan. Since many of these targets were downregulated under simulated microgravity, we wished to identify the potential impact of their downregulation by acquiring their RNAi phenotype from WormBase (Harris et al., 2014). Lifespan variant (WBRNAi00063155 and WBRNAi00063156) for *hsp-12.6* and extended lifespan (WBRNAi00064044) for *sod-3* have been reported. Follow-up experiments are needed to fully understand the role of these genes in the longevity regulation under simulated microgravity.

To further investigate the involvement of the longevity genes in simulated microgravity, we compared the DEGs to the DAF-16-responsive genes (Tepper et al., 2013). We found 144 DAF-16-induced genes were downregulated and this overlap

was statistically significant (hypergeometric test,  $p$ -value < 0.0001). The number of shared genes between the simulated microgravity-induced genes and the genes induced or repressed by DAF-16 are 11 and 35, respectively (hypergeometric test,  $p$ -value > 0.05 for both). Similarly, the number of downregulated DAF-16-repressed genes were insignificant with a total of 57 shared genes (hypergeometric test,  $p$ -value > 0.05). Together, our results suggest that the longevity regulation genes are affected under simulated microgravity and that along with the DAF-16-regulation, other mechanisms have an important function in this process.

#### **4.3.4 Sphingolipid signaling pathway is suppressed in response to simulated microgravity**

Previous studies have suggested that the sphingolipid signaling pathway plays a role in the expression of DAF-16/FOXO-regulated genes (Cui et al., 2017; Y. Kim & Sun, 2012). In line with this, our results showed that the sphingolipid signaling pathway is downregulated under simulated microgravity. That is, putative glucosylceramidase 4 gene (*gba-4*) and putative sphingomyelin phosphodiesterase *asm-3* are downregulated and these downregulation patterns were sustained for eight days after return to ground conditions (Figure 4.7a). This downregulation pattern indicates an attenuation in the ceramide levels through a potential decrease in the degradation of sphingomyelin and glucosylceramide to ceramide. Along with the other biological functions such as autophagy, senescence, and apoptosis, the sphingolipid signaling pathway has critical functions in ageing and longevity regulation (Cui et al.,

2017; Cutler, Thompson, Camandola, Mack, & Mattson, 2014; X. Huang, Withers, & Dickson, 2014). The inhibition of this conserved pathway results in an extension of lifespan in animals from worms to humans (Cutler et al., 2014; X. Huang et al., 2014). For example, the inactivation of *asm-3* in *C. elegans* causes translocation of DAF-16 into the nucleus, promotion of DAF-16 target gene expression, and extension of lifespan by 14% - 19% (Y. Kim & Sun, 2012).

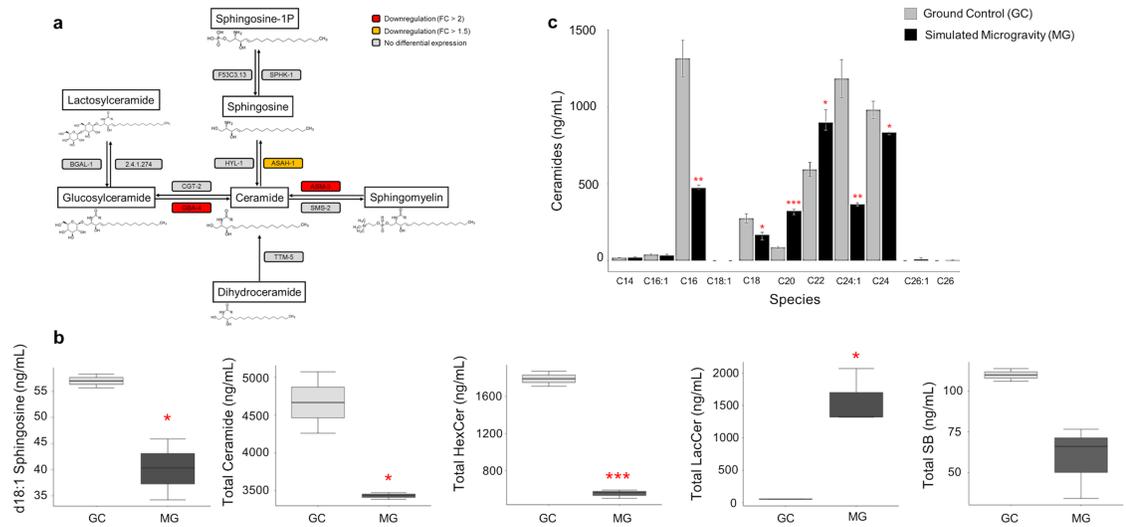


Figure 4.7: Spingolipid signaling pathway is downregulated under simulated microgravity. (a) The spingolipid signaling pathway genes *asah-1*, *asm-3*, and *gba-4* are downregulated under simulated microgravity. The downregulation pattern of *asm-3* and *gba-4* is maintained for eight days after return to ground conditions. (b) The levels of d18:1 sphingosine, total ceramide, hexosylceramide, and sphingoid base reduced while total lactosylceramide level increased under simulated microgravity. (c) The levels of different acyl-chain ceramides show alterations under simulated microgravity. The error bars represent the standard error of the mean (\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ )

To validate the ceramide and sphingosine levels are indeed decreased in response to the aforementioned gene downregulations, we performed mass spectrometry analysis in ground control and simulated microgravity-exposed worms concurrently (Figure 4.7b,c). Our results revealed that total ceramide, total hexosylceramide (HexCer), and d18:1 sphingosine are lower (1.5-, 3.6-, and 1.4-fold, respectively) under simulated microgravity (t-test, p-value < 0.05). Similarly, sphingoid base (SB) levels decreased 1.9-fold under simulated microgravity, but this decrease was not significant (t-test, p-value = 0.055) (Figure 4.7b). Interestingly, the levels of total lactosylceramide (LacCer) increased by 27-fold (t-test, p-value < 0.05). Increase in the LacCer and HexCer levels have been determined as biomarkers of aging in humans and murine (Hernández-Corbacho et al., 2011) while LacCer C18:1 was unaffected during aging in *C. elegans* (Cutler et al., 2014). It is unclear why LacCer levels are elevated when the HexCer levels are decreased, but the overall pattern indicates a downregulation in the sphingolipid signaling pathway under simulated microgravity.

The levels of long acyl-chain ceramides ( $\geq$ C24) are elevated with advancing age and in age-related diseases such as diabetes and cardiovascular disease while C20 and C26 ceramides are unaffected with aging or developmental stage in *C. elegans* (Bismuth, Lin, Yao, & Chen, 2008; Cutler et al., 2014; Strackowski & Kowalska, 2008). To further decipher the potential effect of simulated microgravity on aging-related mechanisms, we quantified the levels of different acyl-chain ceramides (Figure 3.7c). We observed a decrease in d18:1-C24, and d18:1-C24:1 ceramides (1.2-, and

3.2-fold, respectively) under simulated microgravity (t-test, p-value < 0.05). It has been found that the inhibition of C24-C26 ceramide-inducing ceramide synthase HYL-1 causes improvements in neuromuscular function and age-dependent hypoxia and stress response (Chan et al., 2017; Menuz et al., 2009). Hence, the lower levels of C24 and C24:1 ceramides in the simulated microgravity-exposed worms may indicate their positive impact on the stress response and aging of the worms.

Along with their roles in aging and longevity, different acyl-chain ceramides have other distinctive functions. For example, C16-ceramide induces germ cell apoptosis (Deng et al., 2008) and C20-C22 ceramide functions in resistance to hypoxia (Menuz et al., 2009). We determined that ceramides with different acyl-chain showed altered levels in response to simulated microgravity. While d18:1-C16 and d18:1-C18 exhibited a decrease (2.8- and 1.6-fold, respectively), d18:1-C20 and d18:1-C22 ceramides exhibited an increase (3.7- and 1.5-fold, respectively) under simulated microgravity (t-test, p-value < 0.05). Collectively, our findings suggest that sphingolipid and insulin/IGF-1 pathways play a role in simulated microgravity response.

#### **4.3.5 Identification of the “gravitome” through comparison of the microgravity-exposed worm gene expression profiles**

To investigate the “gravitome”, we compared the results from our simulated microgravity experiment to those from previous spaceflight experiments on *C. elegans* reported in NASA’s GeneLab database (<https://genelab.nasa.gov/>). For this analysis,

we used GEO2R to determine the microarray DEGs ( $|\log_2 \text{fold change (logFC)}| > 2$  and  $\text{adj. } p\text{-value} < 0.05$ ) from the first International *C. elegans* Experiment in Space (ICE-First), *C. elegans* RNA Interference in Space Experiment (CERISE) four and eight days, and PRJNA146465 conducted by UKM. We reasoned that the genes that are commonly differentially expressed in all the studies should be the ones responding to microgravity. Surprisingly, the number of shared DEGs were highly limited among the studies (Figure 4.9a). For instance, there were no shared DEGs between ICE-First and the other previous studies, and only five shared DEGs between four and eight days of CERISE. The experimental design of these studies has substantial differences including food source, developmental stage, strain, and exposure time (Table 4.1). These differences can be the underlying reason causing the highly distinct DEG profiles. For example, in the four-day culture of CERISE, approximately 9,000 L1 worms were used whereas in the eight-day culture, the number of worms was 30 to 50 L1 (Higashitani et al., 2009). Additionally, because the worms become gravid in four days in axenic culture (Doh et al., 2016; Samuel et al., 2014), the eight-day culture probably has mixed developmental stage worms as oppose to the synchronized worms in the four-day culture. As expected, the four- and eight-day CERISE cultures showed profound difference in the overall gene expression profiles, as well (Figure 4.9b).

Higashibata et al., have reported reproducible gene expressions for myosin, paramyosin, *unc-87*, *dim-1*, sirtuin and its target genes from four-day CERISE and ICE-First (Higashibata et al., 2016). We wished to evaluate the expression of these genes in eight-day CERISE, PRJNA146465, and our study, as well. Additionally, we

assessed the expression levels of these genes when a different analysis method (i.e., GEO2R) is used and noted the reported expression levels in different manuscripts for the same studies (Table 4.2). None of the genes were in our list of DEGs and most of them showed  $\log_{2}FC < 1$  between the conditions. Furthermore, some genes had high differences in the  $\log_{2}FC$  or discrepancy for up- or downregulation among the experiments. For example, *pqn-5* was reported to be 4.18-fold downregulated in the four-day CERISE while it was 0.9-fold upregulated in the ICE-First. The same gene was upregulated 0.25-fold in PRJNA146465 (Table 4.2). These findings highlight the necessity of uniformity in the experimental design and a detailed evaluation of the currently used techniques for achieving more reproducible and accurate results.

Table 4.2: Previously reported (Higashibata et al., 2016) reproducible gene expression changes from previous spaceflight studies in *C. elegans*. The values represent the logFC of the gene expressions under microgravity versus ground control. NR: not reported, NA: not available, \* (Higashibata et al., 2016), \*\* (Higashibata et al., 2006), \*\*\* (Then et al., 2016), # GEO2R, ## Cuffdiff2, ### DESeq2

Gene Name	CERISE (4 days)		CERISE (8 days)	ICE-First (10 days)			UKM (6 months)		Simulated Microgravity (4 days)	
	*	#	#	*	**	#	***	#	##	###
<i>myo-3</i>	-0.89	-0.78	-0.06	-0.34	-0.3	-0.28	3.12	1.61	-0.18	-0.14
<i>unc-54</i>	-0.49	-0.40	-0.14	-0.45	-0.5	-0.37	NR	-0.47	0.16	0.20
<i>unc-15</i>	-0.62	-0.52	NA	-0.33	-0.2	-0.35	NR	-0.44	0.10	0.15
<i>unc-87</i>	-0.25	-0.30	-0.01	-0.34	NR	-0.29	NR	-0.14	-1.02	-0.98
<i>dim-1</i>	-0.11	-0.39	-0.15	-0.53	NR	-0.34	NR	-0.01	-0.45	-0.41
<i>sir-2.1</i>	0.53	0.59	0.14	0.03*	NR	0.07	NR	0.15	0.56	0.61
<i>abu-7</i>	-1.24	-1.63	-1.18	-0.63	NR	-0.71	NR	-1.05	NA	NA
<i>abu-6</i>	-2.18	-1.91	-1.21	-0.78	NR	-0.72	NR	-0.40	NA	NA
<i>pqn-78</i>	-3.08	-3.11	-1.07	-0.80	NR	-0.73	NR	-0.28	NA	NA
<i>pqn-5</i>	-4.18	-4.89	-1.35	-0.90	NR	-0.41	NR	0.25	NA	NA
<i>lbp-6</i>	-0.47	-0.49	-0.14	-0.61	NR	-0.54	NR	0.43	-1.95	-1.89

We analyzed our RNA-seq data both with the Tuxedo pipeline (Trapnell et al., 2012) and DESeq2 (Love, Huber, & Anders, 2014) to confirm that the results are not dependent upon the analysis method. The results from both the methods were clustered together indicating their similarity (Figure 4.8b). The highest number of DEG overlap was between our experiment and four-day CERISE, and the overlap was significant for both the methods (hypergeometric test,  $p$ -value < 0.05) (Figure 4.8c). Since the overlap between the four-day CERISE and our results from the Tuxedo pipeline is higher, we decided to use the results from this pipeline throughout the manuscript.

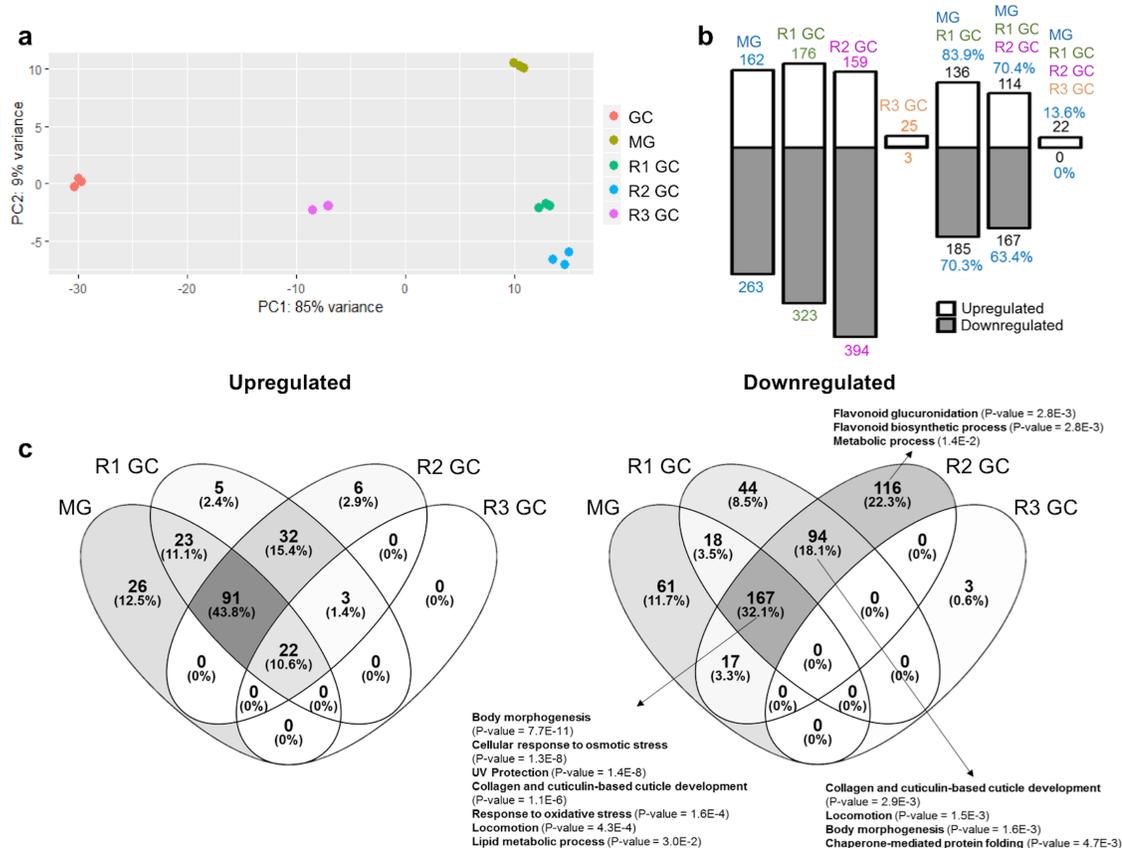


Figure 4.8: RNA-seq data analysis with DESeq2 (a) PCA analysis of the RNA-seq data from the ground control (GC), simulated microgravity (MG), and four-, eight-, and twelve-day after return to ground conditions. (b) The number differentially expressed genes under simulated microgravity and the transmission of the differential expression after return to ground conditions from the analysis with DESeq2. (c) Categorization of the upregulated (left) and downregulated (right) genes in comparison to the ground control animals, and the enriched gene ontology terms assigned to them from the analysis with DESeq2.

We determined a total of 134 common DEGs between our experiment and four-day CERISE, eight of them also being common to PRJNA146465, and categorized these 134 genes as the putative gravitome (Figure 4.9a, c). We discarded

16 DEGs showing conflicting patterns between the experiments (i.e., upregulated in one experiment and downregulated in the other) for further analyses. We reasoned that if the remaining genes are collaboratively involved in a biologically relevant process, they might have protein-protein interactions (PPIs) with each other. To test this hypothesis, we examined the enrichment for PPI of these genes by using the STRING database (Szklarczyk et al., 2017). Among 118 DEGs, 58 had known PPIs with each other (FDR < 1.0e-16) indicating their collaborative involvement in a biological process (Figure 4.9d). The GO analysis revealed that the gravitome genes play a functional role in locomotion, body morphogenesis, and collagen and cuticulin-based cuticle development (Figure 4.9e). Similarly, the gravitome gene products exhibit enrichment for nematode cuticle collagen N-terminal and collagen triple helix repeat domains (Figure 4.9f). Together, our findings suggest that mainly the collagen genes are affected under microgravity and this effect is reproducible between the studies.

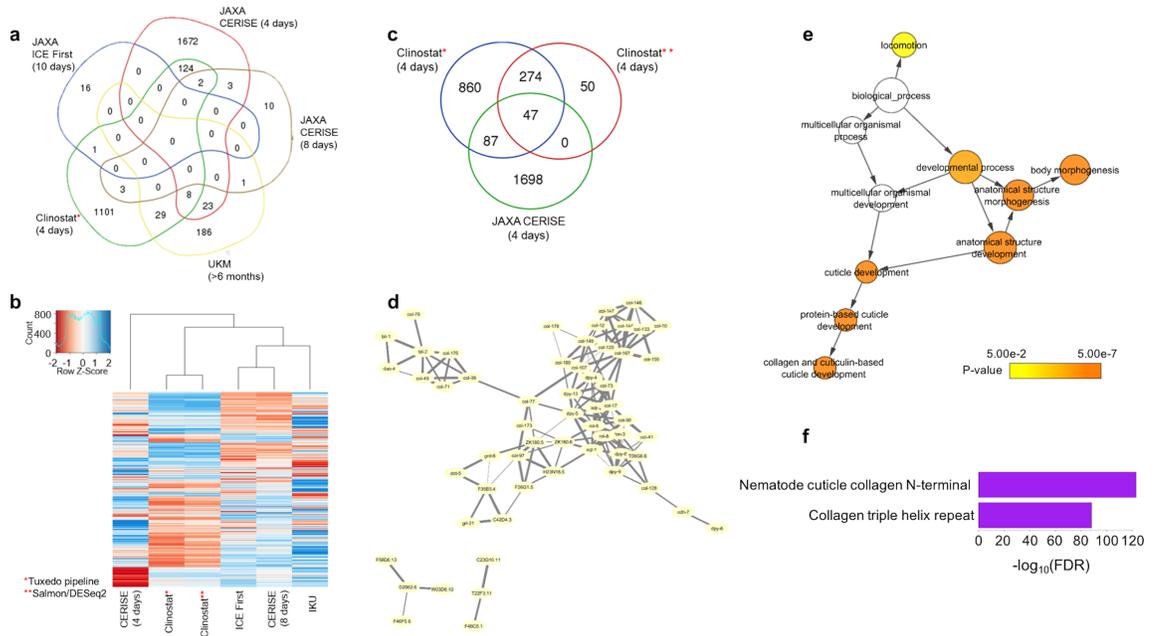


Figure 4.9: Gravitome genes are consistently differentially expressed in the ISS and under simulated microgravity. (a) Shared DEGs among the space-flown worms and our simulated microgravity experiment. The highest number of common DEGs are between our experiment and four-day CERISE. These genes are named as the “gravitome”. (b) The hierarchical clustering of the overall gene expressions among the space-flown worms and our simulated microgravity experiment. Our results analyzed with two different data analysis pipelines demonstrated highly similar patterns. (c) The number common of DEGs were high for our results analyzed with two different pipelines. Both of the pipelines showed common number of DEGs are higher than the values expected by chance (hypergeometric test,  $p < 0.05$ ). (d) Protein-protein interaction of the gravitome genes from STRING database. (e) Gene ontology enrichment of the gravitome genes. (f) Protein domain enrichment of the gravitome genes.

Next, we asked whether the human orthologs of the gravitome genes have relevant functions which might be affected in astronauts. We identified 64 human orthologs to 44 gravitome genes in worms. The human orthologs exhibited

overrepresentation for protein domains including calpain large subunit domain III and calpain family cysteine protease (Figure 4.10). Given the reported upregulation of calpain under microgravity and its function in muscle atrophy (Gao, Arfat, Wang, & Goswami, 2018; Salazar, Michele, & Brooks, 2010), these results provide an additional support for a conserved gravitome in worms and humans.

Our analyses revealed that 20 putative TF genes are upregulated under simulated microgravity conditions. We wondered whether these upregulation patterns are shared in the four-day CERISE. Among the 20 upregulated putative TF genes, Y56A3A.28 and T24C4.2 showed upregulation during the four-day CERISE, indicating a potential transcriptional regulation role for these two in response to microgravity. Our results additionally suggested that the upregulation pattern of these TFs are maintained for eight and twelve days after the return, respectively. We could not find a human ortholog for T24C4.2 while the WormBase-suggested human ortholog of Y56A3A.28 is PRDM16. PRDM16 regulates the switch between skeletal muscle and brown fat cells by inhibiting the skeletal muscle development and gene expression and stimulating brown adipogenesis (Seale et al., 2008). Therefore, Y56A3A.28 can be a strong candidate as a negative regulator of the muscle development under microgravity conditions.

## 4.4 Discussion

In this study, we examined the simulated microgravity responsive gene expression patterns and their maintained levels for four, eight, and twelve days after return to ground conditions. Longevity regulating pathways such as insulin/IGF-1 and sphingolipid signaling pathway were affected under simulated microgravity. We identified a putative gravitome by determining the common microgravity responsive transcriptomic signatures by incorporating our study to the previous ones. Moreover, we revealed that the gravitome is potentially conserved between worms and humans by performing function analysis for the human orthologs of the worm gravitome and identifying some of their potential transcriptional regulators.

Sphingolipid signaling pathway plays crucial biological roles such as lifespan regulation, apoptosis, and oxidative-stress response (Cui et al., 2017; Cutler, Pedersen, Camandola, Rothstein, & Mattson, 2002; Cutler et al., 2014). Our results revealed an overall downregulation of the sphingolipid pathway which is generally related to an increase in longevity (Cutler et al., 2014). A recent study identified decreased ASM-1 and ASM-2 levels in the ISS-housed worms indicating a similar downregulation in response to microgravity (Graveley et al., 2018). Mutant strains or RNAi knockdown of acid sphingomyelinase ASM-1, ASM-2, and ASM-3 genes increase the lifespan, ASM-3 being the most prominent one. Overall, both simulated and the ISS-introduced microgravity seem to contribute to a potential increase in longevity through the downregulation of the sphingolipid signaling pathway. Follow-up experiments are

needed to delimitate the contribution of sphingolipids on longevity under microgravity.

In the Twins Study of NASA, Scott Kelly was sent to the ISS for 340 days while his identical twin Mark Kelly remained on Earth. The preliminary results from this investigation showed that Scott Kelly's telomere length significantly increased during the mission. The telomere length of Mark Kelly, however, remained relatively stable (Edwards & Abadie, 2018). Telomere length and lifespan have a strong relationship. That is, longer telomeres are linked to longer lifespan and increased resistance to environmental stress while shorter telomeres are linked to accelerated aging and reduced longevity (Heidinger et al., 2012; Joeng, Song, Lee, & Lee, 2004; Monaghan, 2010; Sahin & Depinho, 2010). Ceramide functions in the regulation of telomere length (X. Huang et al., 2014; Saddoughi, Song, & Ogretmen, 2008). Considering our findings on decreased ceramide levels and the observations on increased telomere in the ISS, investigation of the ceramide and telomere length relationship might prove important to uncover the mechanisms affecting longevity in space.

Most of the space flight-triggered physiological changes are reverted after return to Earth (Honda, Honda, Narici, & Szewczyk, 2014a). Similarly, we observed that the majority of the simulated microgravity-induced DEGs (over 84%) are reverted 12-days after return to ground conditions. Our results have revealed that while some biological processes are only affected during the exposure, others can be affected for at least eight days (approximately two generations) after the return. For example,

neuropeptide signaling pathway genes were differentially expressed only during the exposure whereas defense response genes remain downregulated for eight days after the return. Approximately 50% of the astronauts from Apollo mission experienced minor bacterial or viral infections at the first week of their return. Later studies reported that space flight-induced reactivation of latent herpes viruses which lasted for a week after the return to Earth (reviewed in (Taylor, 2015)). These consistent observations indicate that the detrimental effects of space conditions (i.e., microgravity) on immune system carry over even after the return. Future investigations could provide insight into whether the duration of the exposure has an impact on the duration of the lasting effects.

The limitation of our study is the inclusion of mixed stage worms and the collection of samples from whole animals. This limitation is common among the aforementioned worm experiments conducted in the ISS. Considering the significantly different gene expression patterns among developmental stages and across cell types (Spencer et al., 2011), it is possible to retrieve biased results from the averaged gene expression levels. It is essential to address these issues in future studies.

Our integrative approach to define the gravitome has shown that the gene expression profiles widely vary among the experiments. Similarly, the comparison of spaceflight responsive genes in *C. elegans* and *Drosophila melanogaster* demonstrated only six common genes from the European Soyuz flights to the ISS (Adenle et al., 2009). The reason for these discrepancies can be the batch effects introduced due to the differences in the experimental designs. For instance, different

exposure times to the space conditions or the usage of different food source or strain can induce highly distinct responses in the worms. Furthermore, during the missions to the ISS, the organisms experience drastic environmental changes including exposure to radiation, different temperature, and gravitational forces (i.e., hypergravity and microgravity). It is crucial to examine methods used for achieving the most accurate and reproducible results and to identify the effects of individual factors (e.g., diet and gravitational force). Towards that end, we previously reported the impact of liquid cultivation of two *C. elegans* strains on gene expression and phenotype (Çelen et al., 2017, 2018) and here, studied the impact of simulated microgravity.

## **4.5 Conclusions**

Given the elevated interest in the human exploration of space, it is crucial to determine the detrimental effects of the space conditions on biological systems. Our study demonstrates that *C. elegans* is an efficient model to identify the impact of space on biological systems. We believe that our results will be a valuable reference for the studies on transcriptomic response to microgravity by controlling for the renowned batch effects from space conditions, allowing the worms to acclimate to liquid cultivation before the experiment, presenting the sustained transcriptomic responses after return to ground along with their putative regulators, integrating the results from the previous studies conducted at the ISS, and providing the response of different ceramide profiles to simulated microgravity.

## 4.6 Materials and Methods

### 4.6.1 *C. elegans* strain and growth conditions

Wild-type N2 strain was obtained from the Caenorhabditis Genetics Center (CGC). The worms were grown at 21°C. Stocks of *C. elegans* were acclimated to CeHR medium for three weeks prior to microgravity experiments. The ground control animals were maintained in CeHR medium in 20 mL scintillation vials as described (Çelen et al., 2018). Approximately 100,000 live worms ( $P_0$  generation) were cleaned by sucrose floatation, their embryos ( $F_1$  generation) were harvested by bleaching and placed in ~40 mL of CeHR culture in sterile VueLife culture bags (“AC” Series 32-C, Saint-Gobain). Culture bags were mounted in the clinostat and rotated at 1 rad/sec (~10 RPM) for four days, which is the measured time to reach L4/adulthood (Doh et al., 2016). The “ $F_2$ ” and subsequent generations were approximated by using four-day intervals. Prior to RNA extraction, only live worms are retrieved using the sucrose floatation method. We note that “ $F_2$ ” generation and beyond will not maintain synchronicity, and all will contain a mixture of generations, but recovering only the live animals enriches the RNA pool for the expected (majority) generation as each adult will produce ~30 offspring. The reason for the usage of mixed stage worms were the difficulties experienced (e.g., contamination of the liquid medium) during the synchronizing of the animals.

#### **4.6.2 RNA isolation, Illumina sequencing**

RNA isolation and RNA-seq was performed with triplicates on Illumina HiSeq 2500 in 50bp single-end reads as described previously (Çelen et al., 2018). As we achieved high correlation for gene expressions with both RNA-seq and qRT-PCR in our previous work (Çelen et al., 2018), we did not perform a separate qRT-PCR for this study.

#### **4.6.3 Gene expression analysis**

Quality control on the RNA-seq data was done with FASTQC (version 0.11.2) (Andrews, 2010). All the reads were in the “very good” quality range. The DEGs were identified by using two separate approaches. First, the Tuxedo pipeline (Trapnell et al., 2012) was used with default parameters. Second, the DESeq2 software (version 1.16.1)(Love et al., 2014) was utilized after quantifying the expression of transcripts with Salmon (version 0.8.2)(Patro, Duggal, Love, Irizarry, & Kingsford, 2017). The reference genome (WBCel235) along with the annotation file were retrieved from Ensembl (Cunningham et al., 2014). We considered genes as differentially expressed if FDR adjusted *p-values* < 0.05, and log<sub>2</sub> fold change > 2 unless otherwise stated. The results from the Tuxedo pipeline was used throughout the paper as explained in the results and the results from DESeq2 were reported in the supplementary data (Figure 4.8). We discarded miRNA, piRNA, and rRNA molecules from the analysis. The ncRNA molecules (WS250) and unconfirmed genes were obtained from WormBase

(Harris et al., 2014). Dendrogram was generated with Ward's clustering criterion by using R software.

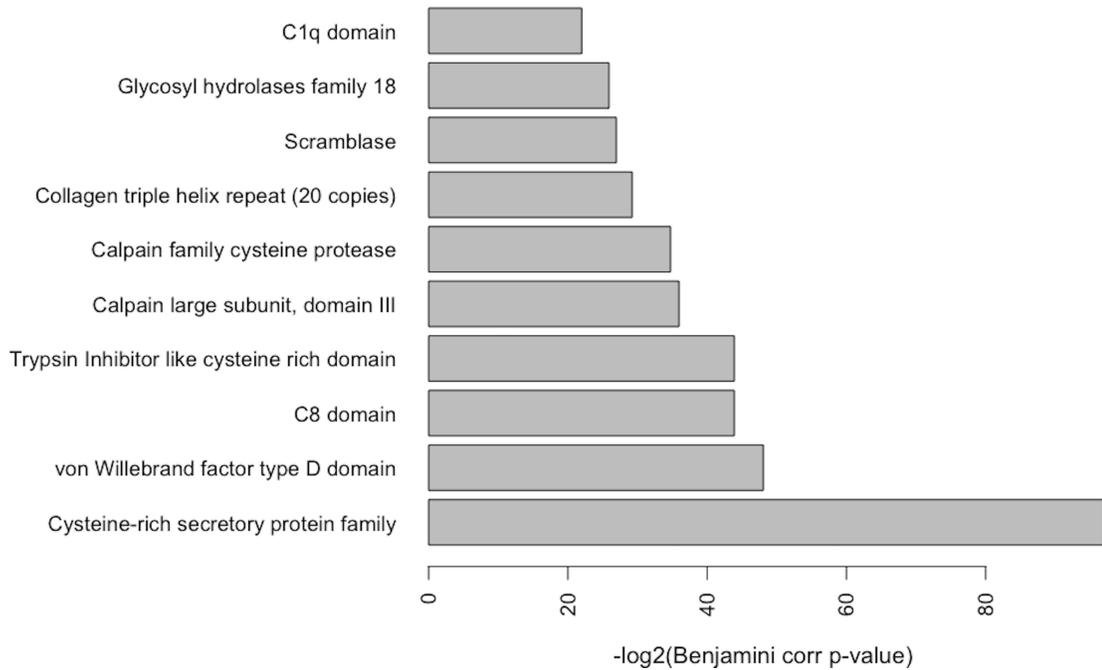


Figure 4.10: Domain enrichment of the human orthologs of the gravitome.

#### 4.6.4 Functional analysis of the genes

Gene ontology (GO) enrichment for biological processes and protein domains were assessed with Database for Annotation, Visualization and Integrated Discovery (DAVID) (v6.8) (D. W. Huang et al., 2008). The GO network for the common DEGs between our simulated microgravity experiment and CERISE4 was created in Cytoscape (version 3.4.0) (Shannon et al., 2003) by using the BINGO plugin (Maere, Heymans, & Kuiper, 2005). Pathway analysis was done using KEGG Mapper ([https://www.genome.jp/kegg/tool/map\\_pathway1.html](https://www.genome.jp/kegg/tool/map_pathway1.html)). Tissue enrichment analysis

was performed by using WormBase Gene Set Enrichment Analysis tool (Angeles-Albores et al., 2016), and FDR adj. p-value < 0.05 considered significant and top ten tissues were reported.

#### **4.6.5 Comparison of DEGs to the previous studies in *C. elegans***

NASA's GeneLab (<https://genelab.nasa.gov/>) platform was used to acquire the Gene Expression Omnibus (GEO) accession numbers for the previous studies on microgravity related gene expressions in *C. elegans*. We identified four such studies with the following GEO accession numbers: GSE71771 (Expression Data from International *C. elegans* Experiment First), GSE71770 (CERISE, 4 days), GSE27338 (CERISE, 8 days), and GSE32949 by UKM (PRJNA146465) from the query made on June 30, 2016 with "*C. elegans*" and "microgravity" terms on GeneLab. We analyzed these publicly available microarray datasets with GEO2R (<https://www.ncbi.nlm.nih.gov/geo/geo2r/>) for gene expressions under microgravity at the ISS versus 1G control in four and eight days CERISE and versus ground control conditions for the other two. The genes with logFC > 2 and Benjamini & Hochberg corrected p-value < 0.05 were considered as differentially expressed.

STRING database (version 10.5) (Szklarczyk et al., 2017) was used to determine whether the commonly DEGs from our simulated microgravity experiment and CERISE4 show enrichment for protein-protein interactions (PPI). We only considered the interactions only from experiments, gene-fusion, databases, and co-expression and selected "high-confidence (0.7)" results. PPI enrichment showed significance at *p*-

*value* < 1.0e-16 with 58 interacting nodes and 155 edges. The PPI network was visualized with Cytoscape (version 3.4.0) (Shannon et al., 2003). Edge width was determined based on the “combined score” for the interactions obtained from the STRING database. We identified the human orthologs of the gravitome with OrthoList 2 with default settings (W. Kim, Underwood, Greenwald, & Shaye, 2018).

#### **4.6.6 Mass spectrometry for ceramides**

The worms were chopped manually with razor blades on ice-cold glass. Three replicates for ground control and simulated microgravity conditions with mixed stage worms were used simultaneously. Each sample was sent to Avanti Polar Lipids Analytical Services as a frozen extract in glass tubes for liquid chromatography with tandem mass spectrometry (LC/MS/MS) experiment for detection of Ceramide levels. Each sample was extracted by modified Bligh and Dyer extraction. The samples were dried and resolvated in 50:50 Chloroform:Methanol and diluted prior to analysis. The resolvated sample was used for analyses and stored at -20°C until assayed. Samples were diluted as needed for analysis with internal standards for Ceramides for quantization by injection on LC/MS/MS. The individual molecular species for each sphingolipid group were measured by reversed phased liquid chromatography tandem mass spectrometry methods which separate the compounds by multiple reaction monitoring *m/z* to fragments and retention time. Quantification was performed by ratio of analyte to internal standard response multiplied by ISTD concentration and *m/z* response correction factor. The results are expressed as ng/ml. The consistency

among the replicates were examined with Pearson correlation. All but one replicates in both the experiments showed high correlation ( $R^2 > 0.9$ ). The ground control replicate which did not show a strong correlation (around 60%) with the other replicates was excluded from the analysis. The data were analyzed with t-test (two-tailed), and  $p$ -value  $< 0.05$  was considered significant. Equal variances were confirmed with Levene's test at  $\alpha = 0.05$  for all the ceramides tested.

#### 4.6.7 Microgravity simulation with clinorotation

We used a lab-built clinostat with integrated microscope to observe the motion of small objects including microspheres, yeast cells and *C. elegans* embryos during clinorotation. The trajectory of a spherical particle in a rotating vessel have been described previously in literature (Briegleb, 1992; Dedolph & Dipert, 1971; Kessler, 1992; Lee Silver, 1976). Dedolph and Dipert (1971) separated the motion of a spherical particle into two parts: a circular motion under the influence of gravity, the radius of which is inversely proportion to the angular velocity ( $\omega$ , rad/sec), and a radial motion due to centrifugal forces, the velocity of which is proportional to  $\omega^2$  (Dedolph & Dipert, 1971). The equation of motion that describes the orbital trajectories is given by:

$$m \frac{d^2 q}{dx^2} + (2i\omega m + \lambda) \frac{dq}{dt} - \omega^2 (m - m_w) q = -ig(m - m_w) e^{-i\omega t}.$$

In the above equation,  $\lambda$  is the viscous drag coefficient,  $m$  is the mass of the object,  $m_w$  is the mass of the displaced water,  $t$  is time,  $g$  is the acceleration due to gravity, and  $q(x,y,z,t)$  is the rotating reference frame defined as:

$$q \equiv x_{rot} + iy_{rot} = ze^{-i\omega t}.$$

As a test of the performance of the clinostat we recorded the motion of fluorescent melamine-formaldehyde microspheres (5.6  $\mu\text{m}$  diameter, 1.51  $\text{g}/\text{cm}^3$  density; Corpuscular Inc., Cold Harbor, NY) suspended in a 3% bovine serum albumin (BSA) solution (Additional file 10: Movie S2). This solution was injected into an 8-mm diameter, 100- $\mu\text{m}$  deep, cover glass topped chamber mounted on the clinostat microscope. Some of the melamine microspheres adhere to cover glass of the chamber, conveniently providing fiducial markers which we were able to use to compensate for the mechanical noise introduced into the micrographs by the rotation of the clinostat. We measured the terminal velocity of the suspended microspheres to be  $3.2 \pm 0.2 \mu\text{m}/\text{s}$  (mean  $\pm$  SEM,  $n = 8$ ). Figure S6 shows a plot of the measured radius of circular motion of suspended microspheres as a function of inverse angular velocity (black dots), and the calculated radius (Kessler, 1992) using the measured mean terminal velocity (blue line). The shaded area encompasses calculated radii for terminal velocities within one standard deviation of the mean.

## Chapter 5

### INTEGRATIVE ANALYSIS OF ENVIRONMENTAL STRESS RESPONSIVE GENE NETWORKS AND THEIR DRIVERS IN *Caenorhabditis elegans*

#### 5.1 Abstract

##### Background

Environmental changes have profound effects on well-being and survival of living organisms. For example, environmental stress factors have been found to account for over 70% of the chronic diseases, yet the knowledge about their genome-wide impacts remains limited. It is crucial to uncover the environmental responsive genetic mechanisms in systems level to identify the underlying mechanisms and putative therapeutic targets.

##### Results

We have built comprehensive gene co-expression networks with in-house and publicly available transcriptome data of *Caenorhabditis elegans* from different environmental stress conditions, such as exposures to microgravity and dietary alterations. Our gene networks showed high accuracy for predicting the known gene networks and are significantly enriched in previously found protein-protein interactions (PPIs) ( $P < 0.01$ ). Moreover, the networks predicted up to 72% more interactions for the known pathways through guilt-by-association. A high correlation was found between the function of the gene networks and the observed phenotype. By

incorporating transcription factor-target interactions into the networks, we identified the putative drivers of the gene co-expression networks.

### **Conclusions**

Overall, our combinatory approach can help facilitate exploration of the disease-driving genetic responses to environmental stimuli and create potential therapeutic targets.

## 5.2 Background

Environmental changes can cause detrimental effects on well-being and survival of living organisms. With the advancements in the scientific and technological developments, additional environmental changes are being introduced. For instance, a product named “soylent” has been commercialized as a dietary supplement claiming to include all the daily required nutrients for human consumption in a liquid form. However, the health impact of such changes in diet is not well understood. Similarly, there is an elevated interest for human exploration of space, yet the long-term consequences still require additional investigations. To overcome this problem, the affected biological mechanisms should be uncovered.

Gene co-expression networks are groups of genes with similar expression patterns across samples. Since genes sharing common biological functions or transcriptional regulation machinery are generally co-expressed, determination of the gene co-expression networks have a biological importance. WGCNA has been successfully used to uncover many systems level processes including gene networks playing essential roles in many diseases including cancer, Alzheimer’s and Huntington’s diseases (Chou et al., 2014; Mehta et al., 2018; Savas et al., 2017).

The inclusion of more data during the construction of genetic networks provides a higher statistical power and allows the retrieval of more accurate results. Thus, an approach integrating data from disparate resources can enable the construction of a more comprehensive network. For example, Talukdar et al., combined 612 gene expression profiles from seven tissues to generate co-expression networks for coronary artery disease patients (Talukdar et al., 2016).

Microgravity causes bone loss through an increase in osteoclasts and a decrease in osteoblast (Nabavi, Khandani, Camirand, & Harrison, 2011). Glucosamine induces osteoblast proliferation and might be a therapeutic agent candidate for astronauts against bone loss (Lv et al., 2018). Due to its positive impact on health including the stimulation of the immune system, vitamin C is supplied to astronauts (Crucian et al., 2018). Finally, the outcome of dietary glucose levels on biological systems is of interest to both ground and space conditions. Hence, we have integrated our RNA-seq data to those from glucosamine, vitamin C, and glucose-supplemented worms to generate co-expression networks.

While previous studies have successfully discovered many genetic networks in *C. elegans*, there still exist gaps in knowledge and discrepancies among these studies. Such fragmentation in knowledge is largely due to the intrinsic complexity of experimental determination of genetic networks. Additionally, the impact of environmental cues on these networks is often overlooked which in turn can cause contradictory results among studies as observed in the study by Kim et al (B. Kim et al., 2016). The significance of this study is multi-fold; the study: (i) computationally predicts the genetic networks with high accuracy and provide data-driven roadmaps for experimental validations; (ii) connects the missing links among previous studies by integrating the existing networks with the newly created ones; (iii) identifies the biological functions of the identified networks; and (iv) determines the drivers of the genetic networks.

### **5.3 Results**

To unravel the space environment-responsive gene networks, we have retrieved publicly available dietary change RNA-seq data in *C. elegans* and integrated them into our in-house RNA-seq data. Briefly, our in-house data included triplicates from OP50 NGM, CeHR (for four days and three weeks), S-Medium-grown and simulated microgravity-exposed worms as well as the worms placed back to OP50 NGM after CeHR and S-Medium cultivations. The three publicly available datasets consisted of triplicates from glucose, D-Glucosamine (GlcN) and vitamin C supplemented dietary changes (Dallaire et al., 2014; Ladage et al., 2016; Weimer et al., 2014). We re-analyzed the RNA-seq data, filtered out FPKM < 1 from the

analysis, and used WGCNA on remaining 8,816 genes to build the gene co-expression networks (Figure 5.1).

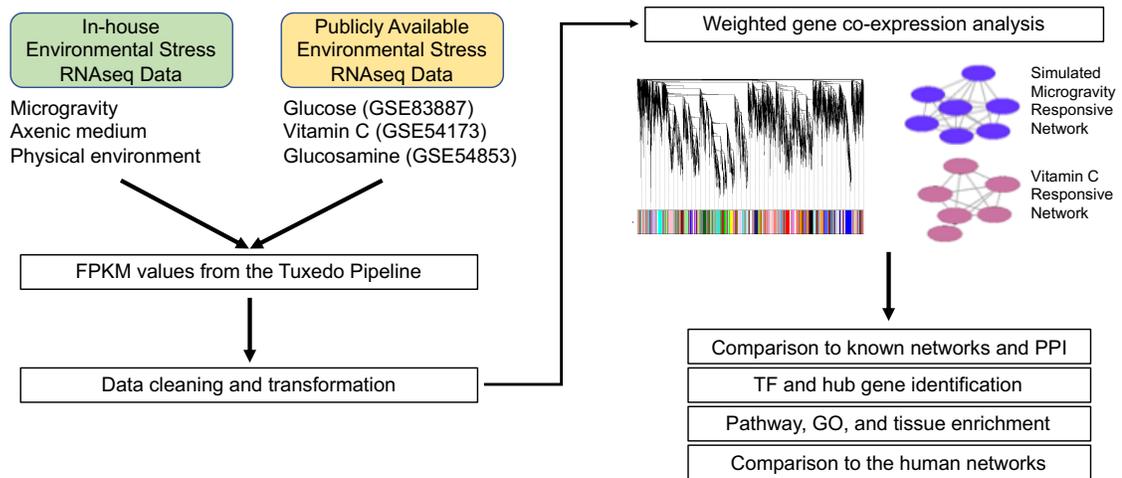


Figure 5.1: Experimental design for the gene co-expression network generation and downstream analyses

A total of 50 modules were identified from the integrated data. Our analysis for the module-trait relationships revealed the modules significantly correlating with the environmental conditions, which we consider here as traits (Figure 5.2 and Table 5.1). The grey module symbolizes the genes ( $n = 58$ ) which did not show high correlation with any of the modules. We found that 48 modules, excluding the grey module, have strong correlation with at least one environmental condition (20 with only one condition; 28 with more than one conditions) (Table 5.1; Table 5.2).

Table 5.1: Number of modules positively correlated with the tested environmental conditions

<b>Environmental Condition</b>	<b>Number of Modules</b>
OP50 NGM	9
CeHR (Four days)	6
S-Medium	2
Vitamin C	18
Glucose	9
Glucosamine	11
Simulated Microgravity	15
CeHR (Three weeks)	11
CeHR Reversion	0
S-Medium Reversion	2

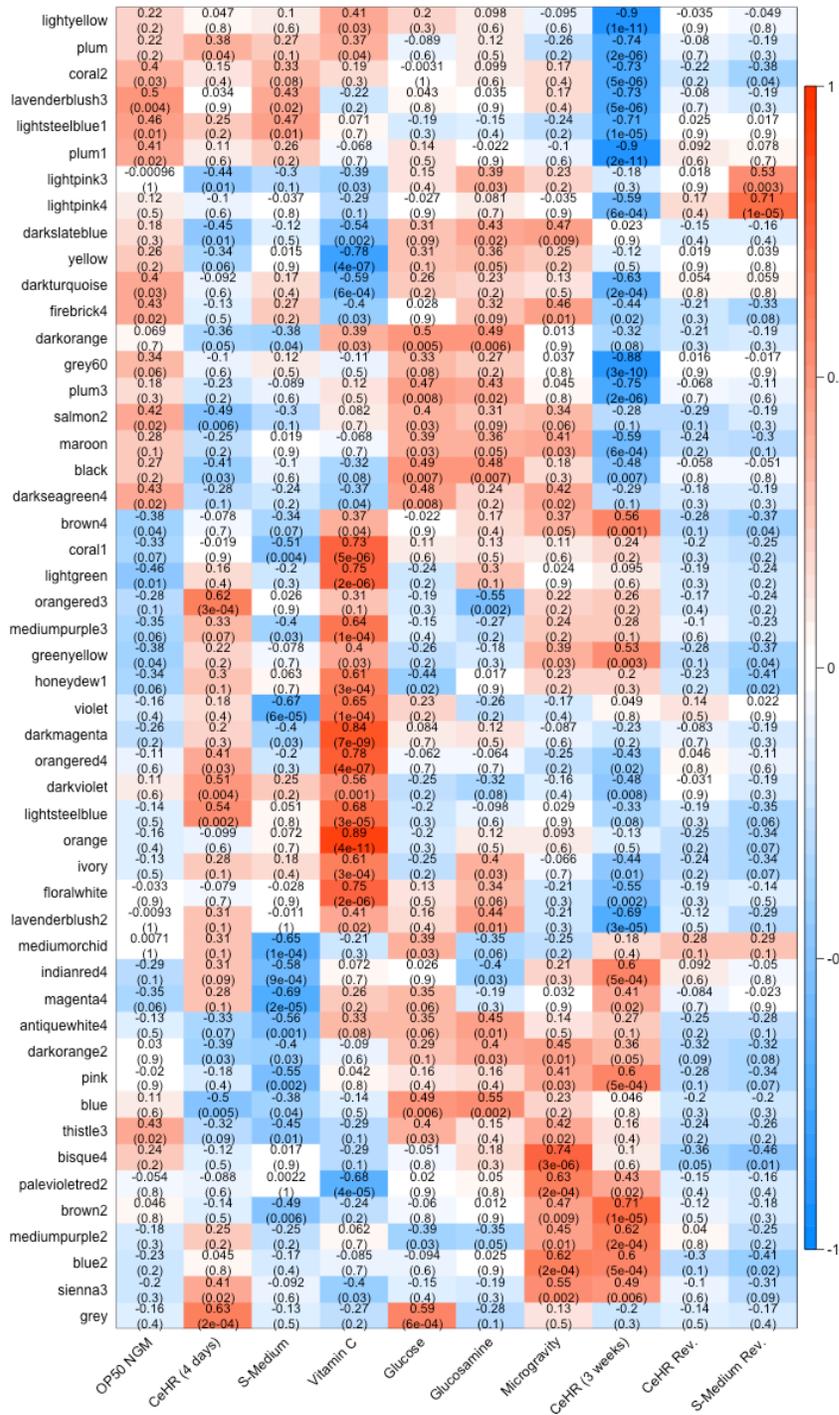


Figure 5.2: Module-trait relationship for the identified modules and the environmental conditions tested (red, blue, and numbers in parentheses represent strong positive and strong negative correlations and the p-values, respectively)

One way to evaluate the accuracy and robustness of the modules is to compare them to known gene networks and find the consistency among them. To evaluate the robustness of the generated modules, we compared the modules to the previously determined gene and protein-protein interaction (PPI) networks. WormNET v3 (Cho et al., 2014) is a gene network prediction server which enables maximum of 500 genes to be queried at a time. Thus, we were able to compare 47 of our modules having less than 500 genes. Our results depicted that 98% (46/47) of our queryable modules were significant in WormNET v3, and 86% (43/50) of the modules were enriched for the known PPIs in comparison to those in the STRING database (Szklarczyk et al., 2017) (Table 5.2). Moreover, our modules predict up to 74.2% and 20.7% more interactions than STRING and WormNET v3, respectively.

Table 5.2: Gene modules and their comparison to previously known gene- and protein-networks show that the majority of the modules have significant overlaps with previously identified networks. Red represents the results that are not significant

Module	# of Nodes	STRING Nodes	STRING P-value	WormNET Node	WormNET P-value	Hub Genes	Transcription Factor
antiquewhite4	615	602	< 1.0e-16	NA	NA	acbp-1	nsy-7
bisque4	85	83	0.0848	76	4.70E-14	smg-8	
black	632	615	< 1.0e-16	NA	NA	nrde-4	pat-1
blue	1152	1127	< 1.0e-16	NA	NA	mrpl-51	
blue2	119	115	1.71E-05	99	1.67E-09	C25G4.2	nhr-84
brown2	56	54	< 1.0e-16	48	2.67E-36	rps-5	
brown4	334	323	0.00142	281	5.88E-24	cal-1	
coral1	395	386	4.44E-15	330	5.32E-24	K07B1.4	elt-3; mxl-3
coral2	71	69	0.00556	58	1.91E-16	zer-1	
darkmagenta	95	94	0.0207	81	2.36E-05	deb-1	
darkorange	244	239	< 1.0e-16	209	2.91E-34	dnj-2	cey-1
darkorange2	389	383	< 1.0e-16	317	4.85E-63	C03H5.3	nhr-86
darkseagreen	74	73	9.45E-05	67	2.89E-38	E04D5.1	
darkslateblue	356	354	< 1.0e-16	309	1.26E-75	air-2	efl-1; hmg-12
darkturquoise	238	231	< 1.0e-16	197	3.27E-66	ced-3	
darkviolet	93	91	< 1.0e-16	84	1.03E-21	dapk-1	daf-16; hlh-30
firebrick4	57	55	6.59E-05	47	1.18E-17	ani-3	
floralwhite	87	86	0.876	71	2.43E-07	die-1	sem-2
greenyellow	166	164	< 1.0e-16	138	3.58E-30	flp-18	sma-4
grey	58	55	0.0208	46	2.27E-01	R02D3.1	
grey60	291	287	1.60E-11	244	6.25E-28	cash-1	athp-1; eor-1
honeydew1	76	74	9.10E-05	64	8.16E-12	jun-1	
indianred4	57	56	1.03E-05	49	1.12E-13	ttr-18	sex-1
ivory	90	86	4.82E-09	78	4.07E-34	ptrn-1	unc-62
lavenderblush2	36	35	0.36	31	4.93E-11	stau-1	skn-1
lavenderblush3	77	76	0.000212	69	1.02E-25	mrp-7	
lightgreen	143	141	0.000373	128	1.87E-26	F11D5.1	nhr-145
lightpink3	36	35	0.00128	28	3.90E-66	K07F5.14	
lightpink4	77	76	< 1.0e-16	70	9.84E-75	mix-1	
lightsteelblue	58	54	0.0371	48	6.04E-12	unc-14	daf-12
lightsteelblue1	171	169	1.59E-10	157	3.55E-33	sel-2	
lightyellow	139	138	< 1.0e-16	115	2.48E-20	csnk-1	nhr-49
magenta4	38	37	0.387	33	4.09E-02	B0410.3	
maroon	159	157	9.01E-05	140	4.04E-21	Y66D12A.1 5	
mediumorchid	64	63	0.00188	59	1.91E-21	col-142	
mediumpurple2	59	55	0.109	48	1.72E-10	T18D3.7	
mediumpurple3	200	198	1.68E-05	168	1.60E-19	arrd-17	hlh-11; nhr-

							177
orange	115	113	0.0911	96	1.44E-06	tes-1	
orangered3	59	58	1.76E-05	52	1.00E-08	ida-1	
orangered4	93	91	5.09E-05	80	1.32E-06	F07A11.4	blmp-1
palevioletred2	43	41	0.482	37	3.95E-01	C47D12.5	
pink	182	181	< 1.0e-16	161	9.54E-66	C30H6.7	
plum	60	57	0.0381	52	5.62E-04	lin-59	dpy-27
plum1	267	263	4.48E-08	218	3.53E-32	F52B5.3	
plum3	230	229	0.00118	200	4.92E-32	cdc-42	eFl-2; ref-1
salmon2	43	41	1.23E-06	39	2.69E-47	arp-1	
sienna3	94	90	1.68E-06	81	1.80E-17	hil-1	
thistle3	49	48	5.34E-10	45	6.55E-52	cif-1	
violet	98	98	< 1.0e-16	93	9.19E-31	dim-1	
Yellow	446	115	< 1.0e-16	391	9.15E-97	C30F12.4	

We next examined whether module membership (MM) and gene significance (GS) exhibit connection for the six mostly correlated condition-assigned modules. A high connection between the MM and GS is a strong indicator that the module has a highly important biological role in the assigned condition.

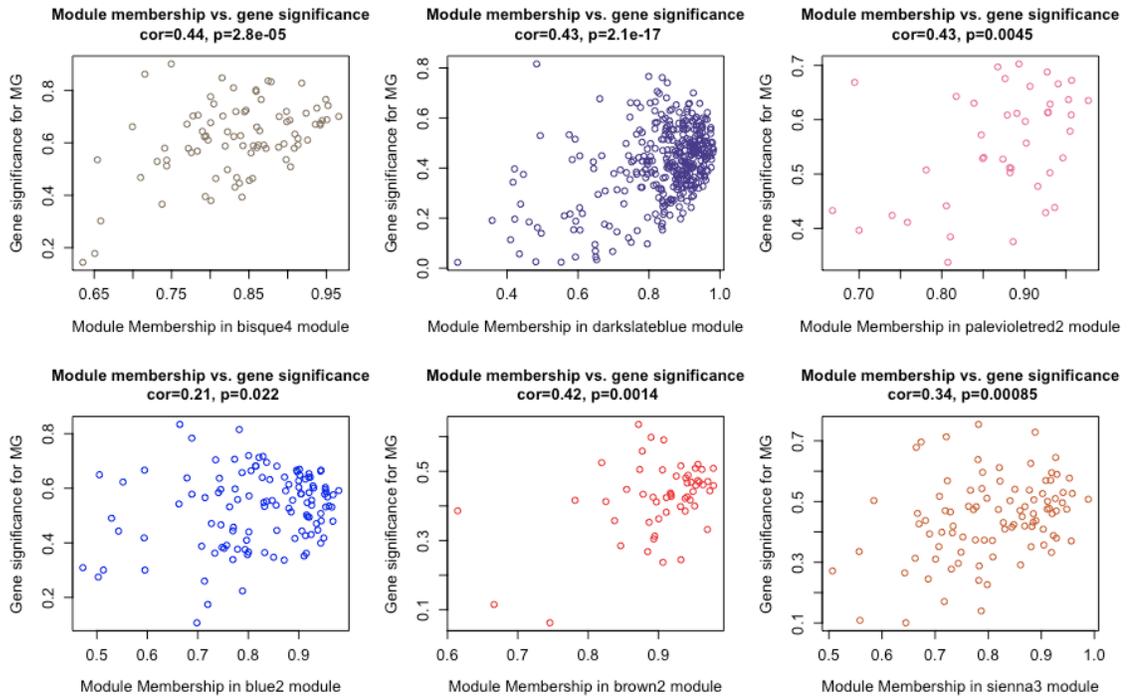


Figure 5.3: Module membership vs gene significance for the six mostly correlated modules with simulated microgravity (MG).

All of the six mostly correlated modules with simulated microgravity, glucose, and vitamin C supplement conditions and one three-week CeHR condition-associated module had significant correlations between the MM and GS indicating that these modules have important functions in response to these environmental conditions (Figures 5.3, 5.4, 5.5, and 5.6). None of the top six CeHR or two S-Medium condition-correlated modules, however, demonstrated significant correlations between the MM and GS while five glucosamine and two OP50 NGM-associated modules showed correlation between the MM and GS (Figures 5.7, 5.8, 5.9, and 5.10).

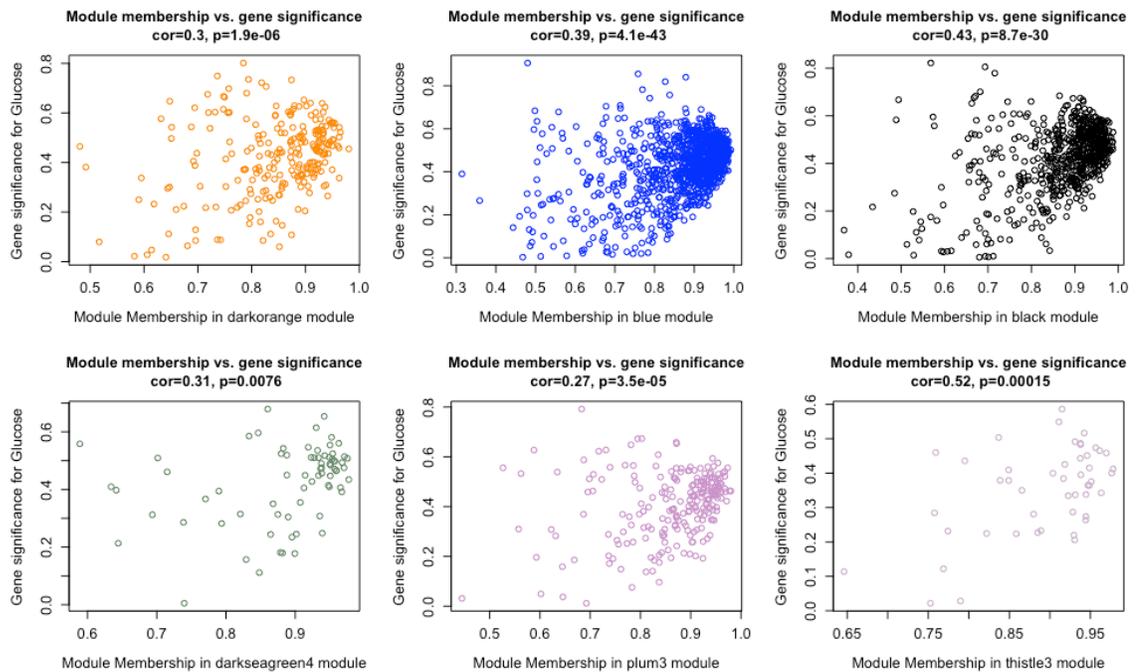


Figure 5.4: Module membership vs gene significance for the six mostly correlated modules with glucose supplemented diet

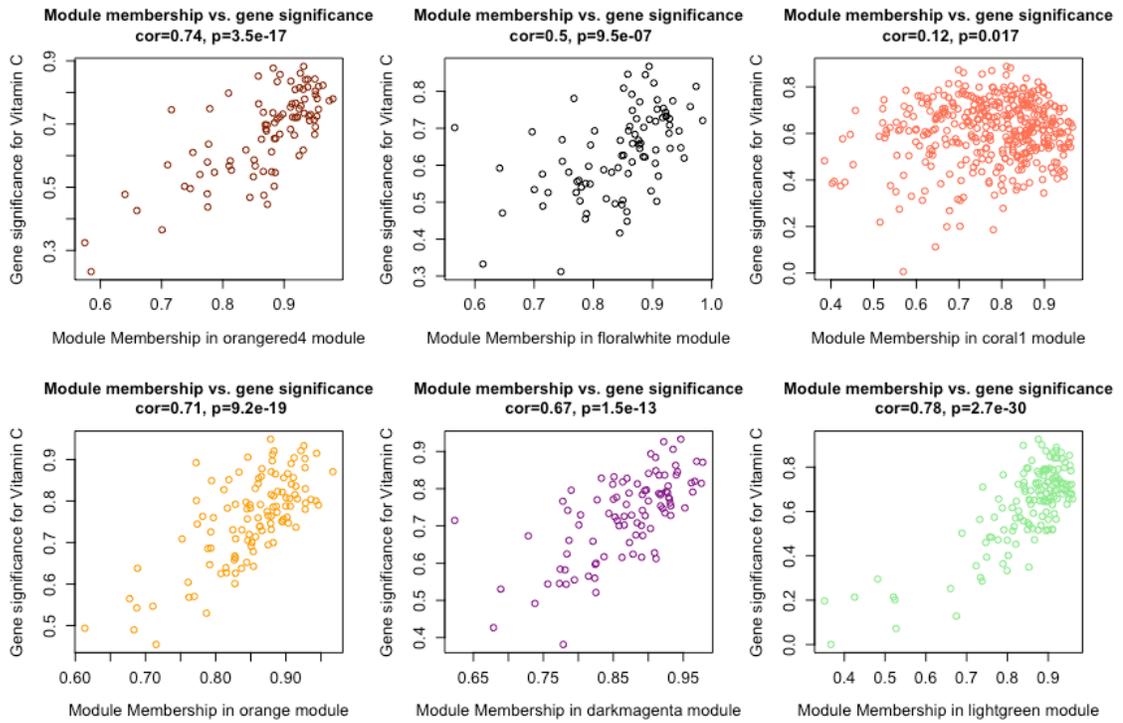


Figure 5.5: Module membership vs gene significance for the six most correlated modules with vitamin C supplemented diet

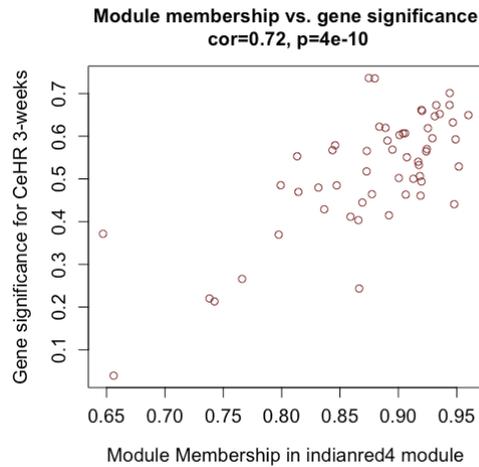


Figure 5.6: Module membership vs gene significance for the six most correlated modules with three-week CeHR cultivation

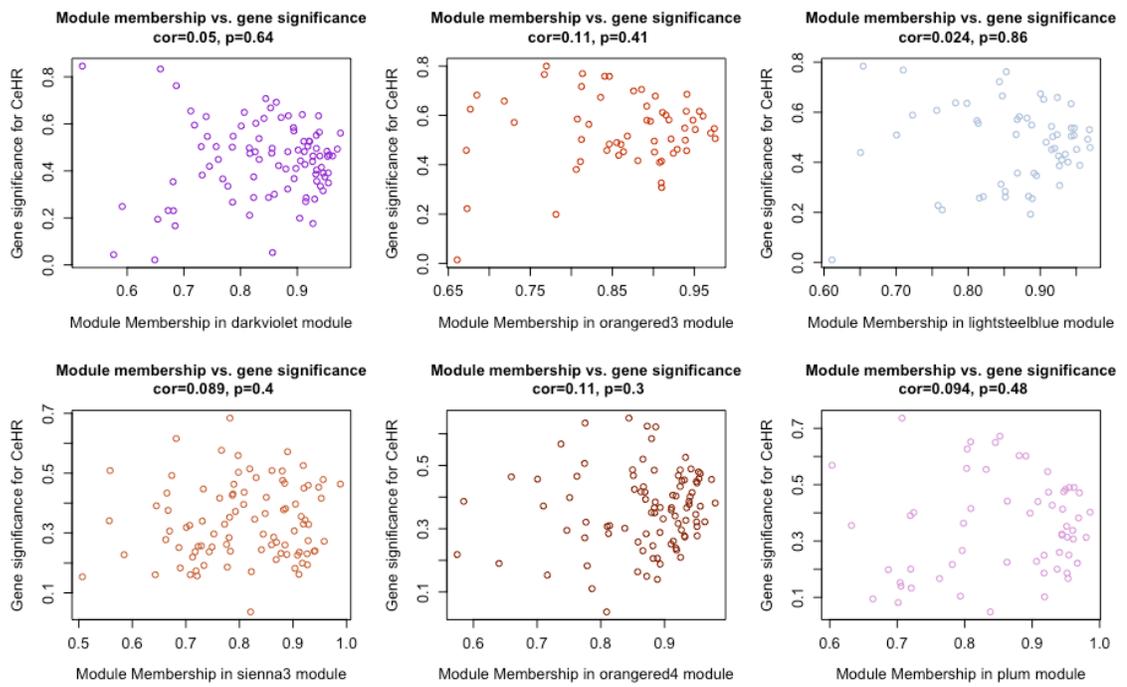


Figure 5.7: Module membership vs gene significance for the six mostly correlated modules with four-day CeHR cultivation

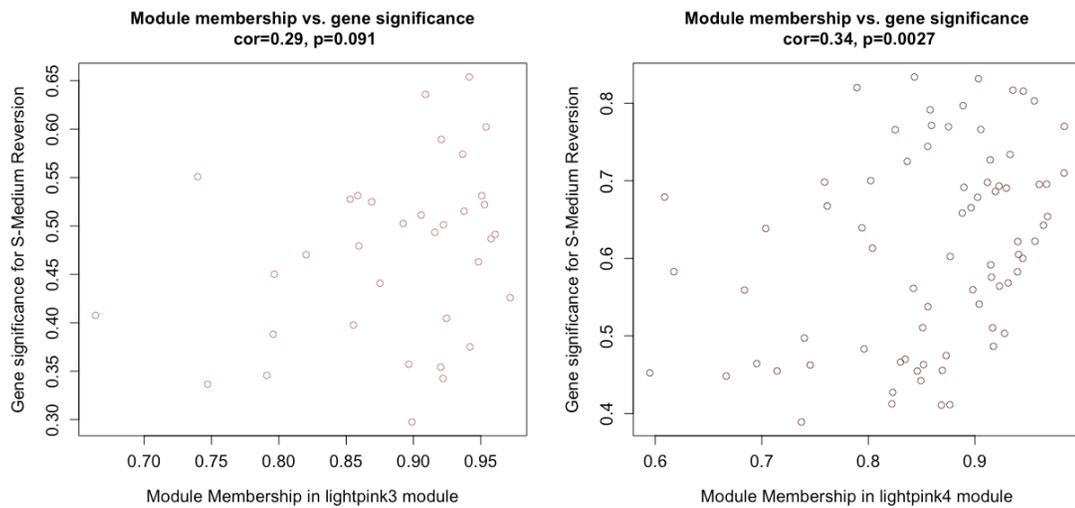


Figure 5.8: Module membership vs gene significance for the six mostly correlated modules with S-Medium cultivation

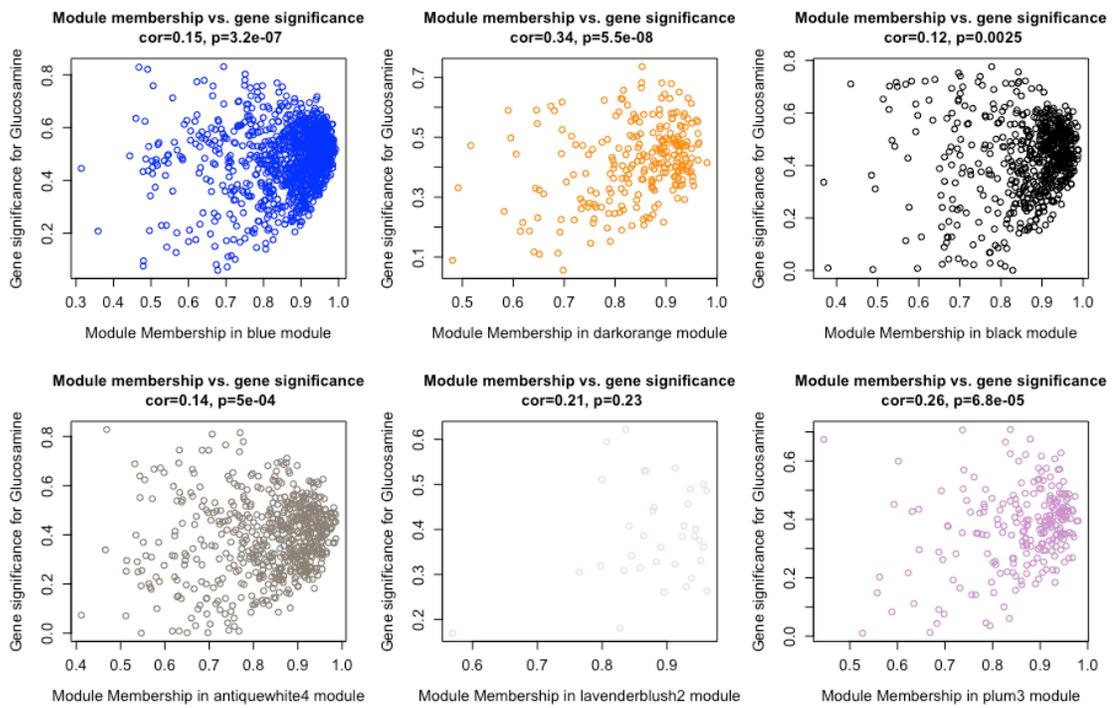


Figure 5.9: Module membership vs gene significance for the six mostly correlated modules with glucosamine supplemented diet

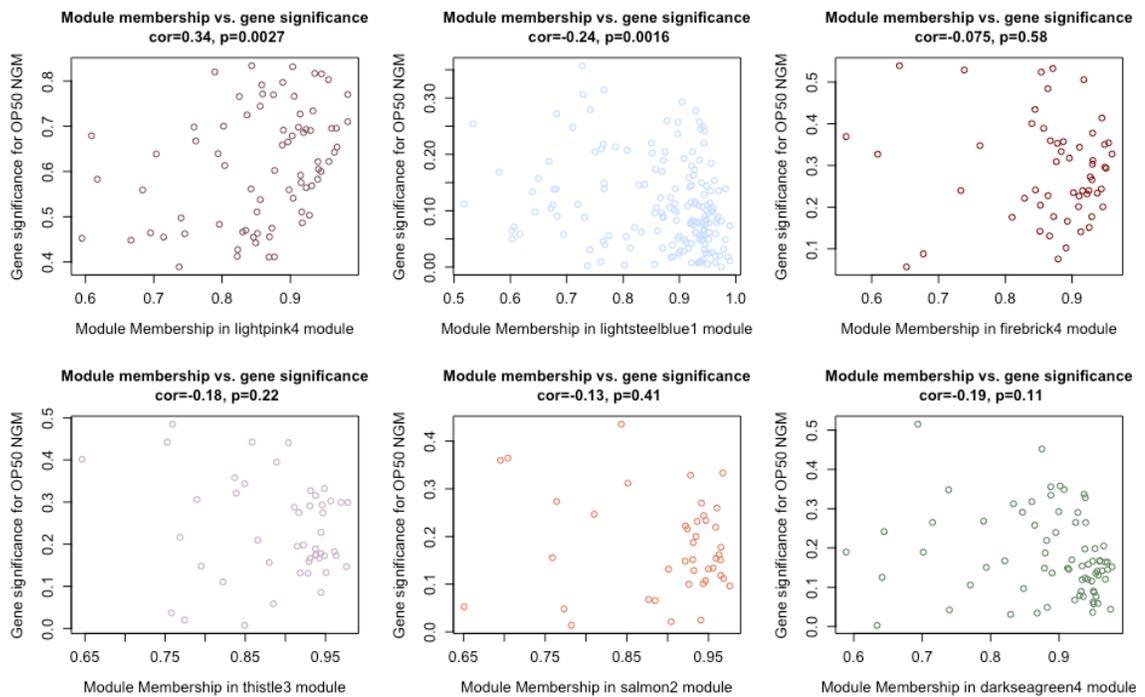


Figure 5.10: Module membership vs gene significance for the six mostly correlated modules with OP50 NGM growth condition

### 5.3.1 Co-expression network genes have biologically relevant functions to environmental stress conditions

To explore potential functions of the environmental condition-associated modules, we conducted further investigations on a set of them. The modules were selected based on either their enriched biological function for the corresponding PPI or significant correlation with an environmental condition or conditions.

Darkviolet module is correlated with vitamin C supplemented diet and CeHR cultivation conditions (0.65 and 0.51, respectively). The investigation of this module for PPIs exhibited that the module is enriched for the known PPIs (91/93 nodes have PPIs;  $p\text{-value} < 0.05$ ). The GO analyses of these PPIs revealed an overrepresentation

for the terms including regulation of response to stimulus, anatomical structure development, and regulation of cell communication (Figure 5.11).

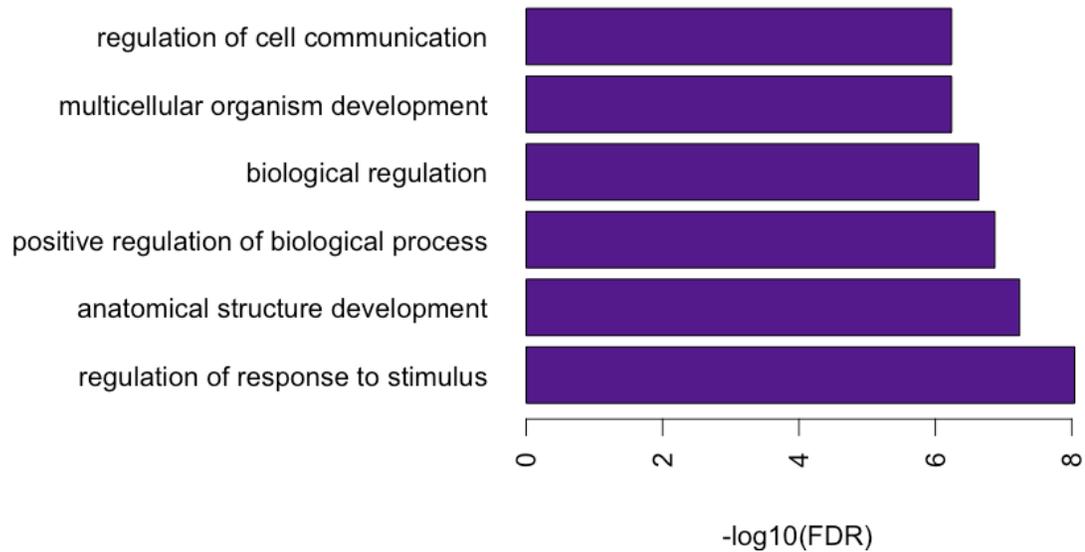


Figure 5.11: Top five GO biological process (BP) enrichments for the PPIs of the darkviolet module.

We also examined the KEGG and Reactome pathway enrichment results for the PPIs in the darkviolet module (Figure 5.12). The results indicated that the genes involved in the Wnt and calcium signaling pathways, and G-protein mediated signaling are enriched in the darkviolet module. Previous studies have shown that vitamin C intake increases the calcium content and inhibits osteoclastogenesis through the Wnt/ $\beta$ -Catenin/ATF4 signaling pathways in ovariectomized Wistar rats (Choi et al., 2019). Thus, further investigations may prove important for the identification of the underlying genetic networks between the vitamin C and calcium relationship.

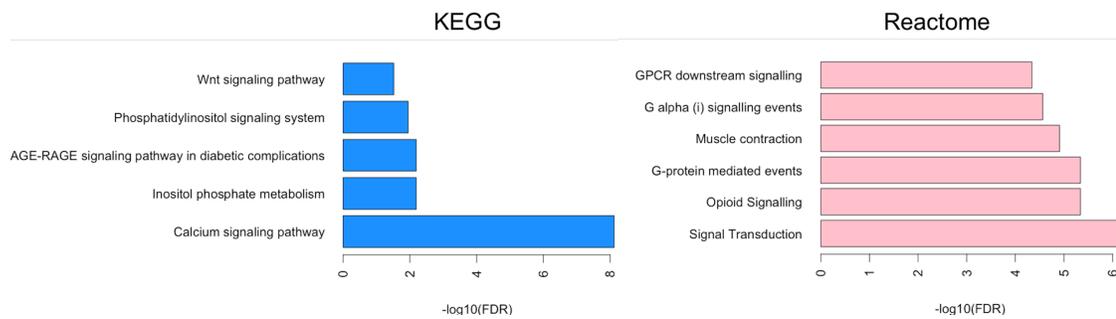


Figure 5.12: Top five KEGG and Reactome pathway enrichments for the PPIs of the darkviolet module.

Hub genes generally play crucial roles in biological processes and they have been categorized as the drivers of the networks for many studies (Langfelder, Mischel, & Horvath, 2013; van Dam et al., 2017). Similarly, transcription factors have regulatory functions in the cell and with high connectivity in a gene network, they are important candidates as the drivers. Identification of the drivers of genetic networks has a great importance as they can be promising drug targets. Hence, we have determined the genes with the highest connectivity as the hub. Furthermore, we mapped the reported transcription factors and their targets from TF2DNA into the modules (Pujato, Kieken, Skiles, Tapinos, & Fiser, 2014) (Figure 5.13). We found that transcription factor genes *daf-16* and *hlh-30* are in the darkviolet module and are connected with 58 and 60 genes, respectively, including each other. While the targets of DAF-16 were enriched ( $n = 17$ ;  $p = 0.002$ ; hypergeometric test), there were only targets (*daf-16* and *ell-1*) of HLH-30 were in the darkviolet module ( $p > 0.05$ ; hypergeometric test). In a recent study, Lin et al., found a crosstalk between DAF-16 and HLH-30 in which these two proteins form a complex and function in longevity

promotion and stress resistance to particular environmental stimuli such as oxidative stress in a conserved manner (X. X. Lin et al., 2018). It has been reported that the cultivation of worms in axenic liquid cultures increases the longevity (Doh et al., 2016; Szewczyk et al., 2006). However, we have not observed a differential expression in *daf-16* or *hlh-30* for the CeHR-grown worms in our previous study (Çelen et al., 2018). Given that the module that is potentially regulated by DAF-16 and HLH-30 is associated with the CeHR condition, it is possible that the targets of these transcription factors play a collaborative role in the longevity promotion in CeHR conditions.

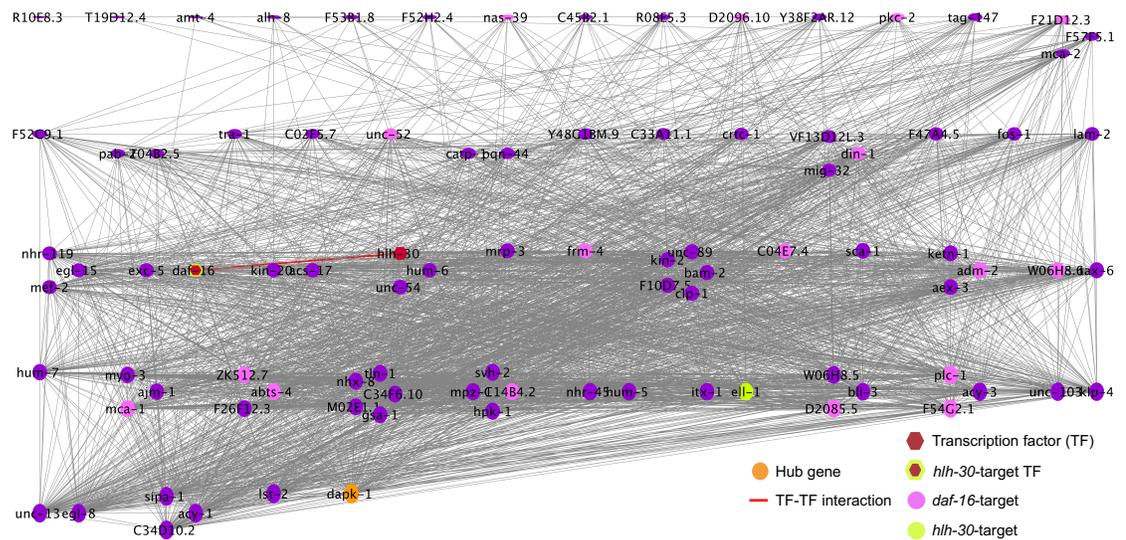


Figure 5.13: Darkviolet module is correlated with the vitamin C supplement and four-day CeHR cultivation conditions. The transcription factors *daf-16* and *hlh-30* and their targets are co-expressed.

However, there are controversial results about the impact of vitamin C on longevity. An earlier study on vitamin C supplement did not observe a longevity extension (Schulz et al., 2007). In a more recent study conducted by using RNA-seq, also used in our module generation, on the other hand, reported a significant increase in lifespan in vitamin C supplemented worms (Dallaire et al., 2014). A study on mice provided additional evidence that vitamin C promotes an extended lifespan (Son et al., 2018). Overall, our data seem to suggest an increased longevity through the transcriptional regulation of DAF-16 and HLH-30 in CeHR-grown and vitamin C-supplemented worm.

By having the highest connectivity, *dapk-1* is the hub of the darkviolet module. The main functions of DAPK-1 include innate immune response, autophagy, and apoptosis (Kang & Avery, 2010). Autophagy is an important mechanism for adaptation to abrupt environmental changes such as starvation or hypoxia and is required for lifespan expansion (Kang & Avery, 2010; Levine & Kroemer, 2008). In this module, *dapk-1* is connected with both *daf-16* and *hlh-30*. Together, our results indicate that darkviolet module functions in longevity expansion and autophagy may play a role in this process.

The darkslateblue module is correlated with the simulated microgravity exposure and glucosamine supplement conditions (0.47 and 0.43, respectively). In this module, proteins of 354 out of 356 genes have known PPIs which suggest that the module has biologically meaningful functions. The GO analysis of the PPIs in the

darkslateblue module exhibited an enrichment for nucleic acid metabolic process (Figure 5.14)

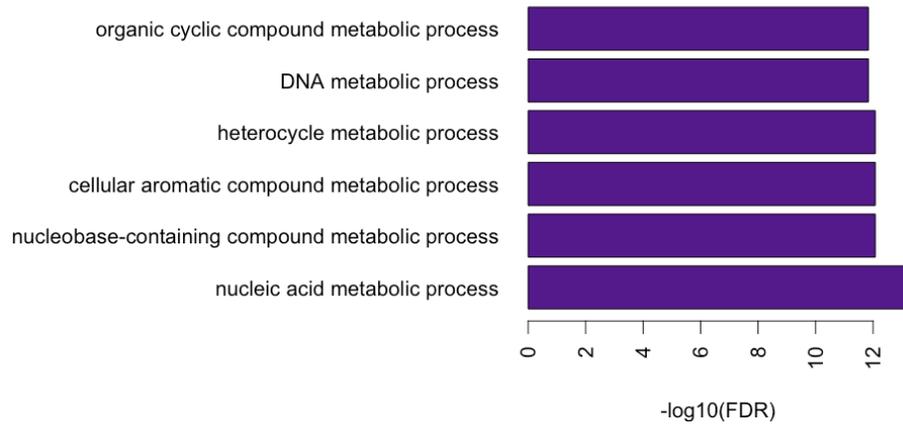


Figure 5.14: GO enrichment analysis for the known protein-protein interactions in the darkslateblue module

The pathway enrichment for the genes in darkslateblue module demonstrated overrepresentation for DNA replication and mismatch repair related pathways suggesting that this module may function in DNA repair mechanism (Figure 5.15). In 2014, a study explored the effect of glucosamine on hydrogen peroxide induced DNA damage in human lymphocytes. The results revealed that glucosamine has protective activity against hydrogen peroxide induced DNA damage even in lower doses (Jamialahmadi, Soltani, Nabavi fard, Behravan, & Mosaffa, 2014). Darkslateblue genes could be strong candidates as the underlying genetic mechanism for the glucosamine and DNA damage protection and repair. On the contrary, simulated microgravity has been found to induce DNA damage response defects in mouse embryonic stem cells and mouse embryonic fibroblast cells (N. Li, An, & Hang,

2015). Additional experiments are needed to uncover the impact of this module genes on DNA damage response under simulated microgravity.

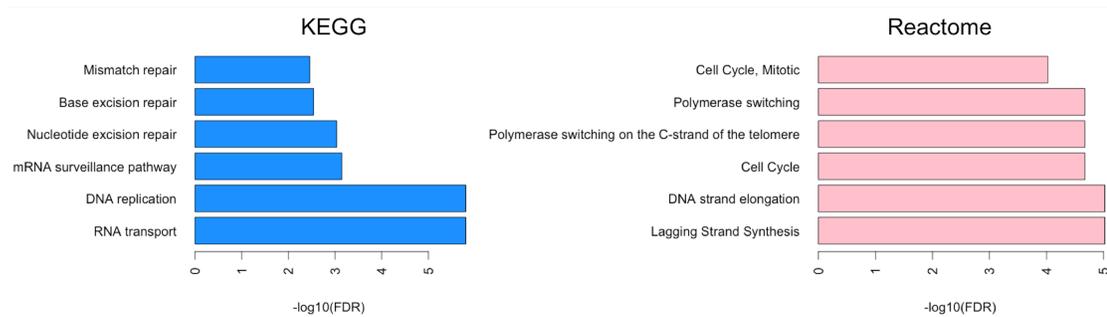


Figure 5.15: Pathway enrichment analysis for the known protein-protein interactions in the darkslateblue module

The genes of transcription factors HMG-12 and EFL-1 are found to be co-expressed in the darkslateblue module and to have high connectivity (323 and 337 out of 356, respectively). The hub gene is *air-2* for this module and it is connected to 342 genes. Interestingly, *air-2* is a target of HMG-12 indicating a potential crosstalk between these two as regulators of a more expanded biological network (Figure 5.16).

The targets of both HMG-12 and EFL-1 are significantly enriched in the darkslateblue module ( $n = 123$ ,  $n = 56$ , respectively;  $p < 0.0001$ ; hypergeometric test). We have identified that 22 of the module genes are the targets of both HMG-12 and EFL-1. Taken together, the transcription factors HMG-12 and EFL-1 are strong candidates as the drivers of the darkslateblue module, and the hub gene *air-2* plays a combinatory function with these two.

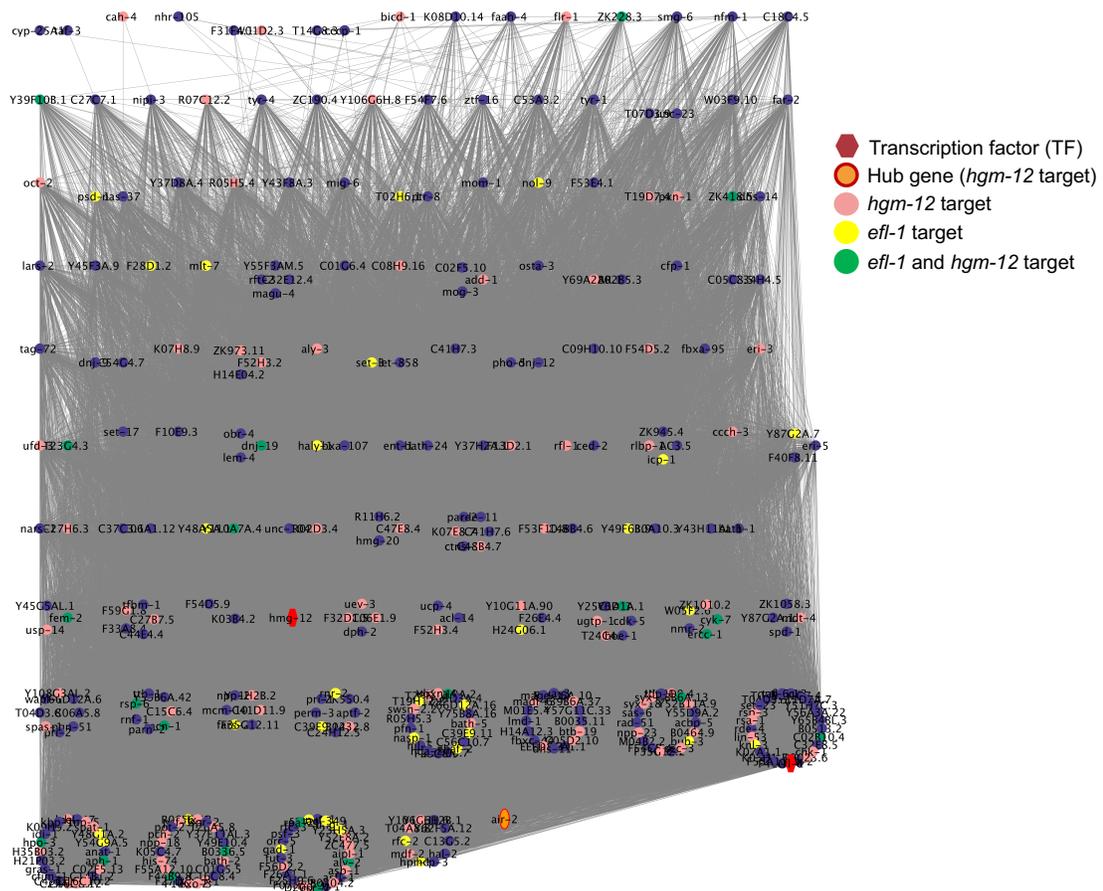


Figure 5.16: Darkslateblue module is correlated with simulated microgravity exposure and glucosamine supplemented diet conditions. The transcription factors *hgm-12* and *efl-1* and their targets are co-expressed in this module.

## 5.4 Discussion

In this study, we conducted a WGCNA analysis by integrating our in-house data to publicly available environmental stress data from the GEO. We identified 50 modules of co-expressed genes which are associated with the environmental stressors we tested. We validated that the modules are enriched for previously determined gene

networks and PPIs and have biologically relevant functions. In addition, we identified the putative drivers of these modules.

Notably, the modules demonstrated enrichment for biologically relevant functions for their associated environmental stressor. For instance, in our previous work, we have not detected any differential expression for the longevity-related genes or pathways in CeHR-grown worms (Çelen et al., 2018). Given that the longevity significantly increases in the CeHR-grown worms, it is expected to observe alterations in the longevity regulating genetic mechanism. Our further investigation with gene co-expression network analysis, we achieved to identify a set of putative genes working collaboratively in longevity promotion in CeHR-grown worms.

The generated modules will have several limitations. To begin with, the gene expression data is from the whole organism. This causes the expression to be an average from different tissues and hence leads to loss of information from lowly expressed genes. Even though the tissue enrichments will be examined, the modules cannot precisely provide tissue-specific information. Next, our RNA-seq data are from different generations and environmental conditions and this can introduce noise into data. The inclusion of the data from other studies can increase the noise even further due to potential batch effects. The modules do not take different transcriptional isoforms into account, which in turn can lead to bias. Together, these factors can confound the network construction and result in increased number of false positives or false negatives.

## 5.5 Methods

### 5.5.1 Data recruitment

In-house data from OP50 NGM, CeHR (four days and three weeks), S-Medium, CeHR reversion, S-Medium reversion-grown, and simulated microgravity-exposed N2 strain worms were generated as described before (Çelen et al., 2017, 2018; Çelen, Jayasinghe, Doh, & Sabanayagam, 2019) and deposited into GEO (<https://www.ncbi.nlm.nih.gov/geo/>) under accession numbers GSE103776 and GSE122125. For publicly available environmental stress data retrieval, GEO was queried with “RNA-seq and *Caenorhabditis elegans*” terms in December 2017. The publicly available datasets were included in the analysis if they meet the following criteria: First, data with triplicates from L4 stage N2 strain grown at 20°C as our data were selected. Second, to minimize the batch effects from sequencing technologies, only RNA-seq data that were generated by Illumina HiSeq 2000 or more recent Illumina technologies were used. Finally, only the experiments testing the effect of environmental changes (e.g., dietary modifications) that can be related to space conditions were included in the analysis. We identified three such studies from the following conditions: glucose supplemented dietary change (GEO: GSE83887) (Ladage et al., 2016), presence and absence of D-Glucosamine (GlcN) (GEO: GSE54853) (Weimer et al., 2014), and presence and absence of vitamin C (GEO: GSE54173) (Dallaire et al., 2014). All the data were analyzed using the Tuxedo pipeline (Trapnell et al., 2012) as described before to ensure uniformity in the data analysis and to be consistent with our previous studies (Çelen et al., 2018, 2019).

### **5.5.2 Data preprocessing**

OP50 NGM-grown worms were used as the control in the publicly available datasets. To avoid the repetitive usage of the data from the same environment, we first confirmed that the data show correlation among the laboratories. That is, we evaluated the Pearson's correlation among the  $\log_2(\text{FPKM})$  values of the RNA-seq data from the OP50 NGM-grown worms in in-house and publicly available data. All but one OP50 NGM condition data demonstrated strong correlation (>86%) while GSE54173 showed moderate correlation (>49%) with the others. To incorporate the OP50 NGM condition data into the co-expression networks, we implemented quantile normalization on the in-house data.

Genes with  $\text{FPKM} < 1$  and the piRNA molecules and were excluded from the analysis. The  $\log_2(\text{FPKM})$  values from 30 RNA-seq libraries were used in the generation of the co-expression networks.

### **5.5.3 Weighted gene co-expression analysis (WGCNA)**

The R package WGCNA was utilized to construct a gene co-expression network with the above processed data (Langfelder & Horvath, 2008). Briefly, a soft threshold power of  $\beta = 16$  was chosen since it is the lowest power for which the scale-free topology fit index reaches 0.8 (Figure 5.17).

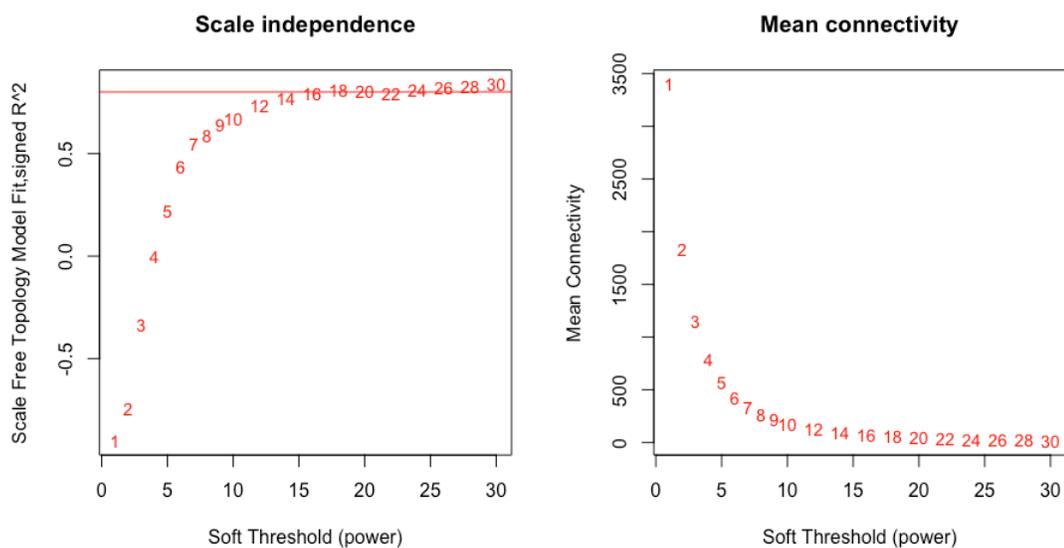


Figure 5.17: Scale free topology model fit and mean connectivity for soft threshold selection.

The adjacency matrix was calculated with Pearson’s correlation and the matrix was transformed into Topological Overlap Matrix (TOM). The corresponding dissimilarity (dissTOM) was calculated by  $1 - \text{TOM}$  and clustered in a dendrogram. The modules for the co-expressed genes were determined by using dynamic tree cut algorithm with a minimum module size of 30, and a representative color was assigned to each module. We then merged the modules showing high correlation (90%) with each other (Figure 5.18).

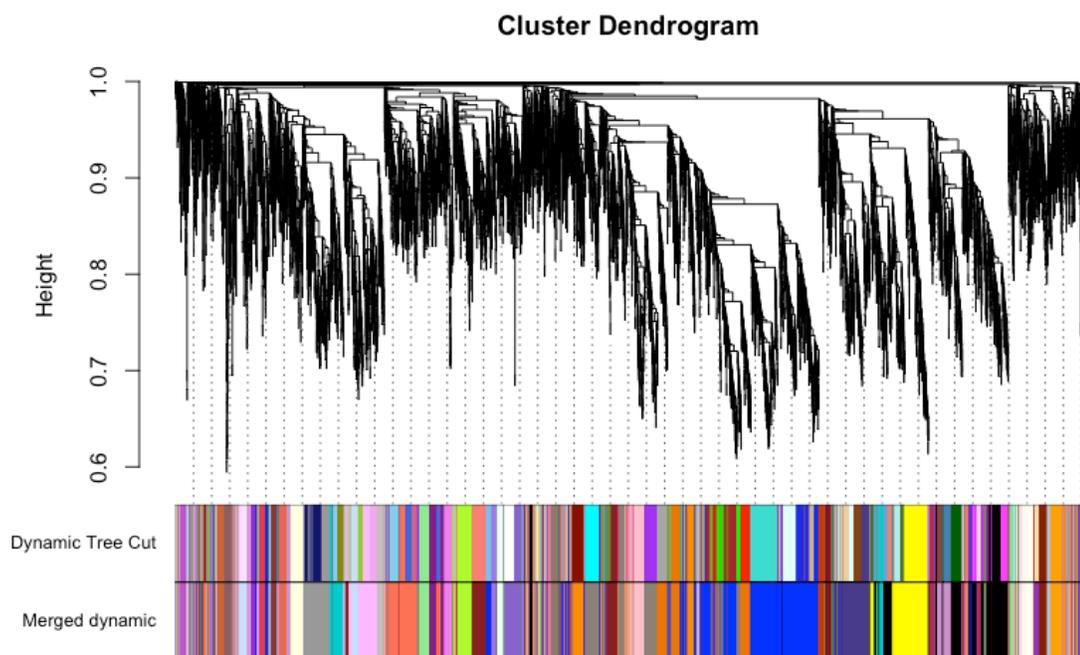


Figure 5.18: Dendrogram of the co-expressed genes for the TOM-based dissimilarity. The colors represent the resulting modules after the dynamic tree cut algorithm and the merge of the highly correlated modules.

To identify the modules that are associated with the tested environmental conditions, we first used one-hot encoding to generate the trait table for the conditions. Then, we correlated the module eigengene, the first principal component, with the conditions to find the important modules for each environmental condition. The  $p$ -value  $< 0.05$  was considered significant.

#### 5.5.4 Network Evaluation

In order to evaluate the performance of the modules for accurately determining the known gene networks, we compared the modules to those from WormNET (v.3) by using the phenotype-centric prediction (Cho et al., 2014). We also used the

STRING database (version 11.0) (Szklarczyk et al., 2017) to assess whether the modules can identify the known PPIs with medium confidence. We only included the results from “experiment”, “database”, and “textmining”. The GO enrichment and pathway analyses for the PPIs were acquired from the STRING database.

### **5.5.5 Network Visualization, Hub Gene and Transcription Factor-Target**

#### **Determination**

We visualized the modules by using Cytoscape (v. 3.7.1) (Shannon et al., 2003) and identified the hub genes, the gene with the greatest number of connectivity, by using the Network Analyzer in Cytoscape.

All the available transcription factors (TF) (n = 54) along with their targets were retrieved from TF2DNA (Pujato et al., 2014) and were mapped to the modules. We then tested whether the targets of the module-associated TFs are overrepresented within the mapped module by using hypergeometric test for the representative modules.

## Chapter 6

### CONCLUSION AND FUTURE DIRECTIONS

In this dissertation, we revealed the individual contribution of simulated microgravity exposure, liquid cultivation, and strain selection into the intergenerational biological responses. We defined a putative gravitome which is conserved in worms and humans through the integration of in-house and publicly available datasets. Our analysis showed that chemically distinct histone modifications demonstrate static or elastic behavior in response to liquid cultivation. Finally, we generated a biologically-relevant gene co-expression network for commonly observed space environmental conditions, such as microgravity and dietary change.

The main limitation of this study is that the ChIP-seq and RNA-seq experiments were performed in populations of whole animals. Therefore, resulting signals are the average of multiple tissues. Another issue is the limitations for comparing the ChIP-seq and RNA-seq data. The fact that these experiments were conducted on different time points can introduce noise to the data and lead to confounding results. Finally, the number of studies with multivalent histone modifications is limited and the existing ones are predominantly from bivalent marks. This issue restricts the evaluation of the model to small number of studies and histone modification combinations. A future study conducting RNA-seq and ChIP-seq

concurrently on synchronized animals can help resolve this issue. To validate the multivalent state of histone modifications, ChIP-reChIP method could be used in future studies (Furlan-Magaril, Rincón-Arano, & Recillas-Targa, 2009). Moreover, a tissue specific study can be conducted to overcome the problems of averaging the data from different tissues.

WGCNA generates undirected networks and does not enable causality inference. Using the module information to generate Bayesian networks would provide useful information about the regulatory gene networks and indicate the direction of gene-gene interactions.

## REFERENCES

- Adenle, A. A., Johnsen, B., & Szewczyk, N. J. (2009). Review of the results from the International *C. elegans* first experiment (ICE-FIRST). *Advances in Space Research*. <https://doi.org/10.1016/j.asr.2009.04.008>
- Adriaens, M. E., Prickaerts, P., Chan-Seng-Yue, M., van den Beucken, T., Dahlmans, V. E. H., Eijssen, L. M., ... Evelo, C. T. A. (2016). Quantitative analysis of ChIP-seq data uncovers dynamic and sustained H3K4me3 and H3K27me3 modulation in cancer cells under hypoxia. *Epigenetics & Chromatin*, *9*, 48. <https://doi.org/10.1186/s13072-016-0090-4>
- Ala, U., Piro, R. M., Grassi, E., Damasco, C., Silengo, L., Oti, M., ... Di Cunto, F. (2008). Prediction of human disease genes by human-mouse conserved coexpression analysis. *PLoS Computational Biology*, *4*(3). <https://doi.org/10.1371/journal.pcbi.1000043>
- Allen, J. D., Xie, Y., Chen, M., Girard, L., & Xiao, G. (2012). Comparing statistical methods for constructing large scale gene networks. *PLoS ONE*, *7*(1). <https://doi.org/10.1371/journal.pone.0029348>
- Alper, S., McBride, S. J., Lackford, B., Freedman, J. H., & Schwartz, D. A. (2007). Specificity and Complexity of the *Caenorhabditis elegans* Innate Immune Response. *Molecular and Cellular Biology*, *27*(15), 5544–5553. <https://doi.org/10.1128/MCB.02070-06>
- Amaral, P. P., & Mattick, J. S. (2008). Noncoding RNA in development. *Mammalian Genome*. <https://doi.org/10.1007/s00335-008-9136-7>
- Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. Available at: [www.bioinformatics.babraham.ac.uk/projects/fastqc/](http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).
- Angeles-Albores, D., N. Lee, R. Y., Chan, J., & Sternberg, P. W. (2016). Tissue enrichment analysis for *C. elegans* genomics. *BMC Bioinformatics*, *17*(1), 366. <https://doi.org/10.1186/s12859-016-1229-9>
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., ... Noble,

- W. S. (2009). MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Research*, 37(SUPPL. 2). <https://doi.org/10.1093/nar/gkp335>
- Barrière, A., Félix, M.-A., Barriere, A., Felix, M.-A., Barrière, A., Félix, M.-A., ... Félix, M.-A. (2005). Natural variation and population genetics of *Caenorhabditis elegans*. *WormBook : The Online Review of C. Elegans Biology*, 1–19. <https://doi.org/10.1895/wormbook.1.43.1>
- Battisti, J. M., Watson, L. A., Naung, M. T., Drobish, A. M., Voronina, E., & Minnick, M. F. (2016). Analysis of the *Caenorhabditis elegans* innate immune response to *Coxiella burnetii*. *Innate Immunity*, 1753425916679255. <https://doi.org/10.1177/1753425916679255>
- Benveniste, D., Sonntag, H.-J., Sanguinetti, G., & Sproul, D. (2014). Transcription factor binding predicts histone modifications in human cell lines. *Proceedings of the National Academy of Sciences of the United States of America*, 111(37), 13367–13372. <https://doi.org/10.1073/pnas.1412081111>
- Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., ... Lander, E. S. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, 125(2), 315–326. <https://doi.org/10.1016/j.cell.2006.02.041>
- Bismuth, J., Lin, P., Yao, Q., & Chen, C. (2008). Ceramide: A common pathway for atherosclerosis? *Atherosclerosis*. <https://doi.org/10.1016/j.atherosclerosis.2007.09.018>
- Blazie, S. M., Babb, C., Wilky, H., Rawls, A., Park, J. G., & Mangone, M. (2015). Comparative RNA-Seq analysis reveals pervasive tissue-specific alternative polyadenylation in *Caenorhabditis elegans* intestine and muscles. *BMC Biology*, 13(1). <https://doi.org/10.1186/s12915-015-0116-6>
- Boettcher, M., & McManus, M. T. (2015). Choosing the Right Tool for the Job: RNAi, TALEN, or CRISPR. *Molecular Cell*. <https://doi.org/10.1016/j.molcel.2015.04.028>
- Boros, J., Arnoult, N., Stroobant, V., Collet, J.-F., & Decottignies, A. (2014). Polycomb repressive complex 2 and H3K27me3 cooperate with H3K9 methylation to maintain heterochromatin protein 1 $\alpha$  at chromatin. *Molecular and Cellular Biology*, 34(19), 3662–3674. <https://doi.org/10.1128/MCB.00205-14>
- Brenner, S. (1974). The genetics of *Caenorhabditis elegans*. *Genetics*, 77(1), 71–94. <https://doi.org/10.1002/cbic.200300625>

- Briegleb, W. (1992). Some qualitative and quantitative aspects of the fast-rotating clinostat as a research tool. *ASGSB Bulletin*, 5(2), 23–30.
- Brinkman, A. B., Roelofsen, T., Pennings, S. W. C., Martens, J. H. A., Jenuwein, T., & Stunnenberg, H. G. (2006). Histone modification patterns associated with the human X chromosome. *EMBO Reports*, 7(6), 628–634. <https://doi.org/10.1038/sj.embor.7400686>
- Brown, A. H., Dahl, A. O., & Chapman, D. K. (1976). Limitation on the Use of the Horizontal Clinostat as a Gravity Compensator. *PLANT PHYSIOLOGY*, 58(2), 127–130. <https://doi.org/10.1104/pp.58.2.127>
- Brückner, A., Polge, C., Lentze, N., Auerbach, D., & Schlattner, U. (2009). Yeast two-hybrid, a powerful tool for systems biology. *International Journal of Molecular Sciences*. <https://doi.org/10.3390/ijms10062763>
- Buske, F. A., Bodén, M., Bauer, D. C., & Bailey, T. L. (2010). Assigning roles to DNA regulatory motifs using comparative genomics. *Bioinformatics*, 26(7), 860–866. <https://doi.org/10.1093/bioinformatics/btq049>
- Cao, J. (2014). The functional role of long non-coding RNAs and epigenetics. *Biological Procedures Online*, 16(1), 11. <https://doi.org/10.1186/1480-9222-16-11>
- Cao, J., Packer, J. S., Ramani, V., Cusanovich, D. A., Huynh, C., Daza, R., ... Shendure, J. (2017). Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science*, 357(6352), 661–667. <https://doi.org/10.1126/science.aam8940>
- Capra, E. J., Skrovanek, S. M., & Kruglyak, L. (2008). Comparative developmental expression profiling of two *C. elegans* isolates. *PLoS ONE*, 3(12), e4055. <https://doi.org/10.1371/journal.pone.0004055>
- Carlson, M. R., Zhang, B., Fang, Z., Mischel, P. S., Horvath, S., & Nelson, S. F. (2006). Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks. *BMC Genomics*, 7, 40. <https://doi.org/10.1186/1471-2164-7-40>
- Carthew, R. W., & Sontheimer, E. J. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell*. <https://doi.org/10.1016/j.cell.2009.01.035>
- Castelein, N., Hoogewijs, D., De Vreese, A., Braeckman, B. P., & Vanfleteren, J. R. (2008). Dietary restriction by growth in axenic medium induces discrete changes in the transcriptional output of genes involved in energy metabolism in

- Caenorhabditis elegans. *Biotechnology Journal*, 3(6), 803–812.  
<https://doi.org/10.1002/biot.200800003>
- Çelen, İ., Doh, J. H., & Sabanayagam, C. (2017). Genetic Adaptation of *C. elegans* to Environment Changes I: Multigenerational Analysis of the Transcriptome. *Doi.Org*, 194506. <https://doi.org/10.1101/194506>
- Çelen, İ., Doh, J. H., & Sabanayagam, C. R. (2018). Effects of liquid cultivation on gene expression and phenotype of *C. elegans*. *BMC Genomics*, 19(1), 562. <https://doi.org/10.1186/s12864-018-4948-7>
- Çelen, İ., Jayasinghe, A., Doh, J., & Sabanayagam, C. R. (2019). Comparative Transcriptomic Signature of the Simulated Microgravity Response in *Caenorhabditis elegans*. *BioRxiv*, 531335. <https://doi.org/10.1101/531335>
- Çelen, İ., Ross, K. E., Arighi, C. N., & Wu, C. H. (2015). Bioinformatics Knowledge Map for Analysis of Beta-Catenin Function in Cancer. *PloS One*, 10(10), e0141773. <https://doi.org/10.1371/journal.pone.0141773>
- Chan, J. P., Brown, J., Hark, B., Nolan, A., Servello, D., Hrobuchak, H., & Staab, T. A. (2017). Loss of sphingosine kinase alters life history traits and locomotor function in *Caenorhabditis elegans*. *Frontiers in Genetics*, 8(SEP). <https://doi.org/10.3389/fgene.2017.00132>
- Chinta, S. J., Lieu, C. A., Demaria, M., Laberge, R. M., Campisi, J., & Andersen, J. K. (2013). Environmental stress, ageing and glial cell senescence: A novel mechanistic link to parkinson's disease? *Journal of Internal Medicine*, 273(5), 429–436. <https://doi.org/10.1111/joim.12029>
- Cho, A., Shin, J., Hwang, S., Kim, C., Shim, H., Kim, H., ... Lee, I. (2014). WormNet v3: A network-assisted hypothesis-generating server for *Caenorhabditis elegans*. *Nucleic Acids Research*, 42(W1). <https://doi.org/10.1093/nar/gku367>
- Choi, H., Kim, G.-J., Yoo, H.-S., Song, D., Chung, K.-H., Lee, K.-J., ... An, J. (2019). Vitamin C Activates Osteoblastogenesis and Inhibits Osteoclastogenesis via Wnt/ $\beta$ -Catenin/ATF4 Signaling Pathways. *Nutrients*, 11(3), 506. <https://doi.org/10.3390/nu11030506>
- Chou, W.-C., Cheng, A.-L., Brotto, M., & Chuang, C.-Y. (2014). Visual gene-network analysis reveals the cancer gene co-expression in human endometrial cancer. *BMC Genomics*, 15(1), 300. <https://doi.org/10.1186/1471-2164-15-300>
- Collin, R., & Cipriani, R. (2003). Dollo's law and the re-evolution of shell coiling. *Proceedings of the Royal Society B: Biological Sciences*, 270(1533), 2551–2555.

<https://doi.org/10.1098/rspb.2003.2517>

- Croll, N. A., Smith, J. M., & Zuckerman, B. M. (1977). The aging process of the nematode *Caenorhabditis elegans* in bacterial and axenic culture. *Experimental Aging Research*, 3(3), 175–189. <https://doi.org/10.1080/03610737708257101>
- Crucian, B. E., Choukèr, A., Simpson, R. J., Mehta, S., Marshall, G., Smith, S. M., ... Sams, C. (2018). Immune system dysregulation during spaceflight: Potential countermeasures for deep space exploration missions. *Frontiers in Immunology*. <https://doi.org/10.3389/fimmu.2018.01437>
- Cui, M., Wang, Y., Cavaleri, J., Kelson, T., Teng, Y., & Han, M. (2017). Starvation-induced stress response is critically impacted by ceramide levels in *Caenorhabditis elegans*. *Genetics*, 205(2), 775–785. <https://doi.org/10.1534/genetics.116.194282>
- Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., ... Flicek, P. (2014). Ensembl 2015. *Nucleic Acids Research*, 43(D1), D662–669. <https://doi.org/10.1093/nar/gku1010>
- Cutler, R. G., Pedersen, W. A., Camandola, S., Rothstein, J. D., & Mattson, M. P. (2002). Evidence that accumulation of ceramides and cholesterol esters mediates oxidative stress - Induced death of motor neurons in amyotrophic lateral sclerosis. *Annals of Neurology*, 52(4), 448–457. <https://doi.org/10.1002/ana.10312>
- Cutler, R. G., Thompson, K. W., Camandola, S., Mack, K. T., & Mattson, M. P. (2014). Sphingolipid metabolism regulates development and lifespan in *Caenorhabditis elegans*. *Mechanisms of Ageing and Development*, 143–144, 9–18. <https://doi.org/10.1016/j.mad.2014.11.002>
- D'Urso, A., & Brickner, J. H. (2014). Mechanisms of epigenetic memory. *Trends in Genetics : TIG*, 30(6), 230–236. <https://doi.org/10.1016/j.tig.2014.04.004>
- Dallaire, A., Proulx, S., Simard, M. J., & Lebel, M. (2014). Expression profile of *Caenorhabditis elegans* mutant for the Werner syndrome gene ortholog reveals the impact of vitamin C on development to increase life span. *BMC Genomics*, 15(1). <https://doi.org/10.1186/1471-2164-15-940>
- Daxinger, L., & Whitelaw, E. (2010). Transgenerational epigenetic inheritance: more questions than answers. *Genome Research*, 20(12), 1623–1628. <https://doi.org/10.1101/gr.106138.110>
- Dedolph, R. R., & Dipert, M. H. (1971). The physical basis of gravity stimulus

- nullification by clinostat rotation. *Plant Physiology*, 47(6), 756–764.
- Dedolph, R. R., Oemick, D. a, Wilson, B. R., & Smith, G. R. (1967). Causal basis of gravity stimulus nullification by clinostat rotation. *Plant Physiology*, 42(10), 1373–1383.
- Deng, X., Yin, X., Allan, R., Lu, D. D., Maurer, C. W., Haimovitz-Friedman, A., ... Kolesnick, R. (2008). Ceramide biogenesis is required for radiation-induced apoptosis in the germ line of *C. elegans*. *Science*, 322(5898), 110–115. <https://doi.org/10.1126/science.1158111>
- Denver, D. R., Morris, K., Strelman, J. T., Kim, S. K., Lynch, M., & Thomas, W. K. (2005). The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nature Genetics*, 37(5), 544–548. <https://doi.org/10.1038/ng1554>
- Denver, D. R., Morris, K., & Thomas, W. K. (2003). Phylogenetics in *Caenorhabditis elegans*: An analysis of divergence and outcrossing. *Molecular Biology and Evolution*, 20(3), 393–400. <https://doi.org/10.1093/molbev/msg044>
- Doh, J. H., Moore, A. B., Çelen, İ., Moore, M. T., & Sabanayagam, C. R. (2016). ChIP and Chips : Introducing the WormPharm for correlative studies employing pharmacology and genome-wide analyses in *C. elegans*. *Journal of Biological Methods*, 3(2), 1–11. <https://doi.org/10.14440/jbm.2016.109>
- Duerr, J. S. (2006). Immunohistochemistry. *WormBook : The Online Review of C. Elegans Biology*, 1–61. <https://doi.org/10.1895/wormbook.1.105.1>
- Duff, M. O., Olson, S., Wei, X., Garrett, S. C., Osman, A., Bolisetty, M., ... Graveley, B. R. (2015). Genome-wide identification of zero nucleotide recursive splicing in *Drosophila*. *Nature*, 521(7552), 376–379. <https://doi.org/10.1038/nature14475>
- Edwards, M., & Abadie, L. (2018). NASA Twins Study Investigators to Release Integrated Paper in 2018. Retrieved October 9, 2018, from <https://www.nasa.gov/feature/nasa-twins-study-investigators-to-release-integrated-paper-in-2018>
- Félix, M.-A., & Braendle, C. (2010). The natural history of *Caenorhabditis elegans*. *Current Biology*, 20(22), R965–R969. <https://doi.org/10.1016/j.cub.2010.09.050>
- Ferraro, T., Esposito, E., Mancini, L., Ng, S., Lucas, T., Coppey, M., ... Lagha, M. (2016). Transcriptional Memory in the *Drosophila* Embryo. *Current Biology : CB*, 26(2), 212–218. <https://doi.org/10.1016/j.cub.2015.11.058>

- Friedman, N., Linial, M., Nachman, I., & Pe'er, D. (2000). Using Bayesian networks to analyze expression data. *Journal of Computational Biology : A Journal of Computational Molecular Cell Biology*, 7(3–4), 601–620. <https://doi.org/10.1089/106652700750050961>
- Fuller, T. F., Ghazalpour, A., Aten, J. E., Drake, T. a, Lusk, A. J., & Horvath, S. (2007). Weighted gene coexpression network analysis strategies applied to mouse weight. *Mammalian Genome : Official Journal of the International Mammalian Genome Society*, 18(6–7), 463–472. <https://doi.org/10.1007/s00335-007-9043-3>
- Fuller, T., Langfelder, P., Presson, A., & Horvath, S. (2011). Review of Weighted Gene Coexpression Network Analysis. In *Handbook of Statistical Bioinformatics* (pp. 369–388). [https://doi.org/10.1007/978-3-642-16345-6\\_18](https://doi.org/10.1007/978-3-642-16345-6_18)
- Furlan-Magaril, M., Rincón-Arango, H., & Recillas-Targa, F. (2009). Sequential chromatin immunoprecipitation protocol: ChIP-reChIP. *Methods in Molecular Biology*. [https://doi.org/10.1007/978-1-60327-015-1\\_17](https://doi.org/10.1007/978-1-60327-015-1_17)
- Gao, Y., Arfat, Y., Wang, H., & Goswami, N. (2018). Muscle atrophy induced by mechanical unloading: Mechanisms and potential countermeasures. *Frontiers in Physiology*. <https://doi.org/10.3389/fphys.2018.00235>
- Gaydos, L. J., Wang, W., & Strome, S. (2014). H3K27me and PRC2 transmit a memory of repression across generations and during development. *Science*, 345(6203), 1515–1518. <https://doi.org/10.1126/science.1255023>
- Gems, D., & Riddle, D. L. (2000). Genetic, behavioral and environmental determinants of male longevity in *Caenorhabditis elegans*. *Genetics*, 154(4), 1597–1610. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10747056>
- Gerstein, M. B., Lu, Z. J., Van Nostrand, E. L., Cheng, C., Arshinoff, B. I., Liu, T., ... Waterston, R. H. (2010). Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science (New York, N.Y.)*, 330(6012), 1775–1787. <https://doi.org/10.1126/science.1196914>
- Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: Scanning for occurrences of a given motif. *Bioinformatics*, 27(7), 1017–1018. <https://doi.org/10.1093/bioinformatics/btr064>
- Graveley, A., Vlasov, A., Freeman, A., Wu, K., Szewczyk, N. J., D’Cruz, R., & Batt, J. (2018). Levels of acid sphingomyelinase (ASM) in *Caenorhabditis elegans* in microgravity. *Gravitational and Space Research*, 6(1)(1), 27–36. Retrieved from <http://gravitationalandspacebiology.org/index.php/journal/article/viewFile/803/809>

- Greenspan, P., Mayer, E. P., & Fowler, S. D. (1985). Nile red: A selective fluorescent stain for intracellular lipid droplets. *Journal of Cell Biology*, *100*(3), 965–973. <https://doi.org/10.1083/jcb.100.3.965>
- Greer, E. L., Maures, T. J., Ucar, D., Hauswirth, A. G., Mancini, E., Lim, J. P., ... Brunet, A. (2011). Transgenerational epigenetic inheritance of longevity in *Caenorhabditis elegans* - nature10572.pdf. *Nature*, *479*(7373), 365–371. <https://doi.org/10.1038/nature10572>
- Gruener, R., & Hoeger, G. (1991). Vector-averaged gravity alters myocyte and neuron properties in cell culture. *Aviation, Space, and Environmental Medicine*, *62*(12), 1159–1165.
- Gunsalus, K. C., Ge, H., Schetter, A. J., Goldberg, D. S., Han, J. D. J., Hao, T., ... Piano, F. (2005). Predictive models of molecular machines involved in *Caenorhabditis elegans* early embryogenesis. *Nature*, *436*(7052), 861–865. <https://doi.org/10.1038/nature03876>
- Haber, M., Schüngel, M., Putz, A., Müller, S., Hasert, B., & Schulenburg, H. (2005). 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), 160–173. <https://doi.org/10.1093/molbev/msh264>
- Häder, D.-P., Hemmersbach, R., & Lebert, M. (2005). *Gravity and the Behavior of Unicellular Organisms*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511546211>
- Harr, J. C., Gonzalez-Sandoval, A., & Gasser, S. M. (2016). Histones and histone modifications in perinuclear chromatin anchoring: from yeast to man. *EMBO Reports*, *17*(2), 139–155. <https://doi.org/10.15252/embr.201541809>
- Harris, T. W., Baran, J., Bieri, T., Cabunoc, A., Chan, J., Chen, W. J., ... Sternberg, P. W. (2014). WormBase 2014: new views of curated biology. *Nucleic Acids Research*, *42*(Database issue), D789-93. <https://doi.org/10.1093/nar/gkt1063>
- Heard, E., & Martienssen, R. A. (2014). Transgenerational epigenetic inheritance: myths and mechanisms. *Cell*, *157*(1), 95–109. <https://doi.org/10.1016/j.cell.2014.02.045>
- Heidinger, B. J., Blount, J. D., Boner, W., Griffiths, K., Metcalfe, N. B., & Monaghan, P. (2012). Telomere length in early life predicts lifespan. *Proceedings of the National Academy of Sciences*, *109*(5), 1743–1748. <https://doi.org/10.1073/pnas.1113306109>

- Hemmersbach, R., Volkmann, D., & Häder, D.-P. (1999). Graviorientation in Protists and Plants. *Journal of Plant Physiology*, *154*(1), 1–15. [https://doi.org/10.1016/S0176-1617\(99\)80311-3](https://doi.org/10.1016/S0176-1617(99)80311-3)
- Hernández-Corbacho, M. J., Jenkins, R. W., Clarke, C. J., Hannun, Y. A., Obeid, L. M., Snider, A. J., & Siskind, L. J. (2011). Accumulation of long-chain glycosphingolipids during aging is prevented by caloric restriction. *PLoS ONE*, *6*(6). <https://doi.org/10.1371/journal.pone.0020411>
- Hibshman, J. D., Hung, A., & Baugh, L. R. (2016). Maternal Diet and Insulin-Like Signaling Control Intergenerational Plasticity of Progeny Size and Starvation Resistance. *PLoS Genetics*, *12*(10). <https://doi.org/10.1371/journal.pgen.1006396>
- Higashibata, A., Hashizume, T., Nemoto, K., Higashitani, N., Etheridge, T., Mori, C., ... Higashitani, A. (2016). Microgravity elicits reproducible alterations in cytoskeletal and metabolic gene and protein expression in space-flown *Caenorhabditis elegans*. *Npj Microgravity*, *2*(1), 15022. <https://doi.org/10.1038/npjmgrav.2015.22>
- Higashibata, A., Szewczyk, N. J., Conley, C. A., Imamizo-Sato, M., Higashitani, A., & Ishioka, N. (2006). Decreased expression of myogenic transcription factors and myosin heavy chains in *Caenorhabditis elegans* muscles developed during spaceflight. *The Journal of Experimental Biology*, *209*(Pt 16), 3209–3218. <https://doi.org/10.1242/jeb.02365>
- Higashitani, A., Hashizume, T., Sugimoto, T., Mori, C., Nemoto, K., Etheridge, T., ... Higashibata, A. (2009). *C. elegans* RNAi space experiment (CERISE) in Japanese Experiment Module KIBO. *Biological Sciences in Space*, *23*(4), 183–187. <https://doi.org/10.2187/bss.23.183>
- Hillier, L. W., Coulson, A., Murray, J. I., Bao, Z., Sulston, J. E., & Waterston, R. H. (2005). Genomics in *C. elegans*: so many genes, such a little worm. *Genome Research*, *15*(12), 1651–1660. <https://doi.org/10.1101/gr.3729105>
- Ho, J. W. W. K., Bishop, E., Karchenko, P. V., Nègre, N., White, K. P., & Park, P. J. (2011). ChIP-chip versus ChIP-seq: Lessons for experimental design and data analysis. *BMC Genomics*, *12*(1), 134. <https://doi.org/10.1186/1471-2164-12-134>
- Honda, Y., Higashibata, A., Matsunaga, Y., Yonezawa, Y., Kawano, T., Higashitani, A., ... Garigan, D. (2012). Genes down-regulated in spaceflight are involved in the control of longevity in *Caenorhabditis elegans*. *Scientific Reports*, *2*, 413–429. <https://doi.org/10.1038/srep00487>
- Honda, Y., Honda, S., Narici, M., & Szewczyk, N. J. (2014a). Spaceflight and Ageing:

- Reflecting on *Caenorhabditis elegans* in Space. *Gerontology*, 60(2), 138–142.  
<https://doi.org/10.1159/000354772>
- Honda, Y., Honda, S., Narici, M., & Szewczyk, N. J. (2014b). Spaceflight and ageing: reflecting on *Caenorhabditis elegans* in space. *Gerontology*, 60(2), 138–142.  
<https://doi.org/10.1159/000354772>
- Hoogewijs, D., Houthoofd, K., Matthijssens, F., Vandesompele, J., & Vanfleteren, J. R. (2008). Selection and validation of a set of reliable reference genes for quantitative sod gene expression analysis in *C. elegans*. *BMC Molecular Biology*, 9(1), 9. <https://doi.org/10.1186/1471-2199-9-9>
- Houthoofd, K., Braeckman, B. P., Lenaerts, I., Brys, K., De Vreese, A., Van Eygen, S., & Vanfleteren, J. R. (2002a). Axenic growth up-regulates mass-specific metabolic rate, stress resistance, and extends life span in *Caenorhabditis elegans*. *Experimental Gerontology*, 37(12), 1371–1378. [https://doi.org/10.1016/S0531-5565\(02\)00173-0](https://doi.org/10.1016/S0531-5565(02)00173-0)
- Houthoofd, K., Braeckman, B. P., Lenaerts, I., Brys, K., De Vreese, A., Van Eygen, S., & Vanfleteren, J. R. (2002b). No reduction of metabolic rate in food restricted *Caenorhabditis elegans*. *Experimental Gerontology*, 37(12), 1359–1369.  
 Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12559405>
- Howe, F. S., Fischl, H., Murray, S. C., & Mellor, J. (2017). Is H3K4me3 instructive for transcription activation? *BioEssays*, 39(1), 1600095.  
<https://doi.org/10.1002/bies.201600095>
- Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1), 44–57. <https://doi.org/10.1038/nprot.2008.211>
- Huang, H.-S., Allen, J. A., Mabb, A. M., King, I. F., Miriyala, J., Taylor-Blake, B., ... Philpot, B. D. (2011). Topoisomerase inhibitors unsilence the dormant allele of *Ube3a* in neurons. *Nature*, 481(7380), 185–189.  
<https://doi.org/10.1038/nature10726>
- Huang, X. A., Yin, H., Sweeney, S., Raha, D., Snyder, M., & Lin, H. (2013). A Major Epigenetic Programming Mechanism Guided by piRNAs. *Developmental Cell*, 24(5), 502–516. <https://doi.org/10.1016/j.devcel.2013.01.023>
- Huang, X., Withers, B. R., & Dickson, R. C. Sphingolipids and lifespan regulation, 1841 *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* § (2014). <https://doi.org/10.1016/j.bbalip.2013.08.006>

- Hughes, T. P., Kerry, J. T., Álvarez-Noriega, M., Álvarez-Romero, J. G., Anderson, K. D., Baird, A. H., ... Wilson, S. K. (2017). Global warming and recurrent mass bleaching of corals. *Nature*, *543*(7645), 373–377. <https://doi.org/10.1038/nature21707>
- Hunt-Newbury, R., Viveiros, R., Johnsen, R., Mah, A., Anastas, D., Fang, L., ... Moerman, D. G. (2007). High-throughput in vivo analysis of gene expression in *Caenorhabditis elegans*. *PLoS Biology*, *5*(9), 1981–1997. <https://doi.org/10.1371/journal.pbio.0050237>
- Jaenisch, R., & Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics*, *33* Suppl(march), 245–254. <https://doi.org/10.1038/ng1089>
- Jamialahmadi, K., Soltani, F., Nabavi fard, M., Behravan, J., & Mosaffa, F. (2014). Assessment of protective effects of glucosamine and N-acetyl glucosamine against DNA damage induced by hydrogen peroxide in human lymphocytes. *Drug and Chemical Toxicology*, *37*(4), 427–432. <https://doi.org/10.3109/01480545.2013.878951>
- Jenuwein, T., & Allis, C. D. (2001). Translating the histone code. *Science (New York, N.Y.)*, *293*(5532), 1074–1080. <https://doi.org/10.1126/science.1063127>
- Joeng, K. S., Song, E. J., Lee, K. J., & Lee, J. (2004). Long lifespan in worms with long telomeric DNA. *Nature Genetics*, *36*(6), 607–611. <https://doi.org/10.1038/ng1356>
- Jørgensen, S., Schotta, G., & Sørensen, C. S. (2013). Histone H4 lysine 20 methylation: key player in epigenetic regulation of genomic integrity. *Nucleic Acids Research*, *41*(5), 2797–2806. <https://doi.org/10.1093/nar/gkt012>
- Kaikkonen, M. U., Lam, M. T. Y., & Glass, C. K. (2011). Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovascular Research*. <https://doi.org/10.1093/cvr/cvr097>
- Kalinava, N., Ni, J. Z., Peterman, K., Chen, E., & Gu, S. G. (2017). Decoupling the downstream effects of germline nuclear RNAi reveals that H3K9me3 is dispensable for heritable RNAi and the maintenance of endogenous siRNA-mediated transcriptional silencing in *Caenorhabditis elegans*. *Epigenetics & Chromatin*, *10*(1), 6. <https://doi.org/10.1186/s13072-017-0114-8>
- Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., ... Ahringer, J. (2003). Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature*, *421*(6920), 231–237.

<https://doi.org/10.1038/nature01278>

- Kang, C., & Avery, L. (2010). Death-associated protein kinase (DAPK) and signal transduction: Fine-tuning of autophagy in *Caenorhabditis elegans* homeostasis. *FEBS Journal*. <https://doi.org/10.1111/j.1742-4658.2009.07413.x>
- Kelly, W. G. (2014). Transgenerational epigenetics in the germline cycle of *Caenorhabditis elegans*. *Epigenetics & Chromatin*, 7(1), 6. <https://doi.org/10.1186/1756-8935-7-6>
- Kenyon, C., Sulston, J., & Hodgkin, J. (1988). The nematode *Caenorhabditis elegans*. *Science*, 240(4858), 1448–1453. <https://doi.org/10.1126/science.3287621>
- Kessler, J. (1992). The internal dynamics of slowly rotating biological systems. *ASGSB Bulletin*, 5(2), 11–21.
- Kim, B., Suo, B., & Emmons, S. W. (2016). Gene Function Prediction Based on Developmental Transcriptomes of the Two Sexes in *C. elegans*. *Cell Reports*, 17(3), 917–928. <https://doi.org/10.1016/j.celrep.2016.09.051>
- Kim, W., Underwood, R. S., Greenwald, I., & Shaye, D. D. (2018). OrthoList 2: A New Comparative Genomic Analysis of Human and *Caenorhabditis elegans* Genes. *Genetics*, genetics.301307.2018. <https://doi.org/10.1534/genetics.118.301307>
- Kim, Y., & Sun, H. (2012). ASM-3 Acid Sphingomyelinase Functions as a Positive Regulator of the DAF-2/AGE-1 Signaling Pathway and Serves as a Novel Anti-Aging Target. *PLoS ONE*, 7(9), e45890. <https://doi.org/10.1371/journal.pone.0045890>
- Kishore, S. (2006). The snoRNA HBII-52 Regulates Alternative Splicing of the Serotonin Receptor 2C. *Science*, 311(5758), 230–232. <https://doi.org/10.1126/science.1118265>
- Kramer, M., Kranz, A. L., Su, A., Winterkorn, L. H., Albritton, S. E., & Ercan, S. (2015). Developmental Dynamics of X-Chromosome Dosage Compensation by the DCC and H4K20me1 in *C. elegans*. *PLoS Genetics*, 11(12). <https://doi.org/10.1371/journal.pgen.1005698>
- Ladage, M. L., King, S. D., Burks, D. J., Quan, D. L., Garcia, A. M., Azad, R. K., & Padilla, P. A. (2016). Glucose or Altered Ceramide Biosynthesis Mediate Oxygen Deprivation Sensitivity Through Novel Pathways Revealed by Transcriptome Analysis in *Caenorhabditis elegans*. *G3: Genes, Genomes, Genetics*, 6(10), 3149–3160. <https://doi.org/10.1534/g3.116.031583>

- Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9(1), 559. <https://doi.org/10.1186/1471-2105-9-559>
- Langfelder, P., Mischel, P. S., & Horvath, S. (2013). When Is Hub Gene Selection Better than Standard Meta-Analysis? *PLoS ONE*, 8(4). <https://doi.org/10.1371/journal.pone.0061505>
- Lee, S., Lee, J., Chae, S., Moon, Y., Lee, H.-Y., Park, B., ... Park, H. (2017). Multi-dimensional histone methylations for coordinated regulation of gene expression under hypoxia. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkx747>
- Lee, S. S., Kennedy, S., Tolonen, A. C., & Ruvkun, G. (2003). DAF-16 target genes that control *C. elegans* Life-span and metabolism. *Science*, 300(5619), 644–647. <https://doi.org/10.1126/science.1083614>
- Lee Silver, I. (1976). The dynamics of a discrete geotropic sensor subject to rotation-induced gravity compensation. *Journal of Theoretical Biology*, 61(2), 353–362. [https://doi.org/10.1016/0022-5193\(76\)90023-0](https://doi.org/10.1016/0022-5193(76)90023-0)
- Lehner, B., Crombie, C., Tischler, J., Fortunato, A., & Fraser, A. G. (2006). Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. *Nature Genetics*, 38(8), 896–903. <https://doi.org/10.1038/ng1844>
- Levine, B., & Kroemer, G. (2008). Autophagy in the Pathogenesis of Disease. *Cell*. <https://doi.org/10.1016/j.cell.2007.12.018>
- Li, C., & Kim, K. (2008). Neuropeptides. *WormBook : The Online Review of C. Elegans Biology*, (212), 1–36. <https://doi.org/10.1895/wormbook.1.142.1>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, N., An, L., & Hang, H. (2015). Increased sensitivity of DNA damage response-deficient cells to stimulated microgravity-induced DNA lesions. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0125236>
- Lichtenstein, P., Holm, N. V., Verkasalo, P. K., Iliadou, A., Kaprio, J., Koskenvuo, M., ... Hemminki, K. (2000). Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *The New England Journal of Medicine*, 343(2), 78–85. <https://doi.org/10.1056/NEJM200007133430201>

- Light, W. H., & Brickner, J. H. (2013). Nuclear pore proteins regulate chromatin structure and transcriptional memory by a conserved mechanism. *Nucleus*, *4*(5), 357–360. <https://doi.org/10.4161/nucl.26209>
- Lin, K., Hsin, H., Libina, N., & Kenyon, C. (2001). Regulation of the *Caenorhabditis elegans* longevity protein DAF-16 by insulin/IGF-1 and germline signaling. *Nature Genetics*, *28*(2), 139–145. <https://doi.org/10.1038/88850>
- Lin, X. X., Sen, I., Janssens, G. E., Zhou, X., Fonslow, B. R., Edgar, D., ... Riedel, C. G. (2018). DAF-16/FOXO and HLH-30/TFEB function as combinatorial transcription factors to promote stress resistance and longevity. *Nature Communications*. <https://doi.org/10.1038/s41467-018-06624-0>
- Liu, T., Rechtsteiner, A., Egelhofer, T. A., Vielle, A., Latorre, I., Cheung, M. S., ... Liu, X. S. (2011). Broad chromosomal domains of histone modification patterns in *C. elegans*. *Genome Research*, *21*(2), 227–236. <https://doi.org/10.1101/gr.115519.110>
- López-Maury, L., Marguerat, S., & Bähler, J. (2008). Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nature Reviews Genetics*, *9*(8), 583–593. <https://doi.org/10.1038/nrg2398>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12). <https://doi.org/10.1186/s13059-014-0550-8>
- Lumey, L. H., Khalangot, M. D., & Vaiserman, A. M. (2015). Association between type 2 diabetes and prenatal exposure to the Ukraine famine of 1932–33: a retrospective cohort study. *The Lancet Diabetes & Endocrinology*, *3*(10), 787–794. [https://doi.org/10.1016/S2213-8587\(15\)00279-X](https://doi.org/10.1016/S2213-8587(15)00279-X)
- Lv, C., Wang, L., Zhu, X., Lin, W., Chen, X., Huang, Z., ... Yang, S. (2018). Glucosamine promotes osteoblast proliferation by modulating autophagy via the mammalian target of rapamycin pathway. *Biomedicine and Pharmacotherapy*. <https://doi.org/10.1016/j.biopha.2018.01.066>
- Maere, S., Heymans, K., & Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics (Oxford, England)*, *21*(16), 3448–3449. <https://doi.org/10.1093/bioinformatics/bti551>
- Matsumura, Y., Nakaki, R., Inagaki, T., Yoshida, A., Kano, Y., Kimura, H., ... Sakai, J. (2015). H3K4/H3K9me3 Bivalent Chromatin Domains Targeted by Lineage-Specific DNA Methylation Pauses Adipocyte Differentiation. *Molecular Cell*,

60(4), 584–596. <https://doi.org/10.1016/j.molcel.2015.10.025>

- McGrath, P. T., Rockman, M. V., Zimmer, M., Jang, H., Macosko, E. Z., Kruglyak, L., & Bargmann, C. I. (2009). Quantitative Mapping of a Digenic Behavioral Trait Implicates Globin Variation in *C. elegans* Sensory Behaviors. *Neuron*, 61(5), 692–699. <https://doi.org/10.1016/j.neuron.2009.02.012>
- Mckay, S. J., Johnsen, R., Khattra, J., Asano, J., Baillie, D. L., Chan, S., ... Moerman, D. G. (2003). Gene Expression Profiling of Cells, Tissues, and Developmental Stages of the Nematode *C. elegans*. *Cold Spring Harbor Symposia on Quantitative Biology*, 68(0), 159–170. <https://doi.org/10.1101/sqb.2003.68.159>
- Mehta, S. R., Tom, C. M., Wang, Y., Bresee, C., Rushton, D., Mathkar, P. P., ... Mattis, V. B. (2018). Human Huntington's Disease iPSC-Derived Cortical Neurons Display Altered Transcriptomics, Morphology, and Maturation. *Cell Reports*. <https://doi.org/10.1016/j.celrep.2018.09.076>
- Menuz, V., Howell, K. S., Gentina, S., Epstein, S., Riezman, I., Fornallaz-Mulhauser, M., ... Martinou, J. C. (2009). Protection of *C. elegans* from Anoxia by HYL-2 ceramide synthase. *Science*, 324(5925), 381–384. <https://doi.org/10.1126/science.1168532>
- Mohd-Sarip, A., van der Knaap, J. A., Wyman, C., Kanaar, R., Schedl, P., & Verrijzer, C. P. (2006). Architecture of a polycomb nucleoprotein complex. *Molecular Cell*, 24(1), 91–100. <https://doi.org/10.1016/j.molcel.2006.08.007>
- Monaghan, P. (2010). Telomeres and life histories: The long and the short of it. *Annals of the New York Academy of Sciences*. <https://doi.org/10.1111/j.1749-6632.2010.05705.x>
- Murphy, C. T., McCarroll, S. A., Bargmann, C. I., Fraser, A., Kamath, R. S., Ahringer, J., ... Kenyon, C. (2003). Genes that act downstream of DAF-16 to influence the lifespan of *Caenorhabditis elegans*. *Nature*, 424(6946), 277–284. <https://doi.org/10.1038/nature01789>
- Murphy, M. O., Cohn, D. M., & Loria, A. S. (2017). Developmental origins of cardiovascular disease: Impact of early life stress in humans and rodents. *Neuroscience and Biobehavioral Reviews*. <https://doi.org/10.1016/j.neubiorev.2016.07.018>
- Nabavi, N., Khandani, A., Camirand, A., & Harrison, R. E. (2011). Effects of microgravity on osteoclast bone resorption and osteoblast cytoskeletal organization and adhesion. *Bone*. <https://doi.org/10.1016/j.bone.2011.07.036>

- Narasimhan, K., Lambert, S. A., Yang, A. W., Riddell, J., Mnaimneh, S., Zheng, H., ... Hughes, T. R. (2015). Mapping and analysis of *Caenorhabditis elegans* transcription factor sequence specificities. *ELife*, 4. <https://doi.org/10.7554/eLife.06967>
- Nass, R., & Hamza, I. (2007). The nematode *C. elegans* as an animal model to explore toxicology in vivo: solid and axenic growth culture conditions and compound exposure parameters. *Current Protocols in Toxicology / Editorial Board, Mahin D. Maines (Editor-in-Chief) ... [et Al.], Chapter 1, Unit 1.9*. <https://doi.org/10.1002/0471140856.tx0109s31>
- Norris, A. D., Gracida, X., & Calarco, J. A. (2017). CRISPR-mediated genetic interaction profiling identifies RNA binding proteins controlling metazoan fitness. *ELife*, 6. <https://doi.org/10.7554/eLife.28129>
- Park, P. J. (2009). ChIP-seq: Advantages and challenges of a maturing technology. *Nature Reviews Genetics*. <https://doi.org/10.1038/nrg2641>
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*. <https://doi.org/10.1038/nmeth.4197>
- Peng, J. C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., & Wysocka, J. (2009). Jarid2/Jumonji Coordinates Control of PRC2 Enzymatic Activity and Target Gene Occupancy in Pluripotent Cells. *Cell*, 139(7), 1290–1302. <https://doi.org/10.1016/j.cell.2009.12.002>
- Pérez-Lluch, S., Blanco, E., Tilgner, H., Curado, J., Ruiz-Romero, M., Corominas, M., & Guigó, R. (2015). Absence of canonical marks of active chromatin in developmentally regulated genes. *Nature Genetics*, 47(10), 1158–1167. <https://doi.org/10.1038/ng.3381>
- Pierce-Shimomura, J. T., Chen, B. L., Mun, J. J., Ho, R., Sarkis, R., & McIntire, S. L. (2008). Genetic analysis of crawling and swimming locomotory patterns in *C. elegans*. *Proceedings of the National Academy of Sciences*, 105(52), 20982–20987. <https://doi.org/10.1073/pnas.0810359105>
- Prabh, N., & Rödelsperger, C. (2016). Are orphan genes protein-coding, prediction artifacts, or non-coding RNAs? *BMC Bioinformatics*, 17(1). <https://doi.org/10.1186/s12859-016-1102-x>
- Pujato, M., Kieken, F., Skiles, A. A., Tapinos, N., & Fiser, A. (2014). Prediction of DNA binding motifs from 3D models of transcription factors; identifying TLX3 regulated genes. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gku1228>

- Qiao, L., Luo, S., Liu, Y., Li, X., Wang, G., & Huang, Z. (2013). Reproductive and locomotory capacities of *Caenorhabditis elegans* were not affected by simulated variable gravities and spaceflight during the Shenzhou-8 mission. *Astrobiology*, *13*(7), 617–625. <https://doi.org/10.1089/ast.2012.0962>
- R Development Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>. *R Foundation for Statistical Computing, Vienna, Austria*.
- Rappaport, S. M. (2016). Genetic factors are not the major causes of chronic diseases. *PLoS ONE*, *11*(4). <https://doi.org/10.1371/journal.pone.0154387>
- Rappaport, S. M., & Smith, M. T. (2010). Epidemiology. Environment and disease risks. *Science (New York, N.Y.)*, *330*(6003), 460–461. <https://doi.org/10.1126/science.1192603>
- Rechtsteiner, A., Ercan, S., Takasaki, T., Phippen, T. M., Egelhofer, T. A., Wang, W., ... Strome, S. (2010). The histone H3K36 methyltransferase MES-4 acts epigenetically to transmit the memory of germline gene expression to progeny. *PLoS Genetics*, *6*(9), e1001091. <https://doi.org/10.1371/journal.pgen.1001091>
- Reinke, V., Krause, M., & Okkema, P. (2013). Transcriptional regulation of gene expression in *C. elegans*. *WormBook : The Online Review of C. Elegans Biology*, 1–34. <https://doi.org/10.1895/wormbook.1.45.2>
- Riesen, M., Feyst, I., Rattanavirotkul, N., Ezcurra, M., Tullet, J. M. A., Papatheodorou, I., ... Gems, D. (2014). MDL-1, a growth- and tumor-suppressor, slows aging and prevents germline hyperplasia and hypertrophy in *C. elegans*. *Aging*, *6*(2), 98–117. <https://doi.org/10.18632/aging.100638>
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*(1), 24–26. <https://doi.org/10.1038/nbt.1754>
- Rosenbloom, K. R., Armstrong, J., Barber, G. P., Casper, J., Clawson, H., Diekhans, M., ... Kent, W. J. (2015). The UCSC Genome Browser database: 2015 update. *Nucleic Acids Research*, *43*(D1), D670–D681. <https://doi.org/10.1093/nar/gku1177>
- Rosenkrantz, J. T., Aarts, H., Abee, T., Rolfe, M. D., Knudsen, G. M., Nielsen, M.-B., ... Pin, C. (2013). Non-essential genes form the hubs of genome scale protein function and environmental gene expression networks in *Salmonella enterica* serovar Typhimurium. *BMC Microbiology*, *13*, 294.

<https://doi.org/10.1186/1471-2180-13-294>

- Ross-Innes, C. S., Stark, R., Teschendorff, A. E., Holmes, K. A., Ali, H. R., Dunning, M. J., ... Carroll, J. S. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*, *481*(7381), 389–393. <https://doi.org/10.1038/nature10730>
- Sachs, J. (1887). *Lectures on the physiology of plants*,. Oxford: Clarendon Press,. <https://doi.org/10.5962/bhl.title.54852>
- Saddoughi, S. A., Song, P., & Ogretmen, B. (2008). Roles of Bioactive Sphingolipids in Cancer Biology and Therapeutics. In *Lipids in Health and Disease* (Vol. 49, pp. 413–440). Dordrecht: Springer Netherlands. [https://doi.org/10.1007/978-1-4020-8831-5\\_16](https://doi.org/10.1007/978-1-4020-8831-5_16)
- Sahin, E., & Depinho, R. A. (2010). Linking functional decline of telomeres, mitochondria and stem cells during ageing. *Nature*. <https://doi.org/10.1038/nature08982>
- Salazar, J. J., Michele, D. E., & Brooks, S. V. (2010). Inhibition of calpain prevents muscle weakness and disruption of sarcomere structure during hindlimb suspension. *Journal of Applied Physiology*. <https://doi.org/10.1152/jappphysiol.01080.2009>
- Samuel, T. K., Sinclair, J. W., Pinter, K. L., & Hamza, I. (2014). Culturing *Caenorhabditis elegans* in Axenic Liquid Media and Creation of Transgenic Worms by Microparticle Bombardment. *Journal of Visualized Experiments*, (90), e51796. <https://doi.org/10.3791/51796>
- Savas, J. N., Wang, Y. Z., DeNardo, L. A., Martinez-Bartolome, S., McClatchy, D. B., Hark, T. J., ... Yates, J. R. (2017). Amyloid Accumulation Drives Proteome-wide Alterations in Mouse Models of Alzheimer's Disease-like Pathology. *Cell Reports*. <https://doi.org/10.1016/j.celrep.2017.11.009>
- Schulz, T. J., Zarse, K., Voigt, A., Urban, N., Birringer, M., & Ristow, M. (2007). Glucose Restriction Extends *Caenorhabditis elegans* Life Span by Inducing Mitochondrial Respiration and Increasing Oxidative Stress. *Cell Metabolism*. <https://doi.org/10.1016/j.cmet.2007.08.011>
- Seale, P., Bjork, B., Yang, W., Kajimura, S., Chin, S., Kuang, S., ... Spiegelman, B. M. (2008). PRDM16 controls a brown fat/skeletal muscle switch. *Nature*, *454*(7207), 961–967. <https://doi.org/10.1038/nature07182>
- Selch, F., Higashibata, A., Imamizo-Sato, M., Higashitani, A., Ishioka, N., Szewczyk,

- N. J., & Conley, C. A. (2008). Genomic response of the nematode *Caenorhabditis elegans* to spaceflight. *Advances in Space Research : The Official Journal of the Committee on Space Research (COSPAR)*, *41*(5), 807–815.  
<https://doi.org/10.1016/j.asr.2007.11.015>
- Serin, E. A. R., Nijveen, H., Hilhorst, H. W. M., & Ligterink, W. (2016). Learning from Co-expression Networks: Possibilities and Challenges. *Frontiers in Plant Science*, *7*. <https://doi.org/10.3389/fpls.2016.00444>
- Seung, W. O., Mukhopadhyay, A., Dixit, B. L., Raha, T., Green, M. R., & Tissenbaum, H. A. (2006). Identification of direct DAF-16 targets controlling longevity, metabolism and diapause by chromatin immunoprecipitation. *Nature Genetics*, *38*(2), 251–257. <https://doi.org/10.1038/ng1723>
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., ... Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11), 2498–2504.  
<https://doi.org/10.1101/gr.1239303>
- Shirayama, M., Seth, M., Lee, H. C., Gu, W., Ishidate, T., Conte, D., & Mello, C. C. (2012). PiRNAs initiate an epigenetic memory of nonself RNA in the *C. elegans* germline. *Cell*, *150*(1), 65–77. <https://doi.org/10.1016/j.cell.2012.06.015>
- Sievers, A., & Hejnowicz, Z. (1992). How well does the clinostat mimic the effect of microgravity on plant cells and organs? *ASGSB Bulletin : Publication of the American Society for Gravitational and Space Biology*, *5*(2), 69–75.
- Son, Y. S., Arif Ullah, H. M., Elfadl, A. K., Chung, M. J., Ghim, S. G., Kim, Y. D., ... Jeong, K. S. (2018). Preventive effects of Vitamin C on diethylnitrosamine-induced hepatotoxicity in Smp30 knockout mice. *In Vivo*.  
<https://doi.org/10.21873/invivo.11209>
- Spencer, W. C., Zeller, G., Watson, J. D., Henz, S. R., Watkins, K. L., McWhirter, R. D., ... Miller, D. M. (2011). A spatial and temporal map of *C. elegans* gene expression. *Genome Research*, *21*(2), 325–341.  
<https://doi.org/10.1101/gr.114595.110>
- Sterken, M. G., Snoek, L. B., Kammenga, J. E., & Andersen, E. C. (2015). The laboratory domestication of *Caenorhabditis elegans*. *Trends in Genetics*.  
<https://doi.org/10.1016/j.tig.2015.02.009>
- Stiernagle, T. (2006). Maintenance of *C. elegans*. *WormBook*, 1–11.  
<https://doi.org/10.1895/wormbook.1.101.1>

- Straczkowski, M., & Kowalska, I. (2008). The role of skeletal muscle sphingolipids in the development of insulin resistance. *Review of Diabetic Studies*.  
<https://doi.org/10.1900/RDS.2008.5.13>
- Strahl, B. D., & Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, *403*(6765), 41–45. <https://doi.org/10.1038/47412>
- Szewczyk, N. J., Kozak, E., & Conley, C. A. (2003). Chemically defined medium and *Caenorhabditis elegans*. *BMC Biotechnology*, *3*, 19. <https://doi.org/10.1186/1472-6750-3-19>
- Szewczyk, N. J., Udranszky, I. A., Kozak, E., Sunga, J., Kim, S. K., Jacobson, L. A., & Conley, C. A. (2006). Delayed development and lifespan extension as features of metabolic lifestyle alteration in *C. elegans* under dietary restriction. *The Journal of Experimental Biology*, *209*(Pt 20), 4129–4139.  
<https://doi.org/10.1242/jeb.02492>
- Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., ... Von Mering, C. (2017). The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Research*, *45*(D1), D362–D368. <https://doi.org/10.1093/nar/gkw937>
- Talukdar, H. A., Foroughi Asl, H., Jain, R. K., Ermel, R., Ruusalepp, A., Franzén, O., ... Björkegren, J. L. M. (2016). Cross-Tissue Regulatory Gene Networks in Coronary Artery Disease. *Cell Systems*.  
<https://doi.org/10.1016/j.cels.2016.02.002>
- Taylor, P. W. (2015). Impact of space flight on bacterial virulence and antibiotic susceptibility. *Infection and Drug Resistance*.  
<https://doi.org/10.2147/IDR.S67275>
- Tee, L. F., Neoh, H. min, Then, S. M., Murad, N. A., Asillam, M. F., Hashim, M. H., ... Jamal, R. (2017). Effects of simulated microgravity on gene expression and biological phenotypes of a single generation *Caenorhabditis elegans* cultured on 2 different media. *Life Sciences in Space Research*, *15*, 11–17.  
<https://doi.org/10.1016/j.lssr.2017.06.002>
- Tepper, R. G., Ashraf, J., Kaletsky, R., Kleemann, G., Murphy, C. T., & Bussemaker, H. J. (2013). PQM-1 complements DAF-16 as a key transcriptional regulator of DAF-2-mediated development and longevity. *Cell*, *154*(3), 676–690.  
<https://doi.org/10.1016/j.cell.2013.07.006>
- Thompson, O. A., Snoek, L. B., Nijveen, H., Sterken, M. G., Volkers, R. J. M., Brenchley, R., ... Waterston, R. H. (2015). Remarkably divergent regions

- punctuate the genome assembly of the *Caenorhabditis elegans* hawaiian strain CB4856. *Genetics*, 200(3), 975–989. <https://doi.org/10.1534/genetics.115.175950>
- Towbin, B. D., González-Aguilera, C., Sack, R., Gaidatzis, D., Kalck, V., Meister, P., ... Gasser, S. M. (2012). Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. *Cell*, 150(5), 934–947. <https://doi.org/10.1016/j.cell.2012.06.051>
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., ... Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, 7(3), 562–578. <https://doi.org/10.1038/nprot.2012.016>
- Turner, B. M. (2002). Cellular Memory and the Histone Code. *Cell*, 111(3), 285–291. [https://doi.org/10.1016/S0092-8674\(02\)01080-2](https://doi.org/10.1016/S0092-8674(02)01080-2)
- van Dam, S., Vösa, U., van der Graaf, A., Franke, L., & de Magalhães, J. P. (2017). Gene co-expression analysis for functional classification and gene–disease predictions. *Briefings in Bioinformatics*, bbw139. <https://doi.org/10.1093/bib/bbw139>
- Vastenhouw, N. L., Zhang, Y., Woods, I. G., Imam, F., Regev, A., Liu, X. S., ... Schier, A. F. (2010). Chromatin signature of embryonic pluripotency is established during genome activation. *Nature*, 464(7290), 922–926. <https://doi.org/10.1038/nature08866>
- Vielle, A., Lang, J., Dong, Y., Ercan, S., Kotwaliwale, C., Rechtsteiner, A., ... Ahringer, J. (2012). H4K20me1 Contributes to Downregulation of X-Linked Genes for *C. elegans* Dosage Compensation. *PLoS Genetics*, 8(9). <https://doi.org/10.1371/journal.pgen.1002933>
- Villa-Vialaneix, N., Liaubet, L., Laurent, T., Cherel, P., Gamot, A., & SanCristobal, M. (2013). The Structure of a Gene Co-Expression Network Reveals Biological Functions Underlying eQTLs. *PLoS ONE*, 8(4). <https://doi.org/10.1371/journal.pone.0060045>
- Voigt, P., Tee, W.-W., & Reinberg, D. (2013). A double take on bivalent promoters. *Genes & Development*, 27(12), 1318–1338. <https://doi.org/10.1101/gad.219626.113>
- Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W. H. A., Lumley, T., ... Venables, B. (2015). gplots: Various R Programming Tools for Plotting Data. Retrieved September 9, 2015, from <https://cran.r-project.org/web/packages/gplots/index.html>

- Watson, E., MacNeil, L. T., Arda, H. E., Zhu, L. J., & Walhout, A. J. M. (2013). Integration of metabolic and gene regulatory networks modulates the *C. elegans* dietary response. *Cell*, *153*(1), 253–266. <https://doi.org/10.1016/j.cell.2013.02.050>
- Watson, E., & Walhout, A. J. M. (2014). *Caenorhabditis elegans* metabolic gene regulatory networks govern the cellular economy. *Trends in Endocrinology and Metabolism*. <https://doi.org/10.1016/j.tem.2014.03.004>
- Wei, H., Persson, S., Mehta, T., Srinivasasainagendra, V., Chen, L., Page, G. P., ... Loraine, A. (2006). Transcriptional Coordination of the Metabolic Network in *Arabidopsis*. *PLANT PHYSIOLOGY*, *142*(2), 762–774. <https://doi.org/10.1104/pp.106.080358>
- Weimer, S., Priebs, J., Kuhlow, D., Groth, M., Priebe, S., Mansfeld, J., ... Ristow, M. (2014). D-Glucosamine supplementation extends life span of nematodes and of ageing mice. *Nature Communications*. <https://doi.org/10.1038/ncomms4563>
- Wuest, S. L., Stern, P., Casartelli, E., & Egli, M. (2017). Fluid dynamics appearing during simulated microgravity using Random Positioning Machines. *PLoS ONE*, *12*(1). <https://doi.org/10.1371/journal.pone.0170826>
- Xue, J., Schmidt, S. V., Sander, J., Draffehn, A., Krebs, W., Quester, I., ... Schultze, J. L. (2014). Transcriptome-Based Network Analysis Reveals a Spectrum Model of Human Macrophage Activation. *Immunity*, *40*(2), 274–288. <https://doi.org/10.1016/j.immuni.2014.01.006>
- Yook, K., Harris, T. W., Bieri, T., Cabunoc, A., Chan, J., Chen, W. J., ... Sternberg, P. W. (2012). WormBase 2012: more genomes, more data, new website. *Nucleic Acids Research*, *40*(Database issue), D735–41. <https://doi.org/10.1093/nar/gkr954>
- Zang, C., Schones, D. E., Zeng, C., Cui, K., Zhao, K., & Peng, W. (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics (Oxford, England)*, *25*(15), 1952–1958. <https://doi.org/10.1093/bioinformatics/btp340>
- Zhang, A. T., Langley, A. R., Christov, C. P., Kheir, E., Shafee, T., Gardiner, T. J., & Krude, T. (2011). Dynamic interaction of Y RNAs with chromatin and initiation proteins during human DNA replication. *Journal of Cell Science*, *124*(12), 2058–2069. <https://doi.org/10.1242/jcs.086561>
- Zhang, B., & Horvath, S. (2005). A General Framework for Weighted Gene Co-Expression Network Analysis. *Statistical Applications in Genetics and Molecular Biology*, *4*(1). <https://doi.org/10.2202/1544-6115.1128>

- Zhang, Y., Chen, D., Smith, M. A., Zhang, B., & Pan, X. (2012). Selection of reliable reference genes in *caenorhabditis elegans* for analysis of nanotoxicity. *PLoS ONE*, 7(3). <https://doi.org/10.1371/journal.pone.0031849>
- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., ... Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, 9(9), R137. <https://doi.org/10.1186/gb-2008-9-9-r137>
- Zhao, S., Fung-Leung, W. P., Bittner, A., Ngo, K., & Liu, X. (2014). Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS ONE*, 9(1). <https://doi.org/10.1371/journal.pone.0078644>
- Zhou, Z., Hartwig, E., & Horvitz, H. R. (2001). CED-1 is a transmembrane receptor that mediates cell corpse engulfment in *C. elegans*. *Cell*, 104(1), 43–56. [https://doi.org/10.1016/S0092-8674\(01\)00190-8](https://doi.org/10.1016/S0092-8674(01)00190-8)
- Zhu, F., Panwar, B., & Guan, Y. (2016). Algorithms for modeling global and context-specific functional relationship networks. *Briefings in Bioinformatics*, 17(4), 686–695. <https://doi.org/10.1093/bib/bbv065>

**Appendix**  
**PERMISSIONS**

Chapter 2 (Çelen et al., 2017, 2018) was published in BMC Genomics and Chapter 3 (Çelen et al., 2019) is under review at the time of the submission of the dissertation. Both the chapters were submitted to bioRxiv as preprints under Creative Commons CC-BY-NC-ND 4.0 International license.

BMC Genomics:

<https://bmcgenomics.biomedcentral.com/submission-guidelines/copyright;>

CC-BY-NC-ND 4.0 International license:

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Neither of the publishing bodies require a permission for reprint of author's articles in their own dissertation.