

أكاديمية كاوست
KAUST ACADEMY



UNIVERSITY OF
CAMBRIDGE



جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology

PROTEOMIC INSIGHTS INTO PROTEIN EXPRESSION DYNAMICS DURING C. ELEGANS LARVAL DEVELOPMENT

What are the key biological processes associated with significant protein expression changes during C. elegans larval development ?

Bioinformatics Summer School Report

Report Prepared by: Raghad Said Algarni

Supervised by: Dr.Hugo Tavares

Table of Contents

INTRODUCTION 3

AIMS AND OBJECTIVES 4

METHODOLOGY AND APPROACHES 4

RESULTS 6

CONCLUSION 7

CHALLENGES AND OBSTACLES 8

FUTURE WORK 8

REFERENCES: 9

Introduction

Caenorhabditis elegans (C. elegans) is a common model organism used in scientific research due to its well-characterized genome, transparent body, and short life cycle. It's ideal for studying developmental biology, genetics, neuroscience, and cell biology. Because of its distinct and well-characterized embryonic and post-embryonic cell lineages, C. elegans is a highly useful model organism for understanding development. Once the worm hatches from the egg, it travels through four larval stages (L1–L4) before becoming an adult *(Figure 1)*. Studying proteins involved in larval development has contributed to understanding cell fate specification, timing of developmental transitions, and responses to environmental signals.

This review examines protein expression changes during the early (L1) and late (L4) stages of larval development in Caenorhabditis elegans using stage-specific proteomic data.

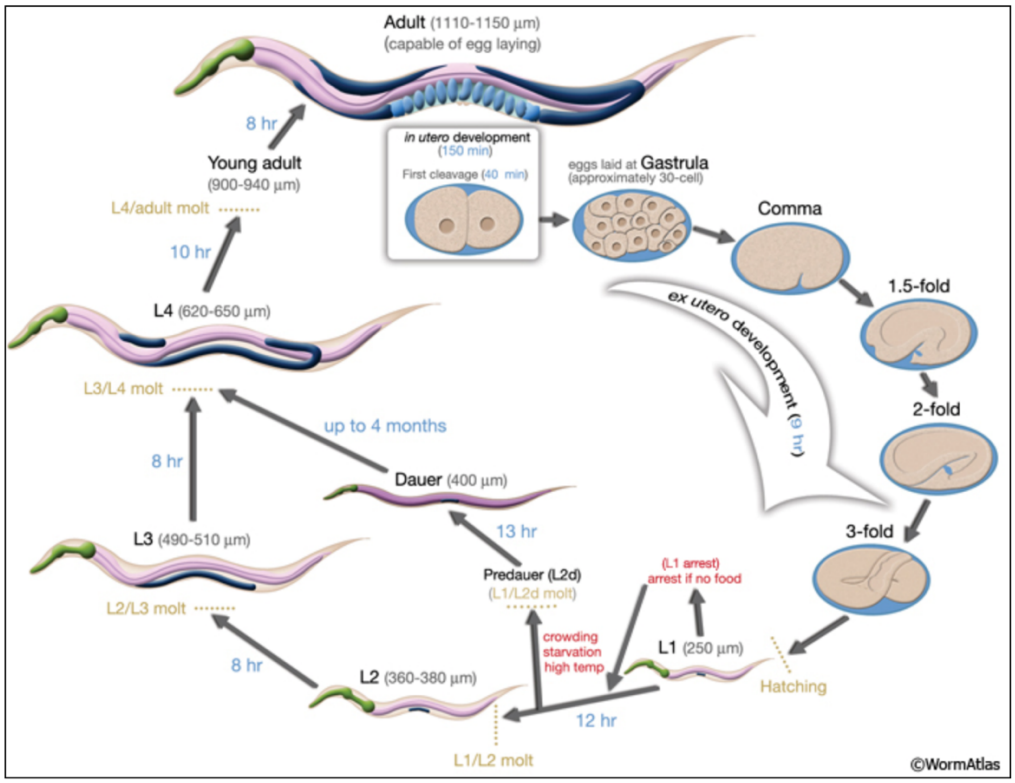


Figure 1. *Developmental stages of C. elegans early embryo to adult.*

Aims and Objectives

The aim of this review is to investigate proteomic changes that occur during larval development of *Caenorhabditis elegans*, focusing on the transition from the L1 to L4 stages.

The specific objectives are:

1. To analyze protein expression changes between L1 and L4 using quantitative proteomic data.
2. To identify proteins involved in developmental regulation, such as timing factors (e.g., AIN-1) and molting proteins (e.g., ACN-1), and compare them with known genetic studies.
3. To link protein expression dynamics to biological processes, including molting, heterochronic regulation, and growth, through clustering and pathway analysis.

Methodology and approaches

The input dataset consisted of quantitative proteomic measurements at the protein level across larval stages of *C. elegans*. As no peptide- or PSM-level data were available, the analysis was performed directly on protein intensities using RStudio (v2025.05.1). Data were imported using the readxl package, and associated sample metadata were processed with dplyr and tidyr. To facilitate consistent organization of assays and metadata, the protein-level data were incorporated into a QFeatures object, even though no aggregation from lower levels was required.

To improve data quality, proteins with more than 40% missing values or with low mean intensity (≤ 5) were excluded. The filtered data were then $\log_2$-transformed to stabilize variance and median-centered for normalization using the normalize function. Boxplots were generated before and after normalization to confirm the reduction of technical bias across samples.

Global variance structure was assessed by Principal Component Analysis (PCA), which confirmed that biological replicates clustered together and developmental stages separated consistently, validating the robustness of the dataset. Differential

protein expression was analyzed using the limma package. Importantly, we applied the global testing method in limma (decideTests with method = "global"), which evaluates all contrasts simultaneously while controlling the false discovery rate across the entire experiment. This approach was chosen over the "separate" testing method, which assesses each pairwise comparison independently and may inflate false positives, in order to achieve a more reliable identification of proteins with consistent developmental changes across multiple stage transitions.

Significantly regulated proteins (adjusted p-value < 0.01, $\log_2$FC > 1) were visualized with volcano plots for each stage comparison (e.g., L2 vs L1, L3 vs L1, L4 vs L1). These volcano plots were used to confirm which developmental stages exhibited the strongest sets of significantly regulated proteins, thereby highlighting stage transitions most critical for larval progression.

To explore stage-specific patterns, significant proteins were clustered using K-means (k = 6), and their expression profiles were visualized as heatmaps with pheatmap *(Figure 2)*. Z-score normalization was applied to compare relative expression dynamics across stages. To interpret biological functions, Gene Ontology (GO) enrichment analysis was conducted with clusterProfiler using C. elegans annotations from org.Ce.eg.db, identifying biological processes enriched in different protein clusters.

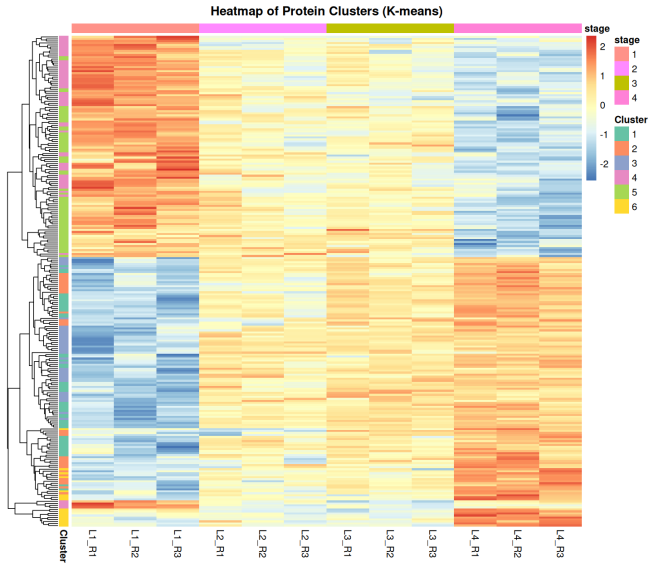


Figure 2. *Heatmap of Protein Clustering*

Protein–protein interaction (PPI) networks were constructed using the STRING database, which provided the significant phenotype annotations, enabling the identification of proteins associated with key developmental outcomes, including organismal growth, larval progression, behavior, and cell physiology.

Finally, from the significantly regulated proteins, we selected those that both showed strong network interactions and were linked to phenotypes affecting organismal development. To further validate these proteins, we compared them against previously published proteomic and genetic literature on C. elegans larval development. This final step ensured that the proteins we highlight revealed functional modules such as mitochondrial metabolism, molting regulation, and timing-related pathways. This approach gave us not only statistically robust findings but also biologically meaningful regulators of developmental progression.

Results

The findings showed that out of a total of 4,130 proteins analyzed across the L1 to L4 larval stages of *C. elegans*, 244 proteins exhibited statistically significant expression changes, highlighting dynamic proteomic shifts during larval development. The observed changes were consistent across all biological replicates, confirming the reliability of the measurements. Volcano plot analysis identified the most significant differences in protein expression at the L1 and L4 stages, indicating that these stages are characterized by pronounced regulatory activity. This trend was further supported by hierarchical clustering of all three replicates, visualized through a heat map, which revealed clear separation of stage-specific expression patterns. To further categorize proteins according to their developmental expression dynamics, K-means clustering of averaged replicate data was performed, resulting in six distinct clusters. Clusters 1, 2, and 3 contained proteins with high expression at the L4 stage, which are predominantly involved in processes associated with late-stage development, such as molting and tissue differentiation. In contrast, Clusters 4 and 5 consisted of proteins with elevated expression at the L1 stage, reflecting functions essential during early development, including cell proliferation and initial organogenesis*(Figure 3)*.

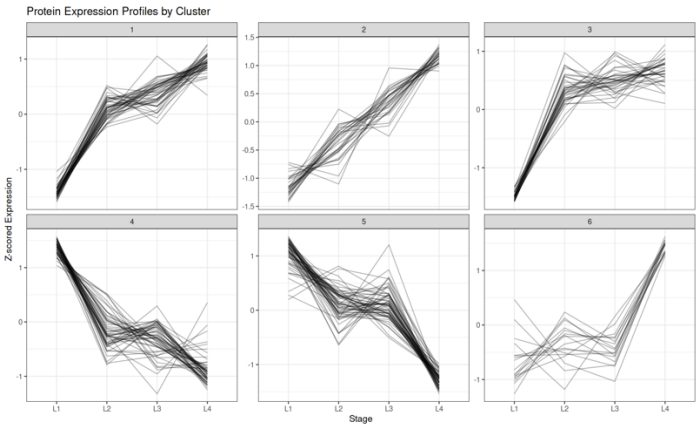


Figure 3. *K-means Clustering*

Beyond statistical analysis, our dataset allowed the identification of proteins with significant phenotypic effects on organismal development, growth, larval progression, behavior, and cell physiology. For example, AIN-1, which falls into Cluster 5, exhibited peak expression at L1 that gradually declined through subsequent larval stages, reaching its lowest level at L4. This expression pattern suggests that AIN-1 plays a critical early regulatory role, likely facilitating miRNA-mediated control of gene expression to ensure proper developmental timing. Foundational studies by Ding et al. (2005) have demonstrated that AIN-1 is essential for heterochronic regulation, with its loss resulting in disrupted larval stage transitions and improper adult differentiation, highlighting the importance of its early activity. In contrast, ACN-1, classified in Cluster 1, displayed low expression at L1 that increased progressively through the larval stages, peaking at L4. This trend indicates that ACN-1's functional importance grows during later development, potentially linked to its enzymatic activity during the final molt. Supporting evidence from Kumar et al. (2024) suggests that ACN-1 possesses enzymatic functions, as inhibition of related pathways can partially mimic acn-1 RNAi phenotypes, underscoring its crucial role during the final developmental stages. Together, these results provide a comprehensive view of proteomic dynamics throughout larval development, linking stage-specific protein expression to critical biological processes and functional phenotypes.

Conclusion

In summary, this study provides a comprehensive analysis of stage-specific proteomic changes during *C. elegans* larval development, emphasizing the regulatory significance of the L1 and L4 stages. The integration of quantitative proteomics with clustering, functional enrichment, and interaction network analysis enabled the

identification of key proteins involved in developmental timing and progression. Findings related to proteins such as AIN-1 and ACN-1 underscore the importance of coordinated protein expression in guiding early developmental events and late-stage transitions, offering deeper insight into the molecular architecture of larval development.

Challenges and obstacles

The project encountered several notable challenges that impacted the efficiency of data analysis and preparation. A significant limitation was the constrained timeframe, as only two to three days were available between receiving the dataset and the poster submission deadline. This limited period restricted the depth of analysis and refinement that could be applied to the final output. Furthermore, access to computing resources was limited; personal laptops could not be used for data processing, necessitating reliance on departmental computers. These machines were not consistently available and presented additional difficulties due to slow performance and insufficient storage capacity, particularly given the large size of the dataset. These factors collectively posed logistical and technical obstacles, which required careful time management and prioritisation to ensure completion of the project within the given constraints.

Future work

Future research could extend this analysis by focusing on the L2 and L3 stages to determine whether significant proteomic or functional differences exist between them. Investigating these intermediate stages may clarify their distinct biological roles and contribute to understanding why *C. elegans* development is divided into four specific larval stages. Integrating transcriptomic data alongside proteomics would offer a more comprehensive view of gene regulation across development. Experimental validation of key proteins, such as AIN-1 and ACN-1, using RNAi or CRISPR, could further confirm their roles in developmental timing. Lastly, enhancing computational resources and incorporating automated analysis pipelines would improve the efficiency and scalability of future proteomic studies.

References:

- Ambros, V. (2011) 'MicroRNAs and Developmental Timing', *Current Opinion in Genetics & Development*, 21(4), pp. 511–517. doi:10.1016/j.gde.2011.04.003.

- Brenner, S. (1974) 'The genetics of *Caenorhabditis elegans*', *Genetics*, 77(1), pp. 71–94. doi:10.1093/genetics/77.1.71.

- Brooks, D.R. et al. (2003) 'An essential role in molting and morphogenesis of *Caenorhabditis elegans* for ACN-1, a novel member of the angiotensin-converting enzyme family that lacks a metallopeptidase active site', *Journal of Biological Chemistry*, 278(52), pp. 52340–52346. doi:10.1074/jbc.m308858200.

- Ding, L. et al. (2005) 'The developmental timing regulator ain-1 interacts with miriscs and may target the Argonaute protein ALG-1 to cytoplasmic p bodies in *C. elegans*', *Molecular Cell*, 19(4), pp. 437–447. doi:10.1016/j.molcel.2005.07.013.

- Frand, A.R., Russel, S. and Ruvkun, G. (2005) 'Functional genomic analysis of *C. elegans* molting', *PLoS Biology*, 3(10). doi:10.1371/journal.pbio.0030312.

- STRING Consortium (2025) 'STRING: functional protein association networks', available at: https://string-db.org/cgi/input?sessionId=bFsQOZ6viyYM&input_page

- Sulston, J.E. and Horvitz, H.R. (1977) 'Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*', *Developmental Biology*, 56(1), pp. 110–156. doi:10.1016/0012-1606(77)90158-0.

- Sundaram, M.V. and Pujol, N. (2024) 'The *Caenorhabditis elegans* cuticle and precuticle: A model for studying dynamic apical extracellular matrices in vivo', *GENETICS*, 227(4). doi:10.1093/genetics/iyae072.

- WormAtlas (2024) 'Introduction to the Hermaphrodite Anatomy', available at: https://www.wormatlas.org/hermaphrodite/introduction/mainframe.htm (Accessed: 21 August 2025).

- Xia, T. et al. (2017) 'Proteomic characterization of *Caenorhabditis elegans* larval development', *PROTEOMICS*, 18(2). doi:10.1002/pmic.201700238.

Codes used in the analysis of this review are available here:

https://colab.research.google.com/drive/1p8lFDNx8f3BhhKynic6qZhk6AQa6hUOM?usp=sharing