Bioinformatics Summer Programme Report

Student name: Afnan Abdullah Abdulwahab
Email: Afnanbdulwahab@gmail.com
Internship Organization: Genetics department, Cambridge University
Project title: Gene Expression dynamics reveal developmental processes during *C. elegans* Embryogenesis using Time-Course RNA-seq

Abstract:
This project analyzed embryonic development in *C. elegans* using time-course bulk RNA-seq. From 12,584 genes, 8,389 were identified as significantly varying across stages through DESeq2 likelihood ratio testing. Genes were grouped into six temporal clusters, each showing distinct activation or repression patterns. Normalization and dimensionality reduction (rlog, UMAP) revealed smooth developmental trajectories. Gene Ontology enrichment linked clusters to key biological processes, from early cell division and DNA replication to later differentiation and tissue formation. These findings reveal how dynamic gene programs guide embryonic progression across developmental stages.

Introduction:

Embryogenesis is a stepwise process where gene activity changes over time to move the embryo from one stage to the next. Understanding these temporal patterns of expression can reveal how cell division, growth, and differentiation are coordinated during development. *Caenorhabditis elegans* is an ideal model because its development is well mapped and many of its molecular pathways are conserved across animals. Using bulk RNA-seq in a time-course design allows us to follow genome-wide expression programs across multiple stages.
In this work, our focus was to identify which sets of genes become active or silent at each developmental stage, and what biological processes they represent. Thus, this project aimed to uncover stage-specific gene expression programs that guide embryonic transitions.

Aim:

The aim of this project was to uncover stage-specific gene expression patterns in *C. elegans* embryogenesis using bulk RNA-seq, linking dynamic gene clusters to biological processes that guide developmental transitions.

Workflow and Analysis:

The analysis started with **Data Import and Preparation**, where raw RNA-seq counts from *C. elegans* embryos were loaded, low-quality samples removed, and low-count genes filtered out. Then, **Differential Expression Analysis** was performed with DESeq2 using the Likelihood Ratio Test (LRT), revealing 8,389 dynamic genes out of 12,584. In **Normalization and Dimensionality Reduction**, data were normalized and transformed with rlog, and UMAP was applied to visualize sample separation by stage. Genes were then grouped by **Clustering of Genes**, with hierarchical clustering identifying six temporal modules showing distinct activity patterns. Finally, **Functional Enrichment Analysis** linked each cluster to major biological processes—such as cell cycle, tissue morphogenesis, and RNA processing highlighting stage-specific regulatory programs in embryogenesis.

Results and Findings:

From the 12,584 genes analyzed, 8,389 showed significant temporal changes across *C. elegans* embryogenesis. To interpret these dynamics, genes were grouped into six expression clusters, each showing stage-specific regulation. Early clusters were enriched in processes such as cell cycle, DNA replication, and genome stability, while later clusters reflected functions including cell migration, tissue morphogenesis, and RNA processing.

These results demonstrate that gene expression programs occur in coordinated waves: some genes are activated early to drive proliferation, while others dominate later to guide differentiation and organ formation. Together, the findings demonstrate how bulk RNA-seq uncovers stage-specific gene programs that shape embryogenesis from a single cell to a complex organism.

Challenges and Obstacles Encountered:

The main challenges during the analysis included selecting the optimal number of clusters, since too few clusters made the patterns overcrowded while too many led to redundant biological terms. Another obstacle was simplifying overlapping GO enrichment terms to extract the clearest biological meaning.
Finally, choosing clear visualizations such as GO heatmap and tree plots, was key effectively communicate stage-specific processes.

Skills and Knowledge Gained

Through this program, I developed both technical skills and conceptual understanding. I gained hands-on experience in RNA-seq data analysis, visualization, and interpretation of gene expression dynamics, alongside strengthening my data analysis abilities. Conceptually, I deepened my understanding of developmental gene regulation. In addition, the program improved my presentation and communication skills, giving me confidence in speaking to audiences and presenting scientific findings clearly in both written and visual formats.

Next Steps and Future Work:

Future analyses could integrate single-cell RNA-seq for higher resolution of gene expression dynamics, distinguishing individual cell types within stages. Another direction is to compare *C. elegans* developmental programs with other model organisms such as *Drosophila* or *zebrafish*. Such comparisons may reveal conserved developmental mechanisms relevant beyond *C. elegans*. Finally, expanding to additional annotations (e.g., KEGG pathways, transcription factor networks) could uncover regulatory interactions driving developmental transitions.

Acknowledgments:

I would like to thank Professor Hugo Tavares for his guidance, as well as the program leads in Cambridge and KAUST Academy for providing this valuable opportunity. I am also grateful to Mr. Sultan Al-Barakati, Ms. Orooba Al-Nasr, and our supervisor Abdulaziz Al-Tailmesani, along with my colleagues in Cambridge, whose presence made this journey more enriching and impactful.

Resources and Tools Used:

The analysis was conducted in R using packages including DESeq2 for differential expression, tximport for count data import, pheatmap and ComplexHeatmap for visualization, bluster for clustering, and clusterProfiler for GO enrichment.
Data and pipeline references were drawn from the *C. elegans* embryogenesis dataset (Hashimshony et al., 2015) and the RNA-seq analysis course (Tavares, H., 2023).
The complete R script used for the analysis is available at the following link: https://drive.google.com/drive/folders/1qUbqx_T1V-40nB5zoGqkRBLeZ4h5F-3T?usp=share_link

References:

1. APSnet (2023) *Caenorhabditis elegans* as a model organism. https://www.apsnet.org/edcenter/apsnetfeatures/Pages/Celegans.aspx
2. Hashimshony et al. (2015) Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer. https://doi.org/10.1038/nature13996
3. Tavares, H. (2023). RNA-seq analysis course: Workflows and differential expression with DESeq2. https://tavareshugo.github.io/rnaseq-course-nxf/index.html