Bioinformatics Summer School Report

Names: Suhaib Alghamdi, Khalid Alnashri

Project Title: Correlation between Embryo Development and Cell Types of the *C. elegans*

Internship Organization: Cambridge University

Supervisor's Names: Dr. Hugo Tavares, Dr Andries Van Tonder

1. Introduction:

Caenorhabditis elegans is a nematode that is about 1 mm in length that lives in temperate soil environments. The reason why this experiment looks at it is due to it having 558 cells at hatching. Despite its limited cell number, however, *C. elegans* contains a wide diversity of cell types and a complex anatomy. This and the fact that its lineage has been fully mapped. allows for a comprehensive study of the animal.

This experiment looks at the *C. elegans* and its embryo development stages and correlates these stages to its different cell types.

2. Materials and Methodology:

Bulk RNA-Seq:

To find the correlation between the cell types and Embryo Development stages we had to first find the genes that correlate to each Embryo Development stage. To do so we looked at a Bulk RNA-Seq dataset that featured samples spanning through all the stages of embryo development. Using a Likelihood Ratio Test to find the best model to represent the dataset we were able to find the 7000 most highly expressed significant genes.

Single Cell RNA-Seq:

The second step in this project is finding genes that are highly variable in each cell type. To find these cell types we investigated a different dataset that featured Single Cell RNA-Seq samples taken at different timepoints. Using the UMAP algorithm, we were able to find 21 different clusters of cells, representing 8 different cell types and in each cluster the most highly variable genes.

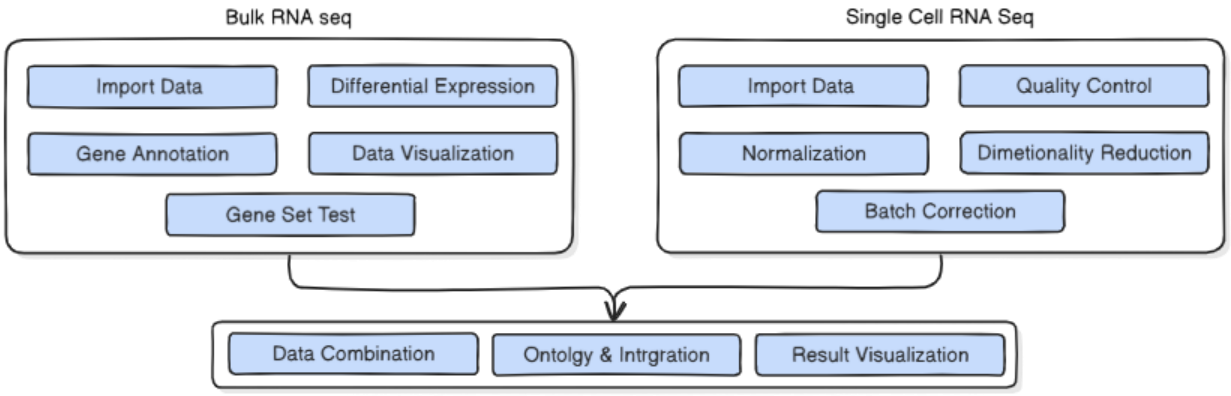


Figure 1.0: Methodology.

3. Results and discussion

This Principal Component Analysis showed a clear segregation of samples based on developmental stage, indicating that distinct transcriptional programs regulate each major phase of embryogenesis.

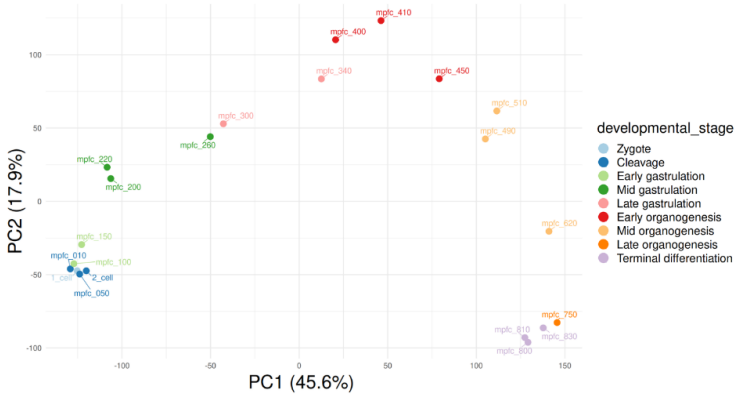


Figure 2.0: Explain the variation of the genes in each development stage.

The examination of the expression heatmap revealed a set of genes that demonstrated stage-specific activity. This included a cluster of genes that were significantly upregulated during early development but subsequently silenced.

The changes in the transcriptome observed across the samples correlate with and reflect the mechanisms facilitating developmental progression.

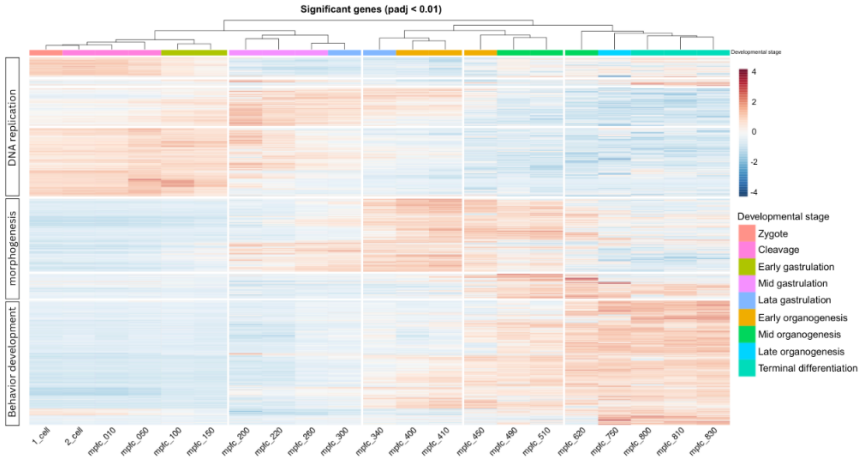


Figure 3.0: (1%) significant genes heat map plot shows the cluster of the sample based on hierarchical clustering and the expression in each development stage.

We used the UMAP Algorithm using the Louvain method to create 21 clusters each cluster was marked with the most significant cell type using anatomical ontology from Wormenrich the following figures show each cluster with its time points and cell types.

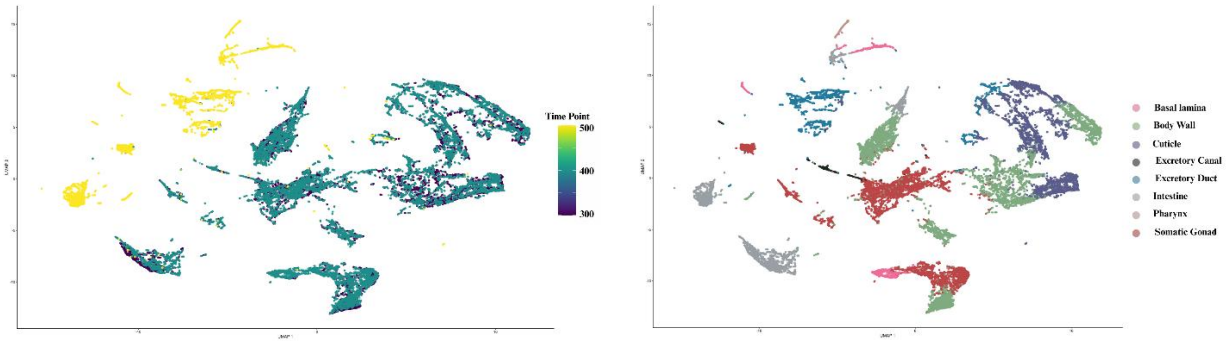


Figure 4.0: UMAP clusters by timepoint and cell type

This heatmap shows the final result of the experiment. Each column represents a cluster of cells, each of which is marked with the most significant cell type and timepoint. The rows here represent the 7000 most significant genes taken from the bulk dataset these genes are marked by their most up-regulated stage. Using this heatmap we can draw conclusions on the relationship between cell types and embryo stages.

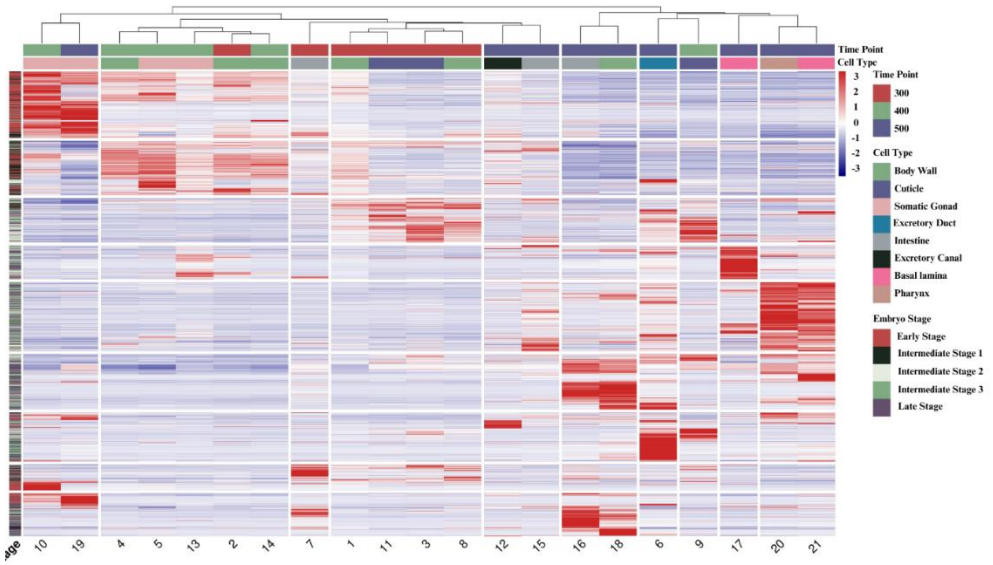


Figure 5.0 Comparing cell types with embryo stages through a heatmap

4. Skills and Knowledge Gained

- Analyzing RNA-Seq Data
- Communication and teamwork
- Biological understanding of organisms
- Visualization of results

5. Challenges and Solutions

The first and biggest challenge that we faced was time, as the course was structured to allow for a very small window of 4 days for the entire project. During these 4 days we had to work overtime each day to find the answers to our questions.

Another challenge we faced was the integration of our two data sets. We had to find a way to categorize each gene, then we had to categorize the cells into clusters and combine both clusters and genes to find our result. All while trying not to contaminate the data.

6. Conclusion

We were able to discern that there is a relationship between the Early stage of embryo development and the development of a specific type of cell, that being the Somatic gonad. But once the embryo starts going into the later stages of development the cell types start getting developed across the board with no specific type of cell being majorly correlated with a specific development stage.

Using this information, we would be able to tell if there is an issue in the child during the early stage of embryo development by checking which of the correct cell types are being developed.

7. Acknowledgement

We would like to acknowledge KAUST Academy for giving us the opportunity to get into this summer program and for supporting us through the previous stage.

We would also like to thank everyone in the Cambridge genetics department for their expertise and for teaching us throughout the summer period.

And special thanks to Dr. Hugo Tavares, Dr Andries Van Tonder, for their guidance throughout the project.

References

Levin, M., Anavy, L., Cole, A. et al. The mid-developmental transition and the evolution of animal body plans. Nature 531, 637–641 (2016). https://doi.org/10.1038/nature16994

Hashimshony, T., Feder, M., Levin, M. et al. Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer. Nature 519, 219–222 (2015). https://doi.org/10.1038/nature13996

Packer, J.S., Zhu, Q., Huynh, C., Sivaramakrishnan, P., Preston, E., Dueck, H., Stefanik, D., Tan, K., Trapnell, C., Kim, J., Waterston, R.H. and Murray, J.I. (2019). A lineage-resolved molecular atlas of C. elegans embryogenesis at single-cell resolution. Science, 365(6459), p.eaax1971. doi:https://doi.org/10.1126/science.aax1971.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G., Clark, N.R. and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics, 14(1), p.128. doi:https://doi.org/10.1186/1471-2105-14-128.