Identifying the Regenerative Organizing Cell in Frog Tail Regeneration

Boiken Jaho (bj2442)

October 2024

Abstract

This study identified the Regenerative Organizing Cell (ROC) using single-cell RNA sequencing (scRNA-seq) data obtained from observations on the regenerating tail of Xenopus frogs. Through data preprocessing, dimensionality reduction, and clustering analysis, several defining marker genes were found. Then, using Gene Ontology (GO) analysis, these marker genes were linked to important biological processes including DNA replication/repair that are essential for the regneration process. The identified genes were found to overlap with supplementary materials from the original data source, which further emphasized the significance of these particular cell types in the regenerative process and also provides insights on the underlying regenerative molecular mechanisms.

1 Introduction

Regeneration-Organizing Cells (ROCs) are a unique cell type that is essential for tail regrowth post amputation as well as to drive regeneration in Xenopus laevis tadpoles. Single-cell RNA sequencing is used to identify ROCs, which move to the amputation plane to generate a specific wound epidermis that is necessary for regeneration. ROCs express important ligands that encourage cell growth, namely Wnt5a and Fgf10. Building on this, the study compares marker gene identification techniques and uses UMAP in conjunction with several clustering techniques (Leiden and Louvain) to analyze scRNA-seq data and find clusters that are enhanced for ROC markers, like PCNA and MCM2. This work adds to our knowledge of the molecular mechanisms behind regeneration by shedding further light on the patterns of gene expression that underpin tail regeneration.

2 Methods

2.1 Data Preprocessing

The project's dataset came from a published study, which included 31,535 genes and 13,199 cell gene expression patterns. We conducted our processing and

analysis in the following steps:

- **Data loading**: The dataset was loaded using the `scanpy` library.
- **Normalization**: We log-normalized the gene expression data using the function and selected the top 2,000 highly variable genes (HVGs) for analysis. Using the `sc.pp.log1p()` function, we log-normalized the gene expression data and subsequently analyzed the top 2,000 highly variable genes (HVGs).
- **Dimensionality reduction**: Principal component analysis (PCA) was applied to reduce the data dimensionality, keeping the first 7 principal components.
- **Clustering**: Leiden and Louvain clustering methods were used to distinguish between the different cell populations. UMAP was then used to visualize the clusters generated.

2.2 Clustering and Metrics

The groups found via Leiden clustering were visualized using UMAP. The quality of clustering was assessed using the Silhouette score. Additionally, the clusters found by the alternate clustering method kNN+Louvain were visualized using UMAP. Then, the Adjusted Rand Index (ARI) was computed to provide a measure of the similarity between the two clustering methods employed.

2.3 Marker Gene Identification

Logistic regression was used to identify marker genes for every cluster. The genes linked to the ROC were identified by comparing these marker genes with those found in the Supplementary materials for an added layer of biological significance in the regeneration process.

2.4 Gene Ontology (GO) Analysis

Using the common genes from our study, we conducted GO enrichment analysis, which revealed important biological processes associated with DNA replication and repair and showed the genes found were, in fact, connected to the process of regeneration.

2.5 Code Availability

The full analysis code can be found in the following Github link: GitHub Repository: STAT5243

3 Results

3.1 Clustering Analysis

Figure 1 displays the UMAP visualization of the clusters. The Leiden and Louvain algorithms was used to divide the cells into discrete clusters; Leiden produced 24 clusters, slightly higher than the 19 clusters identified by Louvain+kNN. This indicates that Leiden produces more refined clusters that are more internally coherent while Louvain maximizes modularity by maximizing smaller clusters into one at times. The Leiden and Louvain clusters were compared using the Adjusted Rand Index (ARI) to evaluate the similarity. The results showed a score of 0.637, showing a respectable level of consistency between the two approaches. Leiden clustering had a Silhouette score of 0.37, whereas Louvain+kNN clustering had a score of 0.42, suggesting that both clustering techniques performed similarly well in separating different cell populations, however there may be some ambiguity or overlap between clusters.

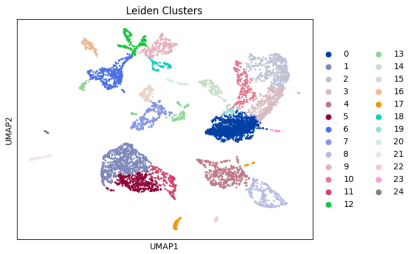

Figure 1: UMAP visualization of Leiden clusters.

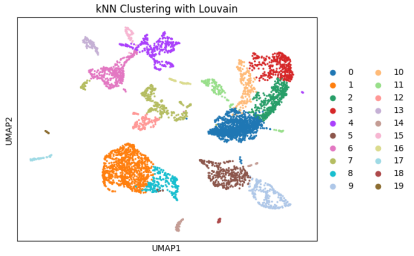

Figure 2: kNN Clustering with Louvain.

3.2 Marker Gene Identification and Comparison

Logistic regression was used to identify marker genes for every cluster. These indicators were contrasted with the genes from the Aztekin et al. study that are included in Supplementary Table 3. The study identified a number of shared genes that are vital to the regeneration process and play important roles in cell proliferation, DNA replication, and repair processes that are needed for tissue regeneration in Xenopus tadpoles.

- **PCNA (Proliferating Cell Nuclear Antigen)**: Involved in DNA replication and repair, marking proliferating cells.
- **MCM2, MCM4, MCM5, and MCM6**: Essential components for the initiation of DNA replication.
- **RPA2 (Replication Protein A)**: Binds single-stranded DNA during replication and repair processes.

- **FEN1 (Flap Structure-Specific Endonuclease 1)**: Involved in processing the 5' ends of Okazaki fragments during DNA replication.
- **RAD51 and BLM (Bloom Syndrome Protein)**: Crucial for homologous recombination and DNA repair mechanisms.
- **USP1 (Ubiquitin-Specific Peptidase 1)**: Plays a role in DNA damage response through.
- **BRIP1 (BRCA1 Interacting Protein C-terminal Helicase 1)**: A helicase involved in DNA repair, interacting with BRCA1.
- **CCNE2 (Cyclin E2)**: Involved in cell cycle regulation and the G1/S transition.
- **RRM1 (Ribonucleotide Reductase M1)**: Involved in nucleotide metabolism and DNA synthesis.

3.3 Gene Ontology (GO) Enrichment Analysis

Gene Ontology enrichment analysis was conducted on the common genes found mainly in clusters 1, 2, and 5 (among others) to be connected to ROCs. The investigation demonstrated an enrichment in biological activities, including DNA replication, repair, and metabolism, which lends additional reliability to the genes' capacity for regeneration. These are the leading procedures:

- DNA metabolic process (GO:0006259)
- DNA-dependent DNA replication (GO:0006261)
- DNA repair (GO:0006281)

The GO enrichment bar plot is shown in Figure 3.

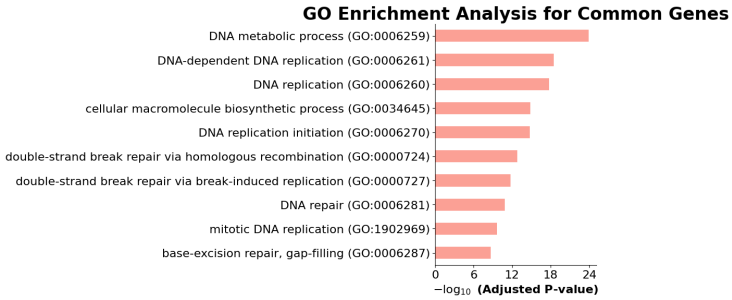

Figure 3: Bar plot of top GO terms enriched in the common genes.

3.4 Limitations

There is a number of limitations that should be taken into account when considering the resutls above. First, only modest cluster separation was indicated by the comparatively low silhouette scores for the Leiden and Louvain clustering (0.37 and 0.42, respectively). This implies that there might not be clear boundaries between groups, and that significant biological differences might be hidden by the the resolution of the clustering. Furthermore, even if marker genes were found using computational techniques like logistic regression, these markers still require experimental approaches for validation. Lastly, the *Xenopus* model was the exclusive focus of the investigation, which would restrict the applicability of these results to other species, such as mammals.

4 Conclusion

In this study, we identified the Regenerative Organizing Cell (ROC) in the regenerating tail of *Xenopus* frogs through comprehensive scRNA-seq analysis. By employing clustering methods and comparing marker genes with those listed in relevant supplementary studies we revealed several key genes, such as PCNA, MCM2, and RAD51, which play essential roles in DNA replication and repair during regeneration. The significant overlap with known regenerative markers further validated our findings. These insights into the molecular underpinnings of tail regeneration not only deepen our understanding of regenerative biology but also pave the way for future therapeutic interventions aimed at enhancing regenerative capacity in other organisms, including humans.

References

[1] C. Aztekin, T. W. Hiscock, J. C. Marioni, J. B. Gurdon, B. D. Simons, J. Jullien, *Identification of a regeneration-organizing cell in the Xenopus tail*, Science, 17 May 2019, 364, 653-658 (2019). DOI: 10.1126/science.aav9996.

[2] N. P. Rougier, et al., *Ten Simple Rules for Better Figures*, PLOS Computational Biology, Public Library of Science, DOI: 10.1371/journal.pcbi.1003833. Accessed 4 October 2024.

[3] Scanpy: Single-cell analysis in Python, accessed 4 October 2024.