Identification and Characterization of Regenerative Organizing Cells in Xenopus laevis Tails at 0 Days Post Amputation Using Single-Cell RNA Sequencing

Abstract

Regenerative Organizing Cells (ROCs) play a crucial role in the regeneration of tissues in amphibians, such as frogs. This study focuses on identifying and characterizing ROC cells using single-cell RNA sequencing data from Xenopus laevis tails sampled at 0 Days Post Amputation (DPA). The data was preprocessed and analyzed using a combination of clustering algorithms and marker gene identification methods. Principal Component Analysis (PCA) was employed to reduce the dimensionality of the dataset, followed by the application of Leiden and Louvain clustering algorithms. Top marker genes were identified using logistic regression and Wilcoxon rank-sum tests, and the results were validated against a known reference gene list. Several key marker genes, including egfl6, lpar3, cpa6, and sp9, were consistently found to be associated with ROC cells. Clustering metrics such as Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), and Silhouette Score were used to evaluate clustering performance, indicating moderate alignment with known categories. The findings of this study provide a robust framework for identifying and characterizing ROC cells, which could contribute to a better understanding of tissue regeneration mechanisms in amphibians.

Introduction

Amphibians, particularly frogs, possess a remarkable ability to regenerate lost tissues, including limbs and tails. This regenerative capacity is facilitated by a unique population of cells known as Regenerative Organizing Cells (ROCs), which orchestrate the repair and regrowth of damaged tissues. Understanding the molecular characteristics of ROCs is critical for deciphering the underlying mechanisms of tissue regeneration and potentially applying this knowledge to regenerative medicine in other species.

Recent advances in single-cell RNA sequencing (scRNA-seq) have enabled the high-resolution profiling of individual cell types, revealing insights into their gene expression profiles and functional states. This study utilizes scRNA-seq data from Xenopus laevis tails sampled at multiple time points post-amputation to identify and characterize ROC cells. While multiple time points were available, the focus of this study is on the initial time point, 0 Days Post Amputation (DPA), representing the baseline state before the regenerative process begins.

Methods

The original dataset contained single-cell RNA sequencing data from frog tail cells sampled at four different time points: 0, 1, 2, and 3 Days Post Amputation (DPA). This analysis was conducted on cells sampled at 0 DPA, representing the baseline state before the onset of regeneration. Focusing on 0 DPA allowed us to establish a foundational understanding of the gene expression profile of Regenerative Organizing Cells (ROCs).

According to the original paper, the cluster names were mapped to four broader categories: ROCs, Epidermis, Neural Progenitor, and Immune. This was for better analyzing in the biological context and easier to compute clustering metrics. The data was first filtered to remove low-quality cells and genes with low expression. A total of 2000 highly variable genes (HVGs) were selected based on their variance. The data was then log-normalized, and scaling was applied to ensure each gene had unit variance.

To reduce the dimensionality and noise, Principal Component Analysis (PCA) was performed, retaining the top 15 principal components. These components captured the majority of the variance within the highly variable genes. The PCA-transformed data was used to construct a k-Nearest Neighbors (kNN) graph.

Two clustering algorithms were then applied to identify distinct cell populations. The first one is Leiden clustering. The Leiden algorithm was applied to the kNN graph using a resolution parameter of 0.5, which controls the granularity of clusters. This resolution yielded several clusters that aligned with the expected cell populations at 0 DPA. The other one is Louvain clustering. It was used as a comparative clustering method. The same resolution parameter (0.5) was applied to ensure a fair comparison with Leiden clustering. Both clustering methods were visualized using UMAP to provide a low-dimensional representation of the identified clusters. Clustering metrics including RAND-index (RI), Adjusted Rand Index (ARI), Silhouette Score, Normalized Mutual Information (NMI), and Homogeneity Score were computed for each clustering method and can be then compared for choosing a better one among the two.

Finally, logistic regression and Wilcoxon rank-sum test were applied to identify marker genes for each cluster. Logistic regression was used to determine marker genes that could best separate ROC cells from other cell types. The logistic regression model assigns a weight to each gene, indicating its importance in distinguishing ROCs. Wilcoxon rank-sum test is a non-parametric test was used to identify genes that were significantly differentially expressed between ROC cells and other cell types. The Wilcoxon test is robust to outliers and does not assume a specific distribution for gene expression, making it suitable for single-cell RNA-seq data. For both methods, top 50 marker genes were displayed and the goal was to find the overlapping marker genes among them. This result was then compared with the genes listed in Supplementary Table 3 from the original paper.

Code Availability: https://github.com/Quinnie2959/GR5243Project1/blob/main/Project1.ipynb

Results

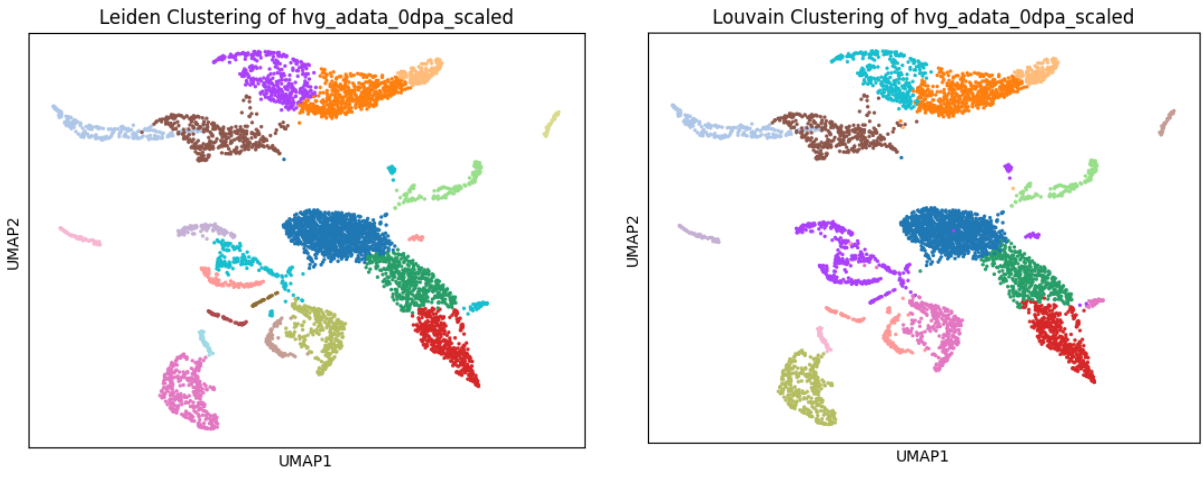

Figure 1. Visualization of Leiden and Louvain clustering using UMAP

1. Leiden Clustering

According to Figure 1, the UMAP visualization of Leiden clustering shows a well-defined separation of clusters at 0 Days Post Amputation (DPA). Each cluster is represented as a distinct

group of points, and the spatial distribution of the clusters suggests that the Leiden algorithm successfully captured the heterogeneity within the cell population at this initial time point.

By creating a contingency table between Leiden clusters and main categories, clustering metrics can be computed. In this situation, it can be obtained that Adjusted Rand Index (ARI): 0.2277, Normalized Mutual Information (NMI): 0.5285, Homogeneity Score: 0.8870, Rand Index (RI): 0.6519, and Silhouette Score (PCA coordinates): 0.1870. The silhouette score for Leiden clustering indicates moderate cluster separation. Clusters that correspond to known cell types, such as ROCs and Neural Progenitors, are well-separated from each other, as shown in the UMAP plot. The clustering pattern reflects the unique transcriptional states of cells at 0 DPA, providing a baseline for further comparison with later time points.

2. Louvain Clustering

Similar to the Leiden clustering, Louvain clustering also resulted in distinct clusters that correspond to various cell types at 0 DPA as shown in Figure 1. However, the UMAP visualization of Louvain clusters shows slightly more compact and cohesive clusters compared to Leiden. This suggests that Louvain clustering may have captured tighter groupings of cells, potentially representing more homogeneous subpopulations.

By creating a contingency table between Louvain clusters and main categories, clustering metrics can be computed. In this situation, it can be obtained that Adjusted Rand Index (ARI): 0.2369, Normalized Mutual Information (NMI): 0.5362, Homogeneity Score: 0.8721, Rand Index (RI): 0.6556, Silhouette Score: 0.1536. The silhouette score for Louvain clustering was slightly lower than that of Leiden, indicating that Louvain clusters, while more cohesive, may have less distinct boundaries compared to Leiden clusters. This might reflect biological reality, where certain cell states are more similar to each other, leading to softer boundaries between clusters.

The tighter clusters formed by Louvain clustering could indicate a finer resolution in capturing the cellular subpopulations present at 0 DPA. For example, clusters representing different subtypes of Neural Progenitors or Immune cells are more compact in Louvain clustering, suggesting that Louvain may be better suited for resolving closely related cell types. Also, according to the computed clustering metrics and the UMAP visualizations, it can be concluded that the two clustering were performing similarly, while Louvain could be said to have a slightly better performance based on its higher ARI, RI, and NMI.

3. Marker Selection and Gene Analysis

After clustering the cells at 0 DPA using both Leiden and Louvain algorithms, marker genes were identified using Logistic Regression and Wilcoxon Rank-Sum Test to ensure the robustness of the identified markers. This step was to characterize each cluster and confirm the presence of distinct cell populations.

Logistic regression was applied to determine genes that best differentiate ROC cells from other cell types. The Wilcoxon Rank-Sum Test, a non-parametric method, was used to identify genes that were significantly differentially expressed between ROC cells and other cell types. There were several overlapping primary ROC markers from logistic regression and the Wilcoxon Rank-Sum such as egfl6, lpar3, cpa6, sp9, nid2, pltp, frem2, bmp2, optc, and snai2. While egfl6, lpar3, cpa6, sp9, nid2, pltp, frem2 overlap with the genes listed in Supplementary Table 3.

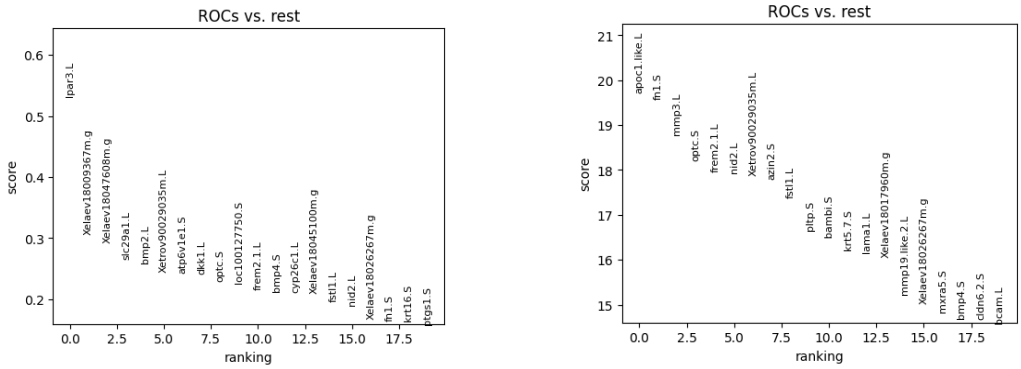

Figure 2. Top ROC Marker Genes Identified Using Logistic Regression (on the left) and Wilcoxon Rank-Sum Test (on the right)

Conclusion

This study successfully identified Regenerative Organizing Cells (ROCs) in frog tail tissues at 0 Days Post Amputation (DPA) using single-cell RNA sequencing. Both Leiden and Louvain clustering algorithms revealed distinct cellular subpopulations, with moderate alignment to known cell types. By comparing clustering metrics, Louvain has a slightly better alignment than Leiden. Marker gene analysis using logistic regression and Wilcoxon rank-sum tests identified key ROC markers, such as egfl6, lpar3, cpa6, sp9, nid2, and pltp. These markers were consistent across both methods and validated against known ROC markers, supporting their role in regeneration. The identified genes are associated with key signaling and cellular processes essential for maintaining the regenerative state.

Overall, our findings provide valuable insights into the molecular characteristics of ROC cells, laying the groundwork for future research on tissue regeneration in amphibians and other species.