STAT5243 Project 1:
Regenerative Organizing Cell in the Frog Tail

Dhruv Jani

October 7, 2024

1 Abstract

This project investigates the Regenerative Organizing Cell (ROC) in Xenopus laevis tadpole tails through single-cell RNA sequencing (scRNA-seq) data analysis. By applying clustering methods, including PCA + Louvain and PCA + Leiden, we identified distinct cell populations, including the ROC, responsible for the tadpoles' regenerative capacity. We used marker gene identification to distinguish ROC from other cell types and performed gene expression analysis. Our analysis highlights specific genes linked to regenerative pathways, which provide insight into how ROCs contribute to regeneration.

2 Introduction

The ability of certain organisms to regenerate lost tissues is a fascinating biological process. Unlike mammals, Xenopus laevis tadpoles exhibit a remarkable ability to regenerate their tails after amputation. This regenerative ability is largely due to the presence of a previously unrecognized cell type called the Regeneration Organizing Cell (ROC), which plays a pivotal role in relocalizing to the amputation site and secreting signals that promote tissue regrowth. To better understand the molecular mechanisms that enable regeneration, this study leverages single-cell RNA sequencing (scRNA-seq) data to identify the ROC and distinguish it from other cell types within the tadpole's tail. By comparing gene expression patterns and clustering cells based on scRNA-seq data, we aim to pinpoint the unique gene markers associated with ROC and uncover insights into the pathways involved in tissue regeneration.

3 Methods

3.1 Data Preprocessing

The single-cell RNA sequencing (scRNA-seq) data used in this study was sourced from the Xenopus laevis tadpole tail dataset after tail amputation. First, we subset the data to include only observations at day 0, focusing on the initial time point. Raw counts were log-normalized, and highly variable genes (HVGs) were selected using the scanpy package. The top 2,000 HVGs were used for further analysis.

3.2 Clustering Analysis

We used PCA to reduce the dimensionality of the data. To uncover cell subpopulations, we employed two clustering algorithms: PCA + Louvain and PCA + Leiden. Both methods revealed distinct clusters, including the ROC population. Clustering quality was assessed using metrics such as the Adjusted Rand Index (ARI) and silhouette score.

3.3 Marker Gene Identification

To identify the genes that distinguish ROC from other cell types, we applied logistic regression-based marker selection and another method to validate marker genes.

3.4 Visualization

UMAP was used to visualize the identified clusters, and the ROC's gene expression patterns were further analyzed to determine the genes unique to this cell type.

3.4.1 Code Availability

The code used for data processing and analysis is available on my GitHub.

4 Results

4.1 Plots

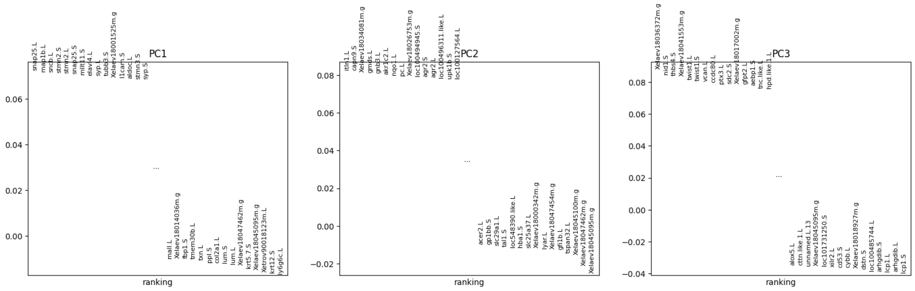

Figure 1: Highly Variable Gene Selection & PCA Anlaysis

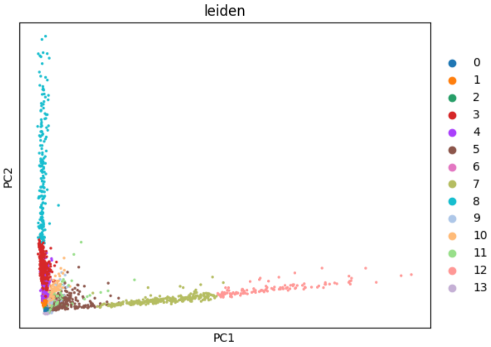

Figure 2: Clustering by Leiden clusters

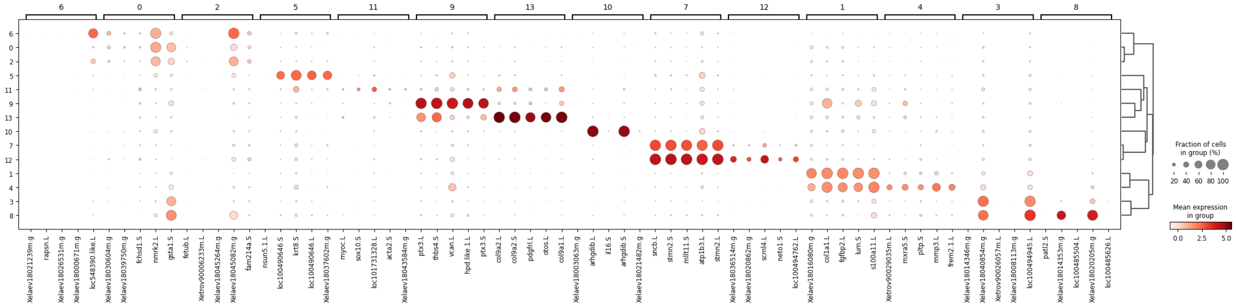

Figure 3: Marker Gene Selection

4.2 Analysis

4.2.1 Clustering Analysis

In this analysis, we used PCA (Principal Component Analysis) and Leiden clustering to segment cells based on their gene expression profiles. The PCA overview (Figure 1) shows the distribution of cells along the first two principal components (PC1 and PC2). Each color in the scatter plot represents a different Leiden cluster, which helps us visually distinguish distinct cell populations. The spread of the cells indicates that most of the variance is captured by PC1, with a small portion captured by PC2, as observed in the highly concentrated points along PC1. This suggests that the major differences between the cell populations are captured within the first component.

4.2.2 Gene Expression Analysis

Gene expression was examined by focusing on the top 2000 highly variable genes across the dataset. The dot plot (Figure 2) shows the mean expression and percentage of cells expressing each gene across different clusters. In this plot, the size of each dot represents the fraction of cells in a cluster expressing a gene, while the color represents the mean expression level. We can infer that certain genes, such as mmp3.1, vcan.1, and krb8.S, are highly expressed in specific clusters, indicating that these genes might play a pivotal role in the functional distinctions between these cell groups.

Furthermore, Figure 3 highlights the ranking of the top genes contributing to the variance in each principal component. For example, in PC1, genes such as smn1.2 and vcam.1 are top contributors, while in PC2 and PC3, different sets of genes (e.g., tleb.1, acad.S, and hpd1_like.1) show higher loadings, implying that these genes drive the variance along those respective components.

5 Conclusion

In conclusion, the analysis effectively identified key gene markers that distinguish the Regeneration Organizing Cell (ROC) from other cell populations within the Xenopus laevis tadpole tail. By applying PCA and Leiden clustering, distinct cell groups were revealed, with specific genes such as mmp3.1, vcan.1, and krb8.S showing high expression in ROC-related clusters. These findings underscore the pivotal role of these genes in regenerative pathways and provide a foundation for further exploration into the molecular mechanisms behind tissue regeneration.

To further validate our findings, additional marker selection methods could be employed to ensure robust identification of ROC-specific genes. Performing a Gene Ontology (GO) analysis could also provide deeper insights into the biological roles of the identified genes across different contexts.