Abstract:

This study utilized a publicly available single-cell RNA sequencing (scRNA-seq) dataset from regenerating Xenopus tail tissue. Data preprocessing included filtering for Day Post Amputation zero samples, applying log-normalization and scaling. Clustering methods, including Leiden, Louvain, and K-means, were implemented to identify distinct cell populations. The results shows that ten prominent markers were identified.

1. Introduction

The capacity of regeneration, which allows for the restoration of lost tissues or body parts has long intrigued researchers. It is usually evident in certain organisms such as amphibians which possess a unique ability to regenerate limbs, tails and also complex structure. Gene expression is the transcription and translation of genetic information into functional proteins or other molecular products that define a cell's role and behavior. By analyzing gene expression, researchers can determine which genes are activated in these regenerative cells, distinguishing them from non-regenerative cells The study aims to identify the Regenerative Organizing Cell (ROC) in the frog's tail and determine the specific genes that distinguish this cell from others in the tissue.

2. Methods

The study used single-cell RNA sequencing (scRNA-seq) dataset which was obtained from publicly available sources and it includes gene expression profiles of individual cells from regenerating Xenopus tail tissue. The dataset consisting of 13,199 rows and 13 columns of single-cell RNA sequencing data including column names. The analysis was conducted using the scanpy library within a Python environment, specifically in Google Colab. Here is the link to the Notebook: https://colab.research.google.com/drive/1snEFXjGTPySRM243a6z_vVT-VwlAsyzw?usp=sharing

2.1. Data Preprocessing

The data processing steps for identifying distinctive marker genes in the frog tail regeneration dataset involved several systematic actions. Initially, the data was filtered to focus on samples from Day Post Amputation (DPA) zero. Log-normalization was applied to adjust for differences in sequencing depth, scaling total counts to a target of 10,000 per cell to ensure comparability. This was followed by log transformation to reduce technical variability and enhance biological relevance. A selection of the top 2,000 highly variable genes (HVGs) was made for further analysis, improving computational efficiency and reducing noise. Finally, the dataset was scaled to center the data and ensure each gene had unit variance, allowing all features to contribute equally to the analysis.

2.2. Code Availability

The code used in this project is publicly available on GitHub at the following link: https://github.com/mch2214/project-1 .This repository includes all code used to reproduce the analyses described in this study.

3. Results:

Clustering methods, including Leiden, Louvain, and K-means was implemented and assessed using Silhouette Scores and the Adjusted Rand Index (ARI). Figure 1a, 1b and 1c provides a summary of the clustering results using UMAP visualization. Among the methods, KNN produced the clearest separation between clusters, while Louvain and K-means showed more overlap in cluster assignments. KNN also achieved the highest Silhouette Score of 0.464, indicating that it formed the most well-defined clusters compared to the other methods. while Louvain and Leiden achieved a score of 0.351 and 0.339 respectively, and the Adjusted Rand Index of Louvain vs. Leiden give an accuracy of 0.817.

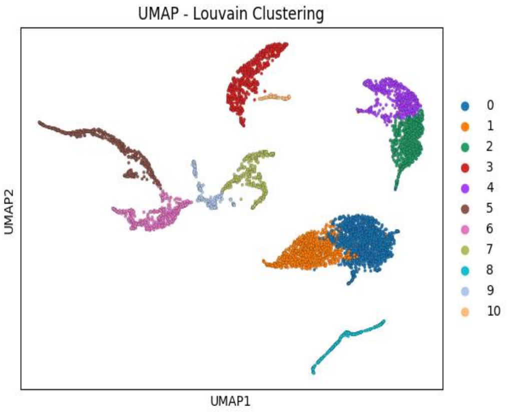

Figure 1a: Louvain Clustering

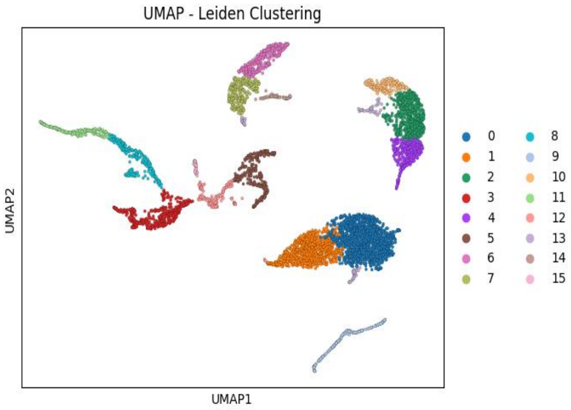

Figure 1b: Leiden Clustering

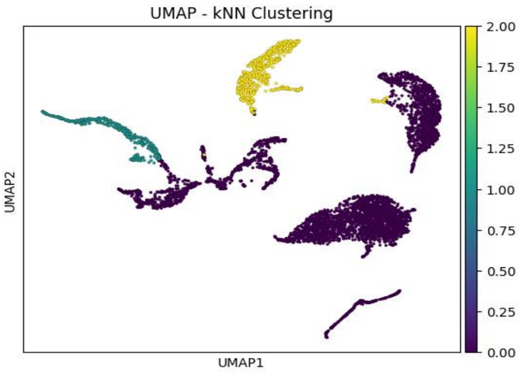

Figure 1c: KNN Clustering

In the gene expression analysis, two distinct methods; Differential Expression and Variance-Based Feature Selection were carried out to identify key marker genes associated with the Regenerative Organizing Cell (ROC). The analysis showed a total of ten important markers, with Xelaev18002241m.g and ca2.L ranking among the highest in scores, indicating their significant roles in the regenerative process.

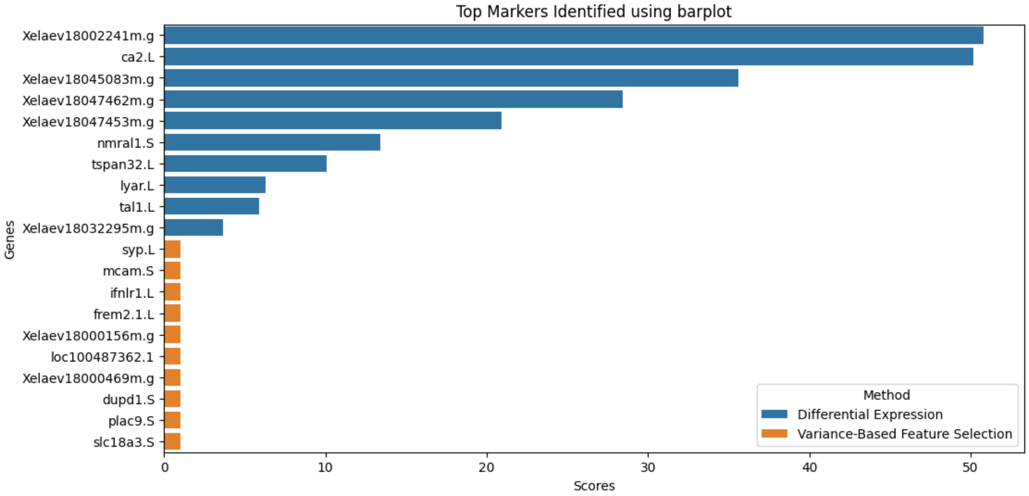

Figure 2: Gene Markers Identification

4. Conclusion:

The exploration of regenerative power in amphibians, particularly frogs, has significant implications for understanding tissue regeneration and potential therapeutic applications in regenerative medicine. This study identifies the Regenerative Organizing Cell (ROC) in the frog's tail using single-cell RNA sequencing data. The analysis, conducted with the scanpy library, involved filtering for Day Post Amputation zero samples and selecting the top 2,000 highly variable genes. Clustering method (Leiden, Louvain, K-means) were evaluated, with KNN showing the best separation and highest Silhouette Score (0.464). Ten key marker genes were identified, including Xelaev18002241m.g and ca2.L.