Frog Tail Regeneration

Qiyuan Zhang (UNI: qz2567)

STAT5243-Applied Machine Learning & Computer Vision for Biologists  
**Abstract**

Single-cell RNA sequencing was used to study tail regeneration in *Xenopus laevis*. After data cleaning, denoising, and clustering with PCA and Leiden methods, a small group of cells showed high expression of *col2a1.L*, *col9a1.L*, and *lect1.L*. These genes are linked to extracellular matrix remodeling and match the Regenerative Organizing Cell (ROC) described in earlier work, indicating this population’s likely role in driving tissue regeneration.

**Introduction**

Frogs such as Xenopus laevis can regrow their tails after they are cut, which makes them a useful model for studying how tissues heal. During regeneration, different cell types become active and change their gene expression as new tissue forms. One group of cells, known as the Regenerative Organizing Cell (ROC), appears to guide this process. However, its exact role and identity are still not fully understood

**Methods**

**Data Acquisition and Pre-processing**. Raw data were downloaded from the ArrayExpress dataset provided via CourseWorks (arrayExpressUpload.zip). Files were unzipped using Python’s zipfile module and processed into an AnnData object using scanpy and anndata.

**Data Denoising**. Two denoising strategies were used:  
(1) Regression-based correction — removed sequencing-depth-related noise (total\_counts, pct\_counts\_mt)  
(2) Scaling and clipping — standardized gene expression and capped outliers at 10.

After denoising, clusters became more compact and biologically meaningful in UMAP.

**Clustering Analysis**. Dimensionality reduction was performed using PCA (30 components), followed by Leiden and Louvain clustering.

**Marker Selection and Gene Analysis**. Marker genes for each Leiden cluster were identified using both t-test and Wilcoxon methods.  
Top 5 marker genes per cluster were visualized, and ROC-specific genes (*col2a1.L*, *col9a1.L*, *lect1.L*) were compared with Supplementary Table 3.

**Code Availability**

<https://github.com/FredaZhang725/FrogROC_project>

**Results**

**Clustering Results**

图示

AI 生成的内容可能不正确。After denoising, PCA + Leiden clustering identified seven cell clusters.  
Regression-based noise removal made clusters tighter and more separable (ARI ≈ 0.82).

**Figure 1. UMAP visualization of single-cell clusters after denoising.**  
Each dot represents a single cell, and colors correspond to Leiden clusters identified after regression-based denoising and scaling.

The embedding reveals 30 distinct clusters (0–29), showing clear separation between major cell populations.  
Clusters are more compact and biologically interpretable after denoising, confirming improved signal-to-noise ratio.

图表

AI 生成的内容可能不正确。

**Figure 2. Differential expression of ROC-associated genes (Cluster 5 vs. rest).**

Cluster 5 showed higher expression of extracellular-matrix-related genes such as *col2a1.L*, *col9a1.L*, *col9a2.L*, and *lect1.L* compared with other clusters. These genes mark the Regenerative Organizing Cell (ROC) population described in the reference study.

**Marker Gene Identification**. Differential-expression analysis identified *col2a1.L*, *col9a1.L*, and *lect1.L* as top markers for one cluster.  
Their expression overlapped with Supplementary Table 3, confirming this cluster as the ROC.

**Conclusion**  
This study identified the Regenerative Organizing Cell (ROC) in frog tail regeneration through single-cell transcriptomic analysis. Regression-based denoising and scaling improved cluster quality, and Leiden clustering highlighted a population expressing ROC marker genes.