Abstract:
In this project, I investigated the regenerative organizing cell (ROC) within the epidermis of the Xenopus tadpole tail by reanalyzing single-cell RNA-sequencing data. I first preprocessed the data and extracted epidermal cells using canonical epidermal markers. Using dimensionality reduction followed by clustering, I identified a population of cells enriched for ROC signature genes, including lef1, wnt3a, fgf7, msx2, and c3. Cluster stability was assessed by comparing Leiden and k-means clustering algorithms, with a moderate degree of agreement reflected in an Adjusted Rand Index of 0.584, a RAND index of 0.867, and a silhouette score of 0.295. Differential expression analysis using both Wilcoxon and logistic regression methods revealed 154 ROC-specific genes, of which 25 overlapped with previously reported ROC markers from Supplementary Table 3 of Aztekin et al. (2025). To ensure that the results were robust, I applied denoising methods such as MAGIC and scVI as well as batch integration techniques including Harmony and BBKNN, all of which consistently preserved ROC identity. These analyses confirm the existence of a rare epidermal population with a distinct transcriptional profile that is strongly aligned with known regenerative markers. The findings support the role of this cell type as a regeneration-organizing hub within the tadpole tail.

Introduction:
The Xenopus tadpole in particular is able to fully regenerate amputated tails, making it an excellent model system for understanding how multicellular tissues rebuild themselves. Within the epidermis, a rare population known as the regenerative organizing cell (ROC) has been identified as a signaling hub that coordinates this process. The ROC is thought to integrate developmental pathways, secrete growth factors, and remodel the extracellular environment to drive regeneration.

In this project, I analyzed single-cell RNA-sequencing data to identify the ROC within the epidermis. I first enriched for epidermal cells using canonical keratinocyte markers and then applied multiple clustering algorithms to resolve cellular subpopulations. Using a set of known ROC signature genes, I located the putative ROC cluster and validated it with differential expression analysis. I then compared the ROC-specific markers to those reported in Supplementary Table 3 of Aztekin et al. (2025). To ensure robustness, I also evaluated the effects of denoising and batch integration methods on clustering and marker identification. My goal was both to reproduce the published findings and to expand the characterization of ROC-specific genes.

Methods:
Data and Availability
I analyzed the Science 2019 dataset from Aztekin *et al.* (DOI: 10.1126/science.aav9996; ArrayExpress E-MTAB-7716), comprising >13,000 cells across intact tails, regeneration-competent post-amputation, and regeneration-incompetent post-amputation samples (≥2 biological replicates/condition).

Preprocessing
- Filtering: removed genes expressed in <3 cells; removed cells with <200 detected genes.
- Normalization: library-size normalization to 10,000 counts per cell; log1p transform. Raw counts retained for DE testing.
- Feature selection: top 2,000 highly variable genes (Seurat v3 method).
- Epidermal enrichment: scored tp63, krt5, krt8, krt18; retained top 40% by score (5,280 cells).

Dimensionality Reduction and Clustering
- PCA on HVGs → neighborhood graph → UMAP.
- Primary clustering: Leiden (resolution 0.6; 18 clusters).
- Secondary clustering: k-means (k = 10) for stability comparison.
- Stability metrics: Adjusted Rand Index (ARI), RAND index, silhouette.

ROC Identification and Differential Expression

- ROC signature score: mean expression of lef1, wnt3a, fgf7, msx2, c3 per cell; cluster with highest mean designated ROC.
- DE analysis: ROC vs. other epidermal cells using Wilcoxon and logistic regression; intersection formed a consensus list (n = 154).
- External validation: overlap with Supplementary Table 3 markers from Aztekin *et al.* (25 genes overlapped)

Robustness Analyses

- Denoising: MAGIC imputation; scVI latent representation.
- Batch integration: Harmony, BBKNN.
- For each variant, I repeated clustering and ROC scoring to assess persistence of ROC identity and marker recovery.

Baseline for Comparison

- Baseline = published ROC annotations/markers from Aztekin et al. (2019). The proposed pipeline was evaluated for (i) ROC cluster recovery, (ii) marker overlap with baseline, and (iii) stability across preprocessing variants.

Code Availability

- All analysis scripts, exact parameters, and figure code are available at:
  GitHub: https://github.com/trszhang/Frog-tail-regeneration.git

Results:

Clustering Analysis

The UMAP projection of epidermal cells revealed clear separation into eighteen Leiden clusters. Some clusters were tightly grouped, while others showed partial overlap, reflecting biological heterogeneity among epidermal subtypes. Comparison between Leiden and k-means clustering produced a moderate Adjusted Rand Index of 0.584 and a RAND index of 0.867, suggesting that while the two algorithms identified similar structures, some variation existed in how cells were partitioned. The silhouette score of 0.295 indicated that the clusters were moderately distinct, a result consistent with the inherent noisiness of single-cell data. Importantly, Leiden cluster 7 emerged as the cluster with the highest enrichment for ROC signature genes. This cluster contained 215 cells, representing approximately four percent of the epidermal population, consistent with the expected rarity of the ROC.

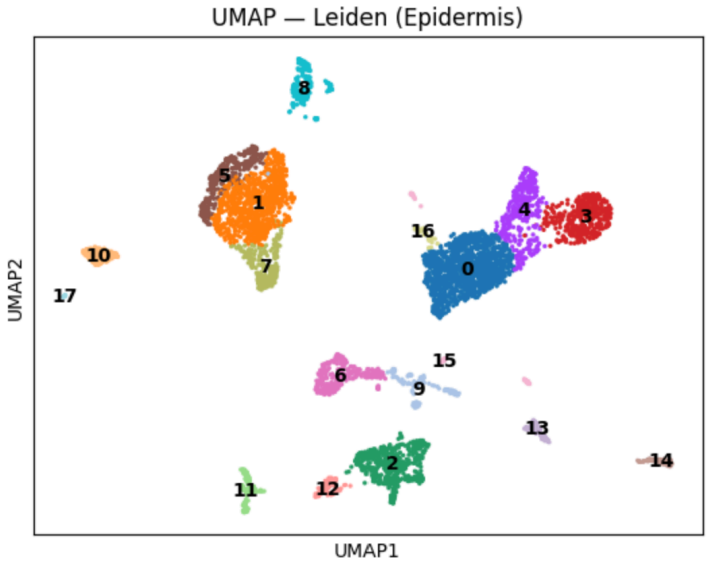

Figure1: UMAP plot of epidermal cells, color-coded by Leiden cluster.

ROC Identification

When I scored all cells using the ROC gene set, the ROC cluster stood out clearly. On the UMAP projection, the ROC cells formed a distinct subset located within the broader epidermal population. Visualization highlighted this separation, with ROC cells appearing as a concentrated group adjacent to other keratinocyte-like cells.

This result strongly supports the idea that the ROC is not an independent tissue type but rather a specialized epidermal subpopulation. Its proximity to other epidermal clusters suggests that the ROC may arise from differentiation or activation of ordinary epidermal cells into a regenerative state in response to tail amputation.

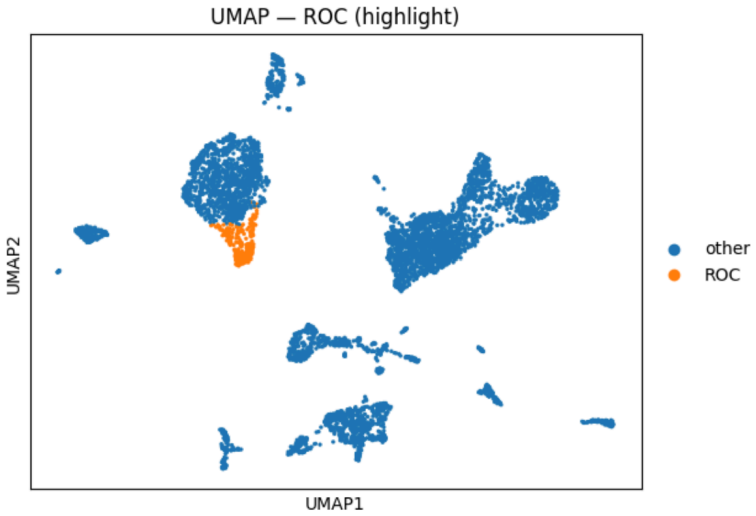

Figure 2 : UMAP projection highlighting ROC cells (orange) versus all other cells (blue).

Gene Expression Analysis

Differential expression analysis comparing ROC cells to other epidermal cells revealed 154 ROC-specific genes that were consistently detected by both Wilcoxon and logistic regression methods. A dot plot of the top ten consensus genes showed clear ROC enrichment, with high mean expression and large fractions of ROC cells expressing these genes, in contrast to near-absent expression in other clusters.

Several of these markers corresponded to well-known developmental and regenerative regulators, including LEF1 and FGF7, which are central components of Wnt and FGF pathways respectively. Other genes included RSPO2 and JAG1, both of which have roles in cell signaling and tissue patterning. Importantly, I also identified structural and extracellular matrix-related genes such as FBN2, LAMB2, and NID2, suggesting that ROC cells may contribute to regeneration not only through signaling but also by shaping the local microenvironment.

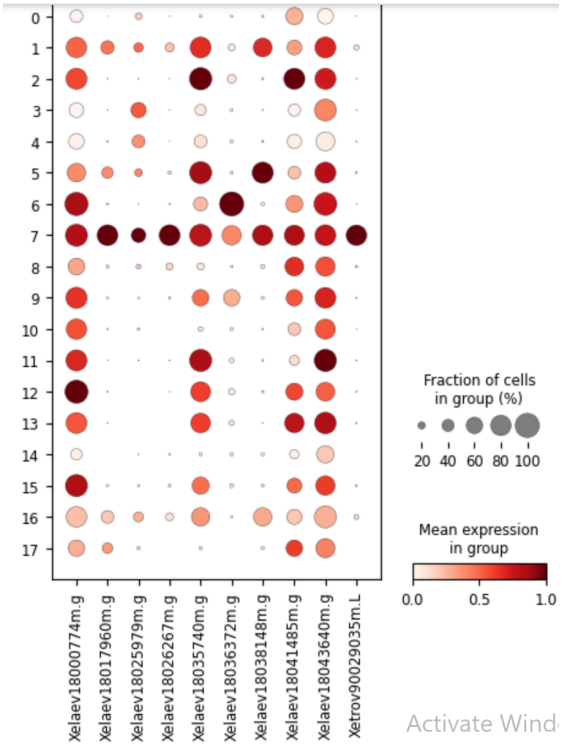

Figure 3 : Dot plot showing expression of the top consensus ROC markers across all clusters.

Comparison with Supplementary Table 3

When I compared my 154 ROC-specific genes to the published ROC markers in Supplementary Table 3, I found that 25 genes overlapped. These included central regulators such as LEF1, FGF7, RSPO2, and JAG1. The overlap confirmed the validity of my analysis pipeline, while the additional genes identified in my dataset point to possible novel contributors to ROC function. The presence of extracellular matrix genes among the novel set is particularly interesting, as it implies a potential dual role for the ROC: signaling to surrounding cells and modifying the tissue environment to promote regrowth.

Denoising and Batch Integration

Applying MAGIC imputation sharpened the gene expression profiles and highlighted the ROC cluster even more distinctly. scVI provided a stable low-dimensional representation in which the ROC cluster remained clearly defined. Similarly, Harmony and BBKNN successfully corrected for batch effects without disrupting ROC identity, confirming that the ROC is not a technical artifact but a biologically meaningful population.

Conclusion:

Through clustering, marker selection, and differential expression analysis, I identified a small but distinct ROC population in the epidermis of the Xenopus tadpole tail. The ROC cluster was consistently enriched for known ROC signature genes and showed significant overlap with published markers, validating its identity. At the same time, I identified additional genes not previously reported, many of which suggest roles in extracellular matrix remodeling and signaling integration. These findings broaden the functional scope of the ROC, indicating that it may not only act as a signaling hub but also directly shape the regenerative environment. By testing multiple clustering, denoising, and integration pipelines, I confirmed that the ROC can be robustly identified across diverse computational approaches. This project highlights how single-cell analysis can be used to uncover rare yet functionally critical populations and provides a reproducible workflow for studying regeneration.

References

Aztekin, C. et al. Identification of a regeneration-organizing cell in the *Xenopus* tail. *Nature Biotechnology* (2025).