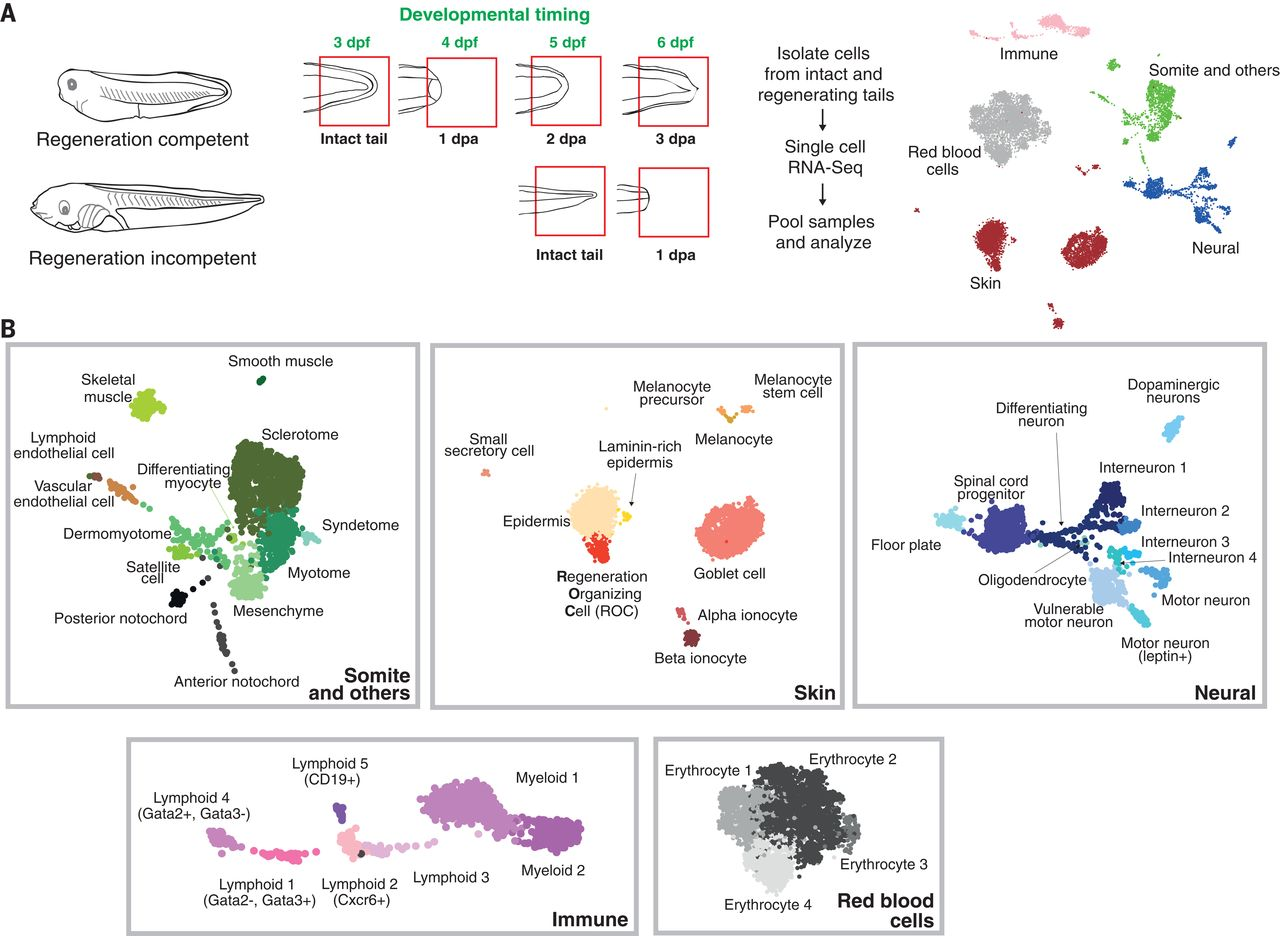
Mini Project 1: Find the frog’s source of regenerative power

(Due October 2nd)

**Project link**: <https://drive.google.com/drive/folders/1fXvEc8HQ-JUPMNK_7fzwdE36OYB5KIg6>

**Prompt**: Find the Regenerative Organizing Cell in the Frog tail (Figure 1, Panel B (Skin). Describe the data processing protocol that has allowed you to identify it. Find the genes that make this cell different from all other cells and compare this set of genes with the ones from Supplementary Table 3.



(Figure adapted from [1])

**Reference Materials:**

* **Main Paper**: [1] [Identification of a regeneration-organizing cell in the Xenopus tail](https://www.science.org/doi/full/10.1126/science.aav9996)
* **Supplementary material:** [Supplementary Material](https://www.science.org/doi/suppl/10.1126/science.aav9996/suppl_file/aav9996_aztekin_sm.pdf)
* **Data:** Available to download from the journal website, but we have added [here](https://drive.google.com/file/d/1boRVP8VCptxxvfEEjlg8OW4xUFQbgn--/view?usp=drive_link) too
* **Starter kit**: [Frog and tail.ipynb](https://colab.research.google.com/drive/1f__Fcj14ZUi9oXbvNZqFqC3mBUAf-CNh?usp=sharing) (includes code for reading and processing the data)
* **Additional reading**: [Frog and tail.ipynb](https://colab.research.google.com/drive/1f__Fcj14ZUi9oXbvNZqFqC3mBUAf-CNh?usp=sharing)[Tutorials Scanpy](https://scanpy.readthedocs.io/en/stable/tutorials/index.html)
* **Help in choosing colors for clustering:** <https://color.adobe.com/create/color-wheel>
* **Good figures**: [Ten Simple Rules for Better Figures | PLOS Computational Biology](https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833)
* **More information on format:** <https://www.nature.com/ncomms/submit/article>
* **Putting together your report:** <https://www.overleaf.com/>
* **Github help:** [Getting started with your GitHub account - GitHub Docs](https://docs.github.com/en/get-started/onboarding/getting-started-with-your-github-account)

### **Expectations and Suggested Steps:**

**Writeup Requirements:**

* **Abstract:** A concise summary of your findings.
* **Introduction:** Brief background information on the problem
* **Methods:** Detailed description of the data processing and analysis steps. Methods should have a section named *Code Availability*, containing a link to the public github account where the code is available.
* **Results:** Presentation of the clustering and gene analysis analysis. What did you find? It should include references to two figures, one summarizing the clustering results, the other one representing the gene expression analysis results (more information below)
* **Conclusion:** Summary of your findings

**Figures and Analysis:**

1. **Clustering Analysis:**
   * Use at least two clustering algorithms (e.g., PCA + Louvain, PCA + Leiden, kNN).
   * Generate figures to visualize the clusters.
   * Compute clustering metrics (e.g., RAND-index, silhouette score, ARI).
   * Optionally, explore a metric not discussed in class for bonus points.
2. **Marker Selection and Gene Analysis:**
   * Use at least two marker selection methods to identify genes that define the ROC.
   * Compare the identified markers across different methods.
   * Compare your gene set with the genes listed in Supplementary Table 3.
   * Optionally, perform Gene Ontology (GO) analysis to explore the biological roles of these genes in other contexts.

**Suggested Steps for Analysis:**

1. **Data Visualization:**
   * Use UMAP or t-SNE to visualize the data and adjust parameters to replicate the visuals of Figure 1B from the main paper.
2. **Clustering:**
   * Perform clustering using different methods learned in class.
   * Compare the clustering results using the metrics discussed.
3. **Marker Identification:**
   * Use logistic regression or other methods to identify marker genes.
   * Annotate the clusters using the markers identified.

**Output:**

* **Project Write-up:** less than 4 pages, see above for additional directions
* **Colab notebook:** it should be able to run in a Google Colab environment (available through Columbia University. A link should be provided in the Write-up.