

Project 1: Frog Tail

Data analysis to understand tail regeneration.

Elias Ludviksson

STATGR5243 - Applied Data Science

Columbia University

October 11, 2025

Abstract

[Write a brief summary of your project, main findings, and conclusions.]

Contents

1	Introduction	2
2	Methods	2
2.1	Code Availability	2
3	Results	2
4	Conclusion	2

1 Introduction

Some *Xenopus laevis* tadpoles have tails which can regenerate after amputation. The tails of these tadpoles contain a variety of cell types, including muscle cells, skin cells, and spinal cord cells. A study by C.Aztekin et al found the regeneration-organizing cells (ROCs) in the tadpole tails, which are crucial for tail regeneration. They performed single-cell RNA sequencing on these tadpoles to study the gene expression profiles of different cell types during tail regeneration. The dataset is publicly available <https://ftp.ebi.ac.uk/biostudies/fire/E-MTAB-716/E-MTAB-7716/Files/arrayExpressUpload.zip>[online]. In order to learn the clustering and gene analysis techniques people use on single-cell genomic data. We will analyze this dataset using scanpy, scikit, pandas and the numpy libraries in order to find meaning in the dataset.

2 Methods

[Describe the data, preprocessing steps, algorithms used, and analysis workflow.] The data is AnnData formatted and contains $13,199 \text{ cells} \times 31,535 \text{ genes}$.

2.1 Code Availability

The code for this project is publicly available at:

https://colab.research.google.com/drive/1HZrv70DnstypcYv1FyD7D5xdXL_oS01C#scrollTo=71a41724

3 Results

[Present your findings. Reference two figures: one for clustering results, one for gene expression analysis.]

Figure 1: Summary of clustering results.

Figure 2: Gene expression analysis results.

4 Conclusion

[Summarize your main findings, their implications, and possible future work.]