STAT 5243 Mini Project 1: Finding the frog's source of regenerative power

Lilly Enekes

October 2024

Abstract

This paper finds that regenerative-organizing cells in the frog/tadpole tails can be identified using single-cell RNA-sequencing data by using clustering and gene market selection techniques. By comparing marker genes with those from Supplementary Table 3, distinct gene signatures fgfbp2.L and gsta1.S cells are found. This analysis uses PCA, Leiden clustering, and differential gene expression techniques to find these key marker genes.

Introduction

The "Identification of a regeneration-organizing cell in the Xenopus tail" article explains that this aspect of regenerative biology in amphibians is a good example to compare to the way that tissues repair and regrow. The Xenopus laevis tadpoles can regenerate their tails and so their process of doing so is a good prototype to evaluate when studying cellular mechanisms of regeneration. The aim of this project is to find the genes that make this regenerative organizing cell different from all other cells and how they compare to this one. Identifying this cell is important because it is theorized that the cell is the main contributing part of conducting this regeneration. With the single-cell RNA sequencing data referenced in the article, this cell can be found using data clustering and differential gene expression analysis.

Methods

The beginning of our analysis is inspired from the starter code provided for the project and homework assignments. This means that we begin by processing and loading the data to see the information aligned with the days post amputation. These days are also graphed in a histogram to show the decrease of the tail growth and then the upward increase of it as it regenerates at around day 3. The next part of the analysis shows the PCA whitening for Leiden clustering.

After that, we go into the gene selection with the top 10 genes in each category. The next part of the analysis shows the K-Means clustering using t-SNE and UMAP for all of the days of regeneration. The final part of the analysis focuses on marker gene identification to identify the gene marker for the Regenerative Organizing Cell (ROC) in the frog tail using violin plot to see the expression distribution of specific marker genes across clusters.

Code Availability

Here is a link to the public Github account where the code is available: `https://github.com/le2363/STAT5243`.

Results

The below figure is a scree plot that indicates that 7 of the principal components captured sufficient variance as they are the clusters included in the 0-5 ranking.

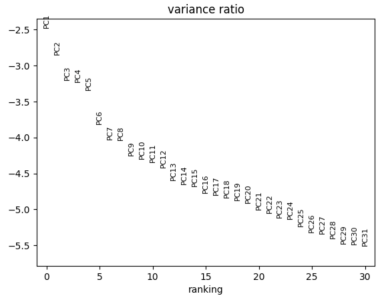

Figure 1: PCA plot of Leiden clusters

I also analyzed using the K-means and t-SNE and UMAP for all the different types of genes discussed (highly variable, mean expression, variance, and random). The clustering can be seen below. I had some trouble with the code when it came to not including day 0, which should not really be included since the text clearly states that regeneration doesn't really happen until after the first 24 hours post amputation. So, that blue dot should try to be ignored on most plots. But, we do see that with the highly variable genes and their variance, they are mainly present within those first 24 hours whereas the mean expression of the genes occurs during the second or third day of regeneration. The random genes are where the cluster starts to show a more varied amount of days, although mainly still staying within the first 0-2 days post amputation. These graphs show us that UMAP creates way more distinct clusters than t-SNE does, so that is the better clustering method to be used. It also shows us that since the Variance and Highly Variable genes were the ones with the most uniform dates (being in day 0), they are more likely to be important when responding

to external stimuli such as regeneration in the Xenopus tail.

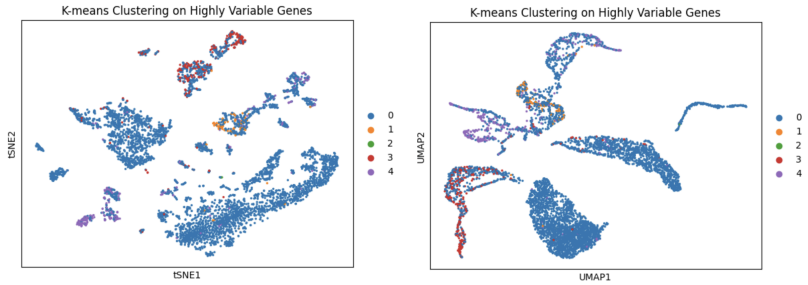

Figure 2 and 3: Cluster of Highly Variable Genes using K-Means and t-SNE (left) and UMAP (right)

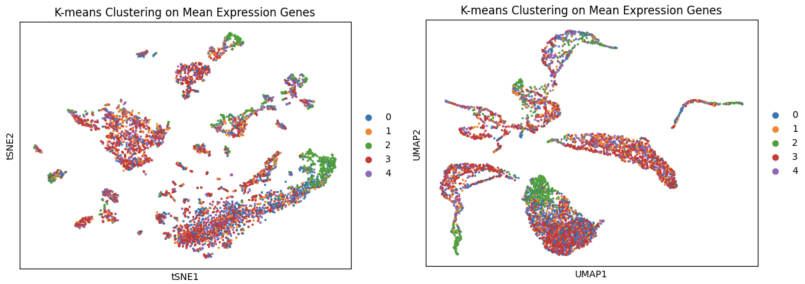

Figure 4 and 5: Cluster of Mean Expression Genes using K-Means and t-SNE (left) and UMAP (right)

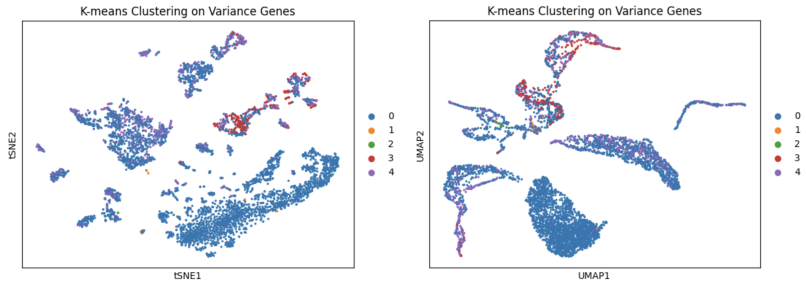

Figure 6 and 7: Cluster of Variance Genes using K-Means and t-SNE (left) and UMAP (right)

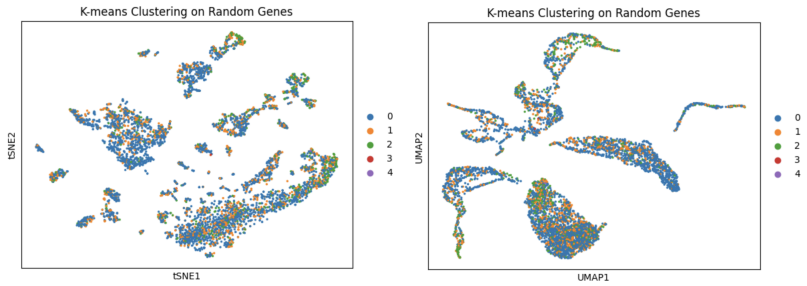

Figure 8 and 9: Cluster of Random Genes using K-Means and t-SNE (left) and UMAP (right)

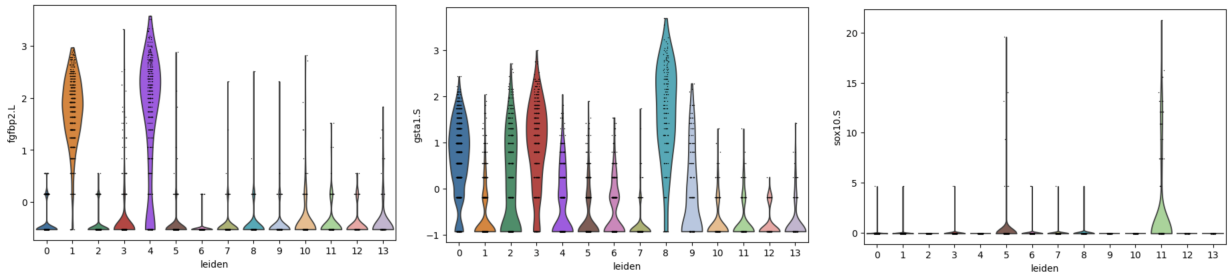

Figure 10, 11, and 12: Violin Plots of Genes (fgfbp2.L, gsta1.S, and sox10.s)

The above violin plots show us the information from 3 different genes. The leftmost one is fgfbp2.L which aligns with Fibroblast Growth Factor Binding Protein 2. I chose to look at this gene because it is known for modulating the activity of firboblast growth factors which are important for cell growth, proliferation, and tissue repair. We discussed before that these qualities apply to regeneration. Since the expression of this gene is only high for about 2 Leiden clusters, this gene may be a marker for cells that are actively taking part in the regenerative process. The middle gene is gsta1.S which is Glutathione S-transferase alpha 1. It is known for being involved in detoxification and protecting cells from oxidative stress by catalyzing the conjugation of glutathione to toxins. This means that since the majority of Leiden clusters are highly expressed, those cells are undergoing oxidative stress or have an increased need for detoxification processes which could be due to an injury or cellular damage during regeneration. The rightmost gene is sox10.s which is a transcription factors in development and maintenance of neural crest ells and is associated with neuronal and glial cell differentiation. Since there is mostly low or no expression of the gene in the clusters, this means that it is not very active in non-neural cell types during regeneration. With these plots, we are able to see that two of the cells fgfbp2.L and gsta1.S, may be involved in the regeneration of Xenopus tails. (These are cells found in Supplementary Table 3).

Conclusion

Using key marker genes identified through differential expression analysis, we have found regeneration-organizing cells using single-cell RNA sequencing data and clustering methods. These genes aligned with the ones in Supplementary Table 3 and confirmed their contribution to the regenerative processes.