Kameneva Lab Data analysis task

Zafer Kosar

April 25, 2025

Abstract :

Kameneva Lab single-cell data analysis task includes a series of single-cell RNAseq datasets that span from week 6 to week 12 of the development of human adrenal glands. These datasets were shown to provide sufficient information on developmental transitions in addition to the gene regulatory programs in adrenal gland development. Here, utilizing multiple single-cell analysis tools and approaches, I demonstrate that Schwann cell precursors in the adrenal gland can potentially give rise to Sympathoblasts. Moreover, Sympathoblasts can transition into chromaffin cells. I also highlight the gene expression dynamics — the possible drivers — of these transitions. Overall, the findings of the task align well with the original study.

Index Terms: Neuroblastoma Origin, Single-cell Transcriptomics, Trajectory Analysis

1 Methods

1.1 Quality-Control, Filtering, Integration

As I am much more proficient in the Python single-cell ecosystem than R's, I started the task with tools like Scanpy and scVI. However, even after a substantial amount of time of trial-and-error, it became clear that I needed to switch ecosystems.

After switching to the R ecosystem, except for memory issues that were resolved by downsampling each dataset to 25% of their original sample size, the integration worked seamlessly after quality-control and filtering steps in Seurat. Integration was validated via visual inspection of the UMAP (Figure 1).

For initial filtering, genes that appear in fewer than 3 cells and cells that have less than 200 genes were filtered out for each dataset. Additionally, cells with fewer than 500 and more than 6000 gene counts were filtered out to remove potentially empty or doublet cells, respectively. Finally, cells with 15% or more mitochondrial genes are filtered out.

Cell cycle scores, using cell cycle markers built into Seurat, were added to the data. Variations in cell cycles were then regressed out during scaling.

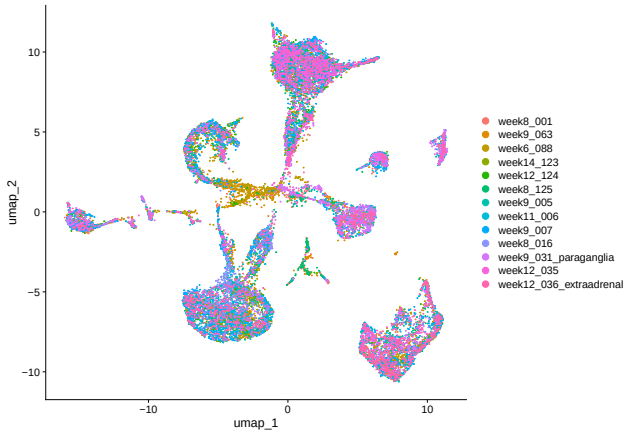


Figure 1. UMAP after integration of the datasets.

1.2 UMAP and Cluster Annotation

UMAP calculation had already been performed after integration using the standard Seurat workflow. Then, clustering has been performed. The cluster annotations have been conducted using markers from Kameneva et. al. 2021, Figure 1b, and visualized on a UMAP plot (Figure 2). Nevertheless, some markers, PAX2 (kidney), HNF4A and AHSG (Liver), AZU1 (HSCs) were absent in the data. Therefore, automatic annotation could not be performed. Instead, manual annotation of clusters through inspecting the marker gene expression of clusters (dot plot, feature plots) was performed. Upon annotation, 5 genes that are differentially expressed for each cluster were visualized using a dot plot (Figure 3).

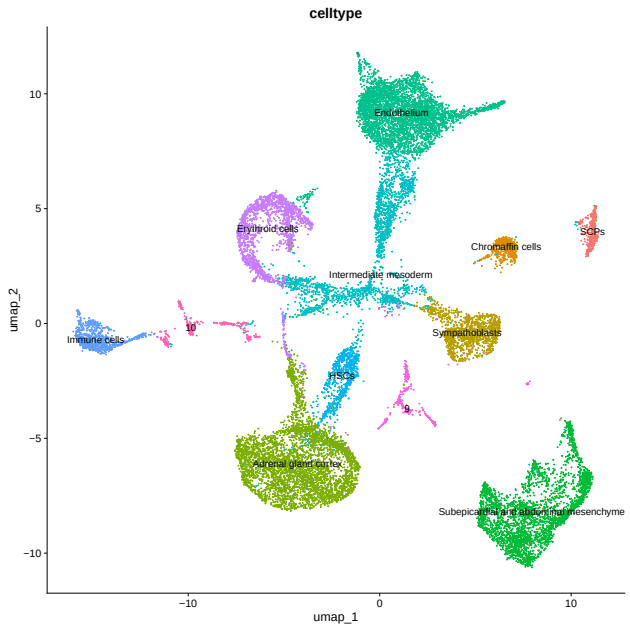


Figure 2. Annotated Clusters.

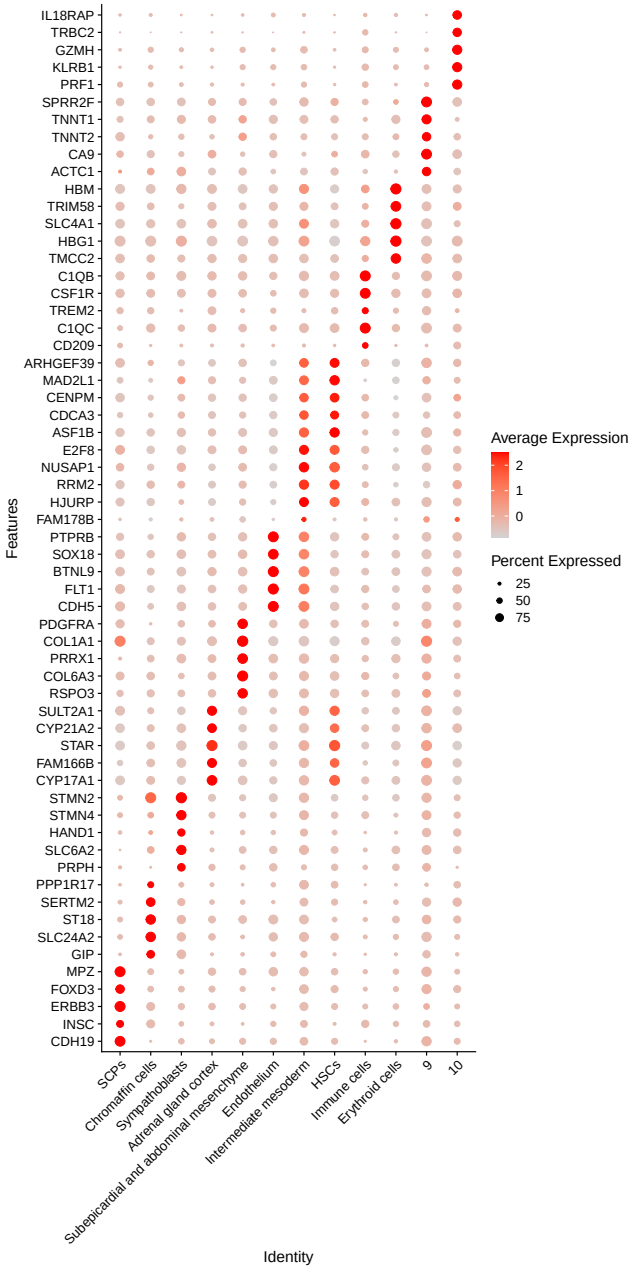


Figure 3. Differentially expressed genes of the clusters.

1.3 Subsetting Adrenal Medulla Clusters

The original data was subsetted to include only SCPs, Chromaffin cells, and Sympathoblasts. The subset was reclustered to improve the resolution of transitions for the upcoming analysis. Clusters were again annotated according to the markers provided in Kameneva et. al. 2021, Figure 2a.

1.4 Trajectory Analysis

For trajectory analysis, I first attempted to utilize monocle3 of the R single-cell data analysis. Monocle3 is an excellent tool for trajectory analysis due to its graph-based approach to trajectory learning and the capacity to handle complex branching. Moreover, monocle3 is well documented, and it has many example use cases due to its popularity. However, after assigning pseudotimes to cells, I was not able to extract the significant gene expression changes along the trajectory.

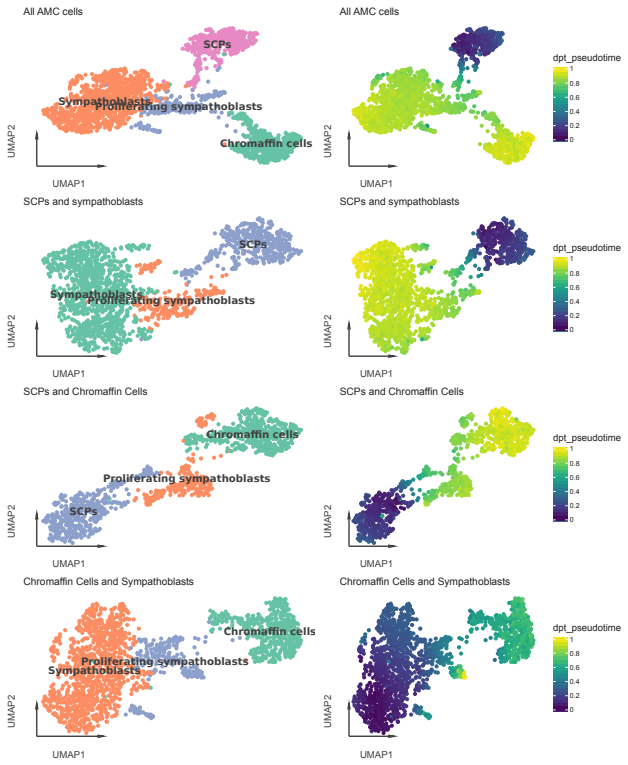
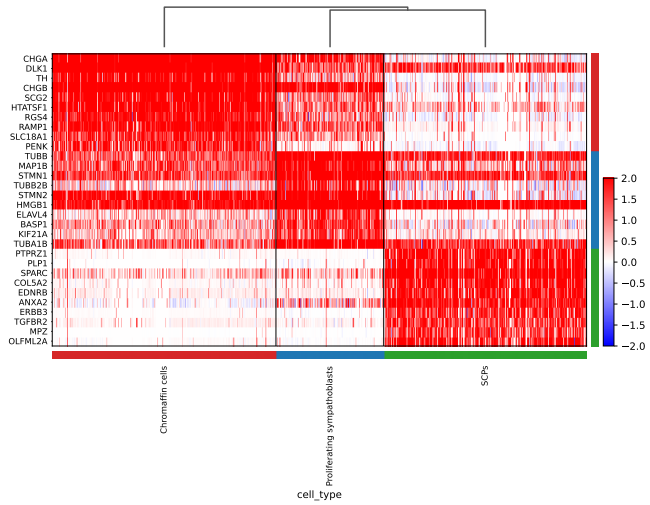


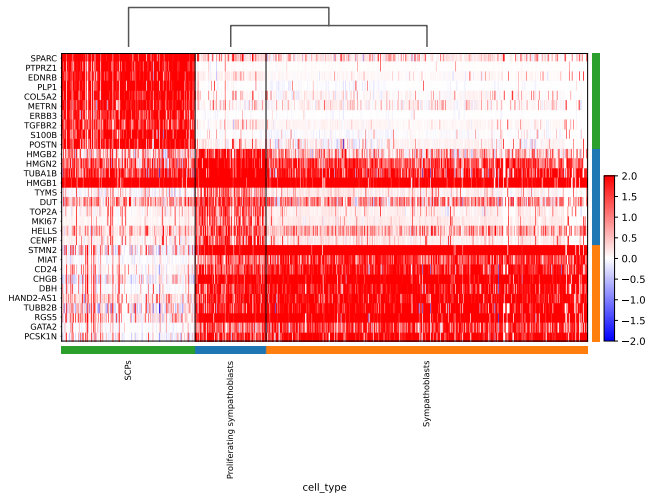
Figure 4. Annotated Clusters.

At that point, I decided to switch to the Python ecosystem, as there are great tools for trajectory analysis within that ecosystem, such as scVelo and paga. Indeed, scvelo requires spliced and unspliced RNA information for such analysis, which this data does not contain. PAGA (Partition-based Graph Abstraction), included in Scanpy, is a scalable tool that allows cluster-level trajectory inference, which is indeed needed here.

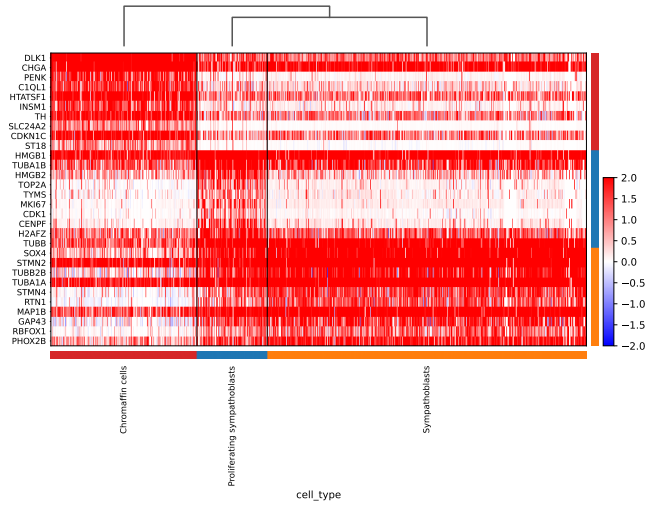
Utilizing PAGA, I was able to get pseudotime trajectories between SCPs and chromaffin cells, SCPs and sympathoblasts, and chromaffin cells and sympathoblasts. Here, I had to opportunity to utilize my own single-cell data visualization tool **Cellestial** (Figure 4). Additionally, PAGA allowed me to obtain differentially expressed genes along the trajectories, which were then visualized



(a) Differentially expressed genes in SCPs and chromaffin cells.



(b) Differentially expressed genes in SCPs and Sympathoblasts.



(c) Differentially expressed genes in Chromaffin cells and Sympathoblasts.

Figure 5. Heatmaps of differentially expressed genes along the trajectories.

on heatmaps (Figure 5a,5b,5c).

2 Discussion

To address task 7c, there are conclusions to draw both from the pseudotime plots of the cells and heatmaps of the differentially expressed genes along the transitions. Firstly, SCPs appear to be the progenitor of both Sympathoblasts and Chromaffin cells (Figure 4). Furthermore, Sympathoblasts might be the progenitors of the Chromaffin cells, establishing a cell lineage from SCPs to Chromaffin cells.

The significant changes in gene expression profiles of the cells along the transition trajectory from SCPs to other cell types, and the transition from Sympathoblasts to Chromaffin cells, reveal important cellular program shifts along such trajectories (Figure 5a,5b,5c). Remarkably, those genes are not only expressed differentially as a result of such transitions, but they are possibly the drivers of cell type transitions in Adrenal Medulla.

In all trajectory analyses, proliferating Sympathoblasts were also employed as they seem to have a role in transitions. This transitional role becomes apparent through the dendograms provided via the heatmaps of the gene expressions (Figure 5a,5b,5c).

It should be noted that there are some minor and some notable differences from the original work (Kameneva et al., 2021). One significant difference is that, due to computational resource limitations, downscaling of the datasets has been performed. This downscaling is expected to influence all the downstream workflow. Nevertheless, there were sufficient numbers of cells to conduct the downstream analysis.

Surprisingly, the datasets did not include some markers that are present in the original work. Markers such as PAX2 (kidney), HNF4A and AHSG (Liver), AZU1 (HSCs) were not present in the data despite a very loose filtering of genes. This difference made manual annotation necessary for cluster annotations.

Overall, the workflow for the task could be completed, and the possible transitions from SCPs to Sympathoblasts and Chromaffin cells were shown, in addition to revealing important cellular programs that drive those transitions. Therefore, the results of this task are in agreement with the original work.

References

Kameneva, Polina et al. (Apr. 2021). "Single-cell transcriptomics of human embryos identifies multiple sympathoblast lineages with potential implications for neuroblastoma origin". In: *Nature Genetics* 53.5, pp. 694–706. ISSN: 1546-1718. DOI: 10.1038/s41588-021-00818-x. URL: http://dx.doi.org/10.1038/s41588-021-00818-x.