Kameneva Lab

Data analysis task

Intro

Some pieces of evidence point to the fact that the origin of pediatric cancer is rooted in developmental processes. The most common extracranial pediatric tumor is neuroblastoma. In 49% of cases, this tumor is found in the adrenal gland area. To shed light on potential reasons why the adrenal gland is one of the tissues of origin for neuroblastoma, we have generated a series of single-cell RNA sequencing datasets covering the development of human adrenal glands across several stages. These datasets should provide us with information about important developmental transitions and programs in adrenal gland development.

Goal

We would like to learn which cell states are present in developing adrenal glands during the collected stages, and to get the information about major transcriptional programs guiding the transitions between these stages.

You can use any programming language, software, and methods you like and consider appropriate. There are many tools/packages out there specifically designed for the analysis of this type of data.

Task includes

1. Download the datasets following the link: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE147821

   Sample name includes the developmental stage in weeks and the internal ID. week8_001 – this sample is collected from week 8 of development and ID is 001. Some developmental stages have replicates.

2. Perform data normalization (consider cell cycle correction) and quality control, data cleaning.

3. Generate the UMAP and perform cluster annotation.

   Take a note of the clusters on Figure 1b https://www.nature.com/articles/s41588-021-00818-x for marker genes.

4. Visualize the data on the dot plot showing the 5 top differentially expressed genes per cluster

5. Sub-select adrenal medulla clusters (Schwann cell precursors (SCPs), Chromaffin cells, Sympathoblasts) and re-cluster them to improve the resolution of transitions.

   Similar to re-clustering on Figure 2 https://www.nature.com/articles/s41588-021-00818-x. Note that not all datasets may be included in the re-clustering.
   If you choose to do so, explain why that may be necessary.

6. For the last task, you can choose between two alternatives

   6a. Perform the trajectory analysis between SCPs and chromaffin cells, SCPs and sympathoblasts, and between chromaffin cells and sympathoblasts.
   Plot the important gene changes along the trajectories on heatmaps.
   You can use any trajectory analysis tool. Please explain your choice.

   6b. Using the SCENIC tool, analyse the regulons in SCPs, chromaffin cells, and sympathoblasts.
   Visualize important regulons whose activity spans the transitions between the clusters. Compare the regulon's expression to the expression of the corresponding transcription factors.

7. Generate the report:
   7a. Please provide annotated code, documentation, results, and plots organized in a PDF file.
   7b. Please provide short explanations of why you chose specific methods of analysis.
   7c. Please provide a short, plausible interpretation of the results of task 6.
   7d. Please provide a small abstract of the overall results of the analysis (max 150 words).