ASSIGMENT GROUP4 – SINGLE CELL MULTIOMICS

Daniel Soto
Laura Sudupe

The integration of multi-omic datasets offers insights into the work of cellular mechanisms. We undertake a comparative analysis between authentic multi-omic datasets vs GEX + ATAC.

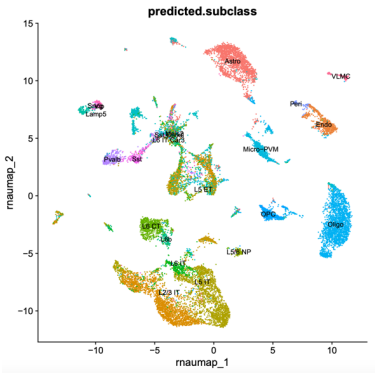
This report shows the differences of synthetic multi-omic datasets in mirroring the complexities captured by genuine multi-omic analyses. By analyzing the clustering and the cell type annotations, we aim to discern the differences between the datasets. the field of systems biology.

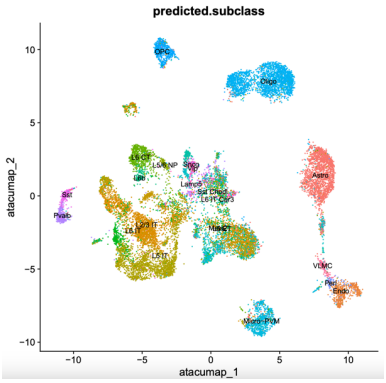
- How is the clustering/annotation of the cell types?

The UMAP plots for both the Multiome and GEX + ATAC datasets reveal distinct clustering patterns, indicative of cellular heterogeneity within the brain samples analyzed. Each cluster is annotated with predicted cell subclasses, providing an understanding of the cell type composition within each dataset.

Multiome

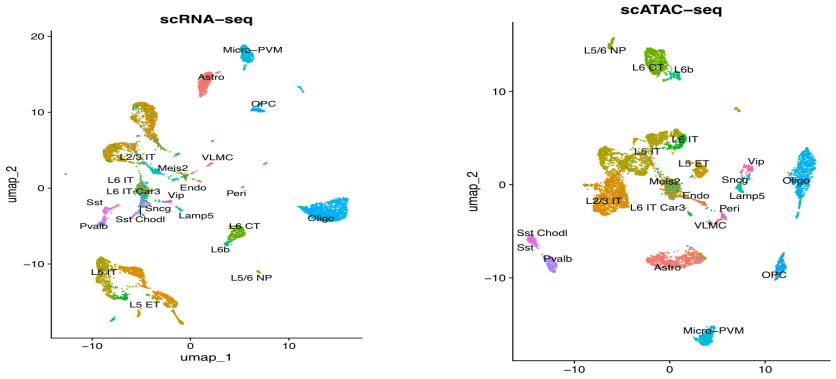
The clusters appear well-defined and segregated, suggesting that the Multiome dataset likely captures a clear and discernible cellular signature. The color-coding corresponds to various cell types, allowing for a visual appreciation of cellular diversity and distribution within the sample.





GEX + ATAC

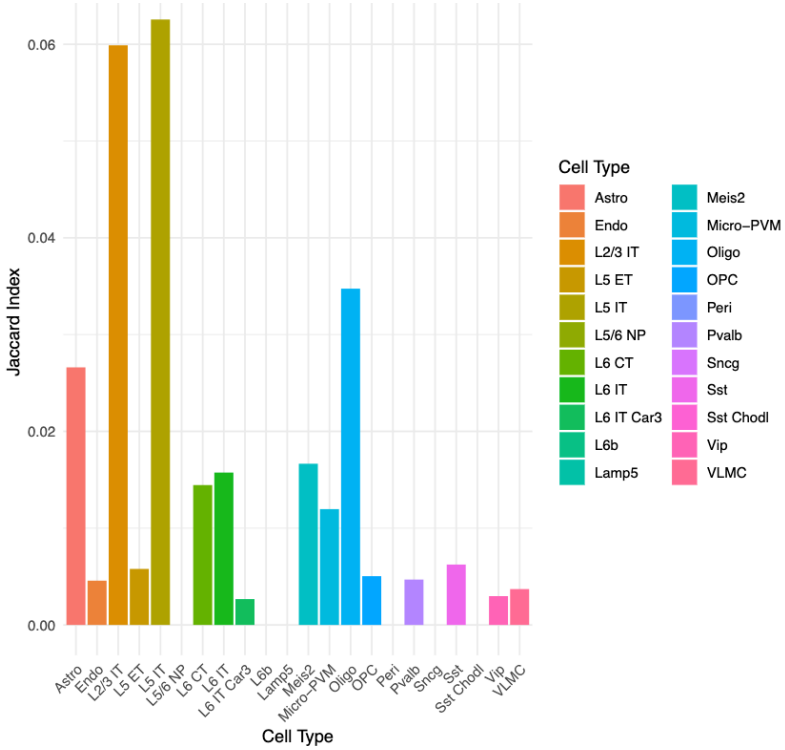
Similarly, this composite UMAP displays a distinct separation of clusters. The clustering patterns offer insights into the cell type variations and the underlying genomic landscape's influence on transcriptional profiles.



Jackard

The Jaccard index plot quantifies the similarity between cell types across the two datasets. Each bar represents a unique cell type, with the height indicating the Jaccard index's value – a measure of overlap between the cell types' distribution in the Multiome and GEX + ATAC datasets.

The Jaccard index values are relatively low, suggesting a small overlap between the datasets. This implies that while there is some similarity in the cell type identification between the Multiome and the GEX + ATAC datasets, there are notable differences. The variability in the Jaccard index across cell types may reflect the degree to which each cell type's signature is captured in the synthetic dataset compared to the Multiome.



- Are the results comparable?

The methods are hardly comparable because ATAC and RNA provide you with a picture of the cell at a given time and given conditions. Taking into account the huge amount of variability inside the cell, when integrating data from cells taken at different times or conditions can be challenging, even if both populations of cells have gone through the same protocol or treatment. In this context, analyzing the state of both RNA and chromatin accessibility from the same cell at the same time provides you with a more accurate picture of the true status of the cell at this given time and conditions. Because of the intrinsic differences of integrated data compared to true multiomics, we consider that both methods cannot be truly compared even when they are comparing the same cell types.

- What are the advantages and disadvantages of each approach?

The authentic multi-omic approach has high accuracy in capturing complex biological interactions due to direct measurement, allowing a detailed understanding of cellular processes. However, it often comes with higher costs and more complex data management and analysis requirements. On the other hand, GEX + ATAC can be more cost-effective and faster by inferring one omic layer from another, but it may sacrifice some accuracy and can be limited by the quality of the prediction models used.

- Why transcriptomics is used as a cornerstone of the different multiome technologies?

Transcriptomics is both simple and informative compared to other omics. Through transcriptomics, the user can rapidly identify the changes in the expression of the cell according to the status or conditions the cell is facing and spot regulatory RNA elements important for gene expression. While genomics analyses of cells are also possible and simple, this assay does not provide clues about the response or behavior of cells in different states. Lastly, proteomics and metabolomics tend to be more complicated. Specifically, proteomics could also provide insights into the response of cells to determined states, but knowing only the protein panorama of a cell could mean that we will not have a glimpse of the important regulation mechanisms that many RNA can have. ATACseq and ChIPseq on the other hand, are excellent complementary techniques for transcriptomics because they provide a more accurate landscape of gene regulation inside the cell.

- Does focusing on single cells within the context of multi-omic data provide a complete picture of biological processes, or are there important aspects that this approach might miss?

Single-cell multiomics allows us to analyze the cellular heterogeneity inside a sample. However, it does not take into account that the cells are located in complex systems in which there is constant communication between the elements that compose the system. An example is the crosstalk

between cells in a tissue that can also impact the cell phenotype. This kind of stimulus can come not only from cells in close vicinity but also from hormones or other molecules that travel long distances.

- How might the future of multi-omic single-cell technology contribute to our understanding of complex biological phenomena?

Multi-omic single cell has the potential to allow us to understand cell dynamics in every layer of complexity (i.e DNA, RNA, protein and gene regulation) and integrate the information to have a global landscape of complex processes carried out inside the cell that at the end contribute to the phenotype we are observing. In this sense, we will be able to unravel the complex heterogeneity of cells within tissues, organs and altered states like a disease. The characterization of health and disease, and the mechanisms underlying these states could be discovered. Finally, this new knowledge can be translated into new insights towards developing new and effective therapies.