Single-cell RNA-seq in a choroid plexus from a multiple sclerosis mouse model

Armindo Machado[1], Ana Mendanha Falcão[2]

[1]Informatics Department, University of Minho

[2]Life and Health Sciences Research Institute (ICVS), University of Minho

Abstract. There is a great amount of cell types, states and intercommunications in human tissues. In order for us to improve our comprehension in current tissue and cellular processes, single cell RNA sequencing (scRNAseq) can enable us to understand which genes are expressed at a certain stage in a cell. Besides revealing cellular heterogeneity, this technique is also capable of detecting new cell types or biomarkers related with diseases. Considering that scRNAseq contributes to a more complete knowledge about the structure of the cells in the tissue, it can provide a full perception of disease pathology. Here this tool is used to delve into the cells in the central tissue, the choroid plexus (CP), in the context of multiple sclerosis (MS). The development of central nervous system (CNS) demyelinating disorder can cause major problems in understanding diseases and improving treatment. MS is defined by an inflammatory, neurodegenerative and demyelinating condition caused by invading immune cells. Although the focus of MS research has generally been neuroinflammation in the brain parenchyma, recent studies have begun to uncover the complex role of other brain structures, particularly the CP, in the pathogenesis of this disease. Traditionally known for its role in the production of cerebral spinal fluid (CSF), CP has emerged as a key player in the immune system and management, after all it can serve as a site for immune system entry into the CNS. However, a better understanding of the role of CP in MS pathology is still unknown and requires further research. For these questions to be answered, single cell RNA sequencing will be used to analyze the choroid plexus from a multiple sclerosis mouse model so that it can be possible to clarify the cause of inflammation and identify changes in cellular composition.

1. Introduction

1.1 Understanding the dynamic functions and responses of the choroid plexus

The choroid plexus, situated in the ventricles, is responsible for various tasks, the primary one being the production of the cerebrospinal fluid, that offers the brain more protection as well as better covering. This fluid also regulates nutrient network between the blood and the brain, providing other supplements that preserve and maintain its development. It orchestrates a lot of significant and important operations in the brain that control its chemical balance, such as the secretion of

peptides and other substances which allows to monitor nerve function and neurological mechanisms. In addition to this, it provides supervision, coordinates the immune response of the CNS in health and disease and on top of that it regulates the exchange of immune cells between the blood and the central nervous system. The choroid plexus acts as an entry point that allows immune cells to enter, influencing both beneficial and harmful reactions in disorders of the central nervous system like multiple sclerosis. This tissue responds dynamically to neurological diseases, mounting protective responses to CNS damage. Although durable, it can experience changes in the structure that affect the function of the central nervous system. Exploring these changes, especially in conditions such as multiple sclerosis, offers hope for treatments aimed at leveraging the neuroprotective functions of the choroid plexus.

1.2 Myelin deterioration in Multiple Sclerosis

MS poses a significant challenge within autoimmune disorders and is defined by the progressive deterioration of myelin (demyelination), the protective layer of nerve fibers. It is formed by oligodendrocytes and it' important to highlight the fact that myelin facilitates a rapid axonal conduction, because it acts like an insulator, and supports its integrity. This fast conduction is critical for some motor and cognitive abilities as well as perception. Regarding myelin function and multiple sclerosis pathology a continuous breakdown of myelin disturbs this transmission and leads to the degeneration of nerve fibers which reinforces the progression of the condition.

1.3 Involvement of the Choroid Plexus in Multiple Sclerosis Pathogenesis

Given this consideration, it's important to dissect the role of the choroid plexus in multiple sclerosis. MS is a gradual chronic condition and severe immunosuppression may be linked with the risk of CP associated diseases. Grasping the complex mechanisms that determine CP induced neuroinflammation is clinically important and offers opportunities for therapies aimed at curing MS. The choroid plexus may act as a gateway for the immune system to enter the brain, potentially impacting the progression of MS. Furthermore, changes in the structure and function of the choroid plexus can upset the equilibrium of neuroimmune reactions, potentially playing a role in multiple sclerosis' development. Recognizing the importance of CP induced immune dysregulation in MS, presents promising affiliation. Interventions targeting CP-specific immunity, hold potential for decreasing neuroinflammatory processes and reduce disease progression in MS patients. Furthermore, ongoing research into CP involvement in other neuroinflammatory conditions may shed light on fundamental pathological mechanisms, aiding the development of effective treatment. Ultimately, clarifying the role of CP in MS pathogenesis represents a critical advancement in combating this disorder.

1.4 Animal models in Multiple Sclerosis Research

Various animal models have been refined as a way to gain a better understanding of demyelinating diseases and the main objective is to target the resemblance of multiple sclerosis' pathology. There are many ways to mimic this condition, one being the induction of chemicals that aim specifically at cells that produce myelin or even break its structure. Another one implicates the induction of an immune reaction directed at myelin antigens which prompts inflammatory demyelination. Other model involves the use of viral infections to induce demyelination. Even though this helps us get a better grasp of the disease's pathology it also has confrontations since it becomes difficult to separate and identify the precise impacts on either oligodendrocytes or axons. In chemical-induced models, for example, while the main focus may be targeting myelin, it's important to consider that there could be additional impacts on axonal health and neuroinflammatory responses, which might complicate the analysis and understanding of results. For instance, Diphtheria toxin (DT) has become a valuable asset for targeting and eliminating oligodendrocytes in animal models of demyelination. The toxin operates by disabling elongation factor 2, a critical protein in the process of protein synthesis, resulting in the death of cells that possess the DT receptor.

1.5 Objective

For this project it was analyzed a new mouse model that allows for the specific ablation of oligodendrocytes and demyelination induction in adult animals by inducing tamoxifen. It enables the exploration of the processes involved in restoring myelin and regenerating axons within the CNS. This model's ability to target specific cells for demyelination offers a promising platform to treat this kind of disorders. Computer-based techniques will be executed and with that data the purpose is to collect important information about the development and progression of multiple sclerosis, demonstrating the role in disease pathology. Along with that, our goal is also to describe and analyze the cellular makeup and the gene profiles within the choroid in response to this condition. This study is dedicated to further research on the structural and functional changes of CP in MS and aims to clarify its contribution,

2. Methods

The technique that will be used for this project will be RNA sequencing, more specifically, single cell RNA sequencing**.** With this method it's possible to identify different arrangements of gene expression and that makes it easier to spot these potential variations in the choroid plexus of multiple sclerosis models which extends our input regarding the biological explanations behind this condition. In order for us to execute, scRNAseq requires several operations, such as preprocessing and visualization and also identifying cellular structure. Each of the next steps is carefully designed to ensure the accuracy and reliability of the analysis and to extract meaningful insights into cellular heterogeneity and disease biology.

2.1 Preprocessing and visualization

2.1.1. Quality Control

This first step is essential to guarantee the accuracy of scRNAseq data. During this analysis stage, low quality cells will be removed from the dataset depending on count depth, number of genes and fraction of mitochondrial counts. By using sequencing depth, the indication of a low count depth implies that there may be poor sequencing or even dead cells. Besides, a low number of identified genes might indicate corrupted RNA, derived from cell degradation. Along with that if a high fraction of mitochondrial counts is detected that also suggests dying cells and cellular stress. However, its crucial to take into account these three control variables at the same time, since, individually, they are not enough to deduce the cell's quality. So, filtering out cells based solely on one parameter is risky because suitable cells might be isolated.

2.1.2. Normalization

This next step is designed to correct the raw counts while taking account various sampling effects whether throughout the capturing procedure, the reverse transcription or even sequencing RNA molecules. Therefore, if we set aside this practice, these sampling effects will be confrontational regarding differences in gene expression between cells.

2.1.3. Feature Selection

For this one the focal point is to target valuable genes and reject the ones that compromise the biological variation across the samples (uninformative genes) which helps to condense the dimensions of the data. Various methods are employed for feature selection, each tailored to capture distinct aspects of gene expression profiles. These methodologies commonly involve filtering genes based on specific criteria such as expression level, dispersion, and detection rate. For instance, genes may be filtered based on their expression level surpassing a minimum threshold and their dispersion measured through metrics like detection rate indicated by the percentage of cells expressing the gene.

2.1.4. Dimensionality Reduction

By picturing scRNA-seq data it's easy to recognize its high volume and scale, so this technique is applied to lower the dimensions while maintaining viable data for analysis. One approach is Principal Component Analysis (PCA) that will generate uncorrelated variables named principal components, obtaining the best variance possible without losing significant information. The t-SNE method offers an idea into the organization of data in higher dimensions, allowing the visualization of complex datasets in two or three dimensions, as it enhances our comprehension of structures and correlations within the data. Uniform Manifold Approximation and Projection or UMAP it's a similar execution, yet presents advantages including a better preservation of the data's overall structure since its used for larger datasets.

2.2 Identifying cellular structure

2.2.1. Clustering

The objective here is to categorize cells into different clusters based on shared patterns within gene expression profiles which helps get a better understanding of cell heterogeneity and identifying cell populations with similar transcriptional profiles. For this to work algorithms are used to organize cells into clusters by measuring how close they are to each other in a lower dimension obtained from previous techniques.

2.2.2. Annotation

The main purpose is to connect the raw dataset with existing biological information. The previous step is relatively easy to execute, however determining prior biological knowledge and assigning it to each cluster it's more demanding. This method often requires manual interpretation of the data, which takes time and can vary based on personal judgment. One way we can achieve this implicates the comparison to existing datasets which have been identified and labeled already. Annotation helps to illustrate the cellular composition of the analyzed tissue.

2.2.3. Data Integration

This final step allows for the comparison of scRNA-seq sequencing datasets obtained from diverse experimental setups or samples. Additionally, with data integration we can identify shared cell types. Regardless, this process involves alterations in observed gene expression levels, or batch effects, that can manipulate data analysis and biological indicators. These effects may derive from changes in sample processing or some biological reasons. By removing such effects, we are capable of creating cooperative datasets for analyses. Integrated analysis provides a comprehensive view of cellular heterogeneity and disease-associated transcriptional changes.

3. Results and Analysis

ScRNAseq was selected to explore the cellular composition and gene expression profiles within the choroid plexus. By using the previous techniques for data analysis, including dimensionality reduction, clustering algorithms, and differential gene expression analysis, we aim to identify distinct cell populations and characterize their molecular signatures in the context of MS. The results presented here provide an examination of cell types identified, their differential gene expression patterns, and potential implications for disease progression.

Initial Data Exploration and Quality Control

To ensure the reliability of our single-cell RNA sequencing data from the choroid plexus of a multiple sclerosis mouse model, we conducted comprehensive quality control analyses. Violin plots were generated to visualize the distribution of key metrics across cells, including total counts, number of genes per count, and percentages of mitochondrial, ribosomal and hemoglobin gene counts (Figure 1). These plots allowed us to assess data quality, identify potential outliers, and detect any systematic biases in gene expression profiles

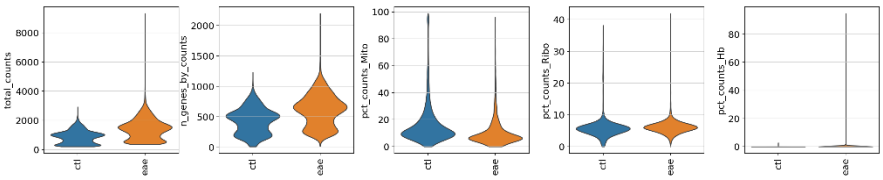

***Figure 1**: Violin plot analysis of key metrics across cells: total counts, genes per count, and mitochondrial, ribosomal, hemoglobin gene percentages.*

Cell Count Distribution Across Samples and Highly Expressed Genes

We proceeded to illustrate the distribution of cells across different experimental samples or conditions. This analysis provided insight into the distribution of sample sizes and ensured that each experimental condition was adequately represented in the dataset. We then plotted a bar chart to identify genes with consistently high expression levels across all cells. This step was crucial for identifying potential marker genes associated with specific cell types or disease states within the choroid plexus and for evaluating overall data quality.

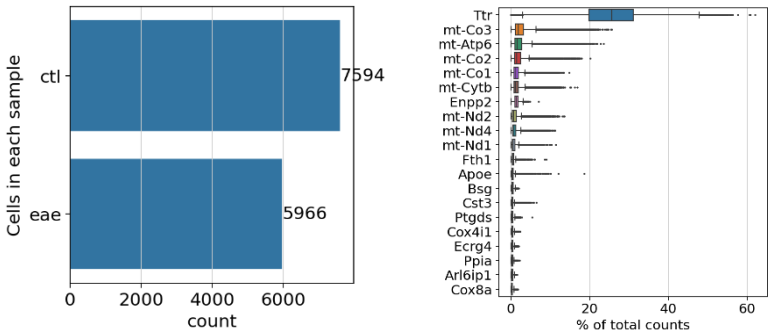

***Figure 2:** Box plot cell distribution of both samples (left). Bar chart of highly expressed genes (right).*

Normalization of Gene Expression Data

To facilitate accurate comparisons of gene expression levels across cells, we normalized the gene expression data so that each cell's total count of expression was regulated to 10,000. This normalization step corrected for differences in sequencing depth among cells, ensuring robust and reliable downstream analyses.

Identification and Comparison of Duplets and Singlets

Using a predefined threshold (0.43), we classified cells as singlets or duplets based on their gene expression profiles. A histogram was generated to visualize the distribution of duplet scores, illustrating the threshold used for distinguishing between singlets and duplets. A violin plot was also created to compare the distribution of detected gene counts between singlets and duplets. This analysis helped assess the effectiveness of duplet removal and ensured that only high-quality cells were included in subsequent analyses.

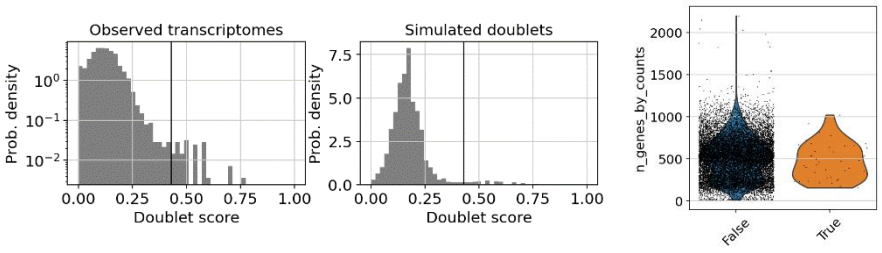

Figure 3: *Histograms showing the doublet scores (left). Violin Plot comparing detected gene counts between singlets and doublets (right)*

Dimensionality Reduction and Visualization

Principal Component Analysis (PCA) was performed to reduce the dimensionality of the gene expression data and extract principal components that captured the variation within the dataset. Subsequently, Uniform Manifold Approximation and Projection (UMAP) was applied for further dimensionality reduction and visualization of the data in two-dimensional space (Figure 4). UMAP enhanced the visualization of complex datasets, aiding in the identification of clusters of cells with similar gene expression profiles.

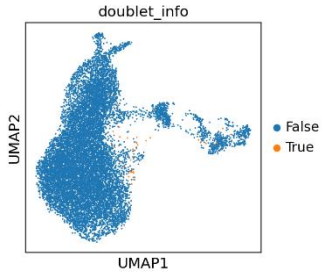

Figure 4*: UMAP showing gene expression data*

Expression of Marker Genes in UMAP

Multiple UMAP plots were generated (Figure 5), each colored according to the expression levels of specific marker genes (e.g., Folr1, Clic6, Ptprz1, C1qa, C1qb, Fh1, Ly6a, Dcn) . These plots visualized the spatial distribution of marker genes within the dataset, facilitating the identification of cell populations and their potential roles in disease pathology.

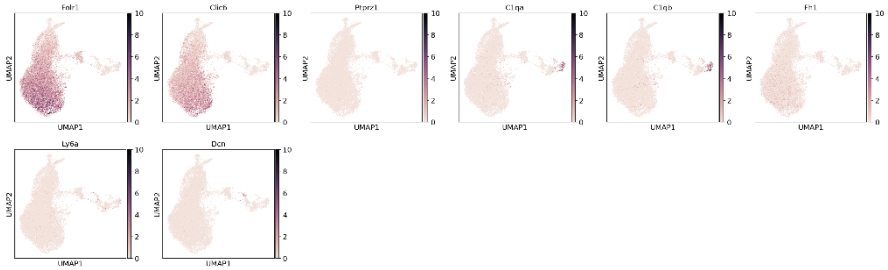

Figure 5: *Series of UMAP charts that illustrate the expression levels of the respective marker genes.*

Quality Filtering and Cell Filtering

Cells identified as duplets were filtered out to retain only singlets for subsequent analyses. Additional filters were applied based on minimum gene expression (300 or more genes per cell) and minimum cell detection per gene to refine the dataset and ensure the inclusion of high-quality cells. There was also filtration based on different quality metrics and specific upper limits for each metric. This is adjusting the criteria for including cells based on their quality, limiting the number of genes and the total number of counts a cell can have. In this case, only cells with up to 5000 genes and 2000 counts were kept.

Post-Filtering Quality Control

Violin plots were used again to visualize quality control metrics (total counts, number of genes per count, percentages of mitochondrial, ribosomal, and hemoglobin gene counts) after applying filtering steps (Figure 6). These plots confirmed the effectiveness of data cleaning procedures and verified that only high-quality cells and genes were retained for following analyses

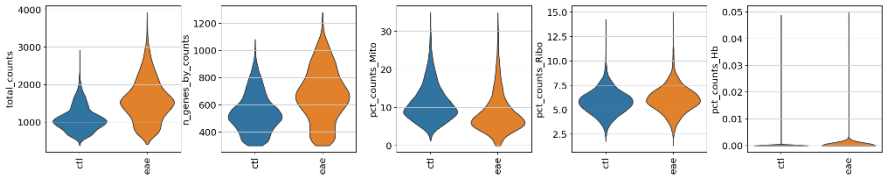

Figure 6*: Violin Plots displaying QC metrics after filtration.*

Identification of Highly Variable Genes

Genes exhibiting high variability in expression across cells were identified, and a graph was plotted to show the dispersion of mean gene expression and expression variability (Figure 7). This analysis highlighted genes potentially involved in cellular heterogeneity and disease processes within the choroid plexus.

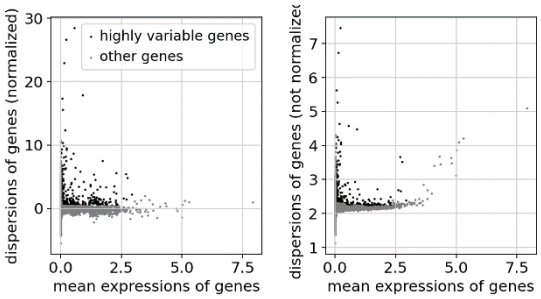

Figure 7: *Charts that identify the genes with highest expression.*

Clustering Analysis

Using the Leiden algorithm, cells were clustered at different resolutions based on their gene expression profiles (Figure 8A). Results identified distinct cell populations within the choroid plexus, with each UMAP plot colored by cluster revealing the spatial distribution of cell clusters and their relationships based on gene expression similarities. The expression of marker genes in the identified clusters was then visualized (Figure 8B).

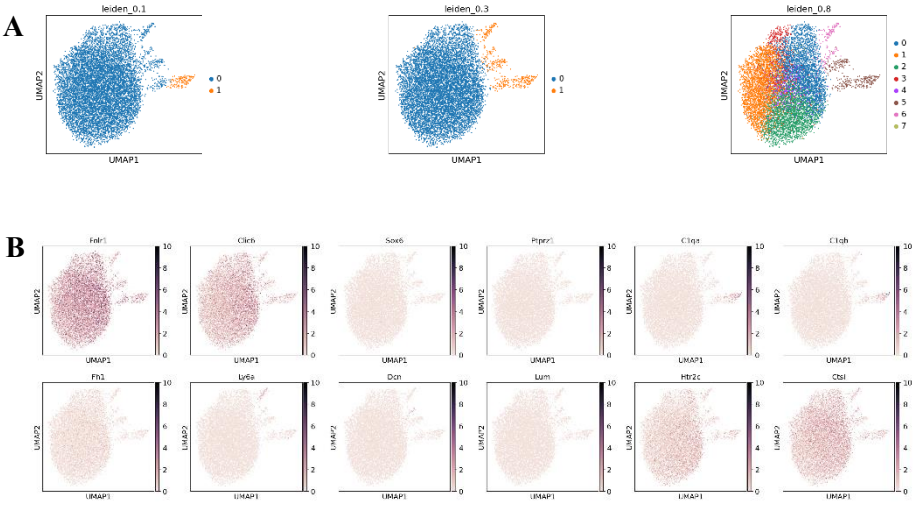

Figure 8: *UMAP visualization of clusters of different cell populations (A) and marker gene expressions (B).*

Differential Expression Analysis

Differential expression analysis between identified cell clusters was conducted using the Wilcoxon test (Figure 9). This analysis identified genes that were significantly differentially expressed across clusters, providing insights into molecular pathways and potential biomarkers associated with multiple sclerosis pathology in the choroid plexus.

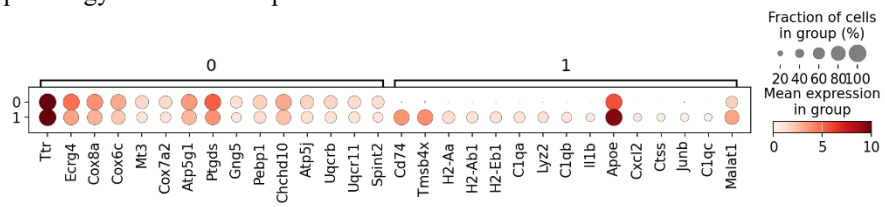

Figure 9: *DotPlot that helps to identify differentially expressed genes.*

Cell Distribution and Gene Expression

This next representation makes it possible to examine the distribution of cells in the UMAP space in relation to the expression of specific genes. These genes can act as markers for different cell types and their expression can help to understand the composition and function of the cell populations identified in the data.

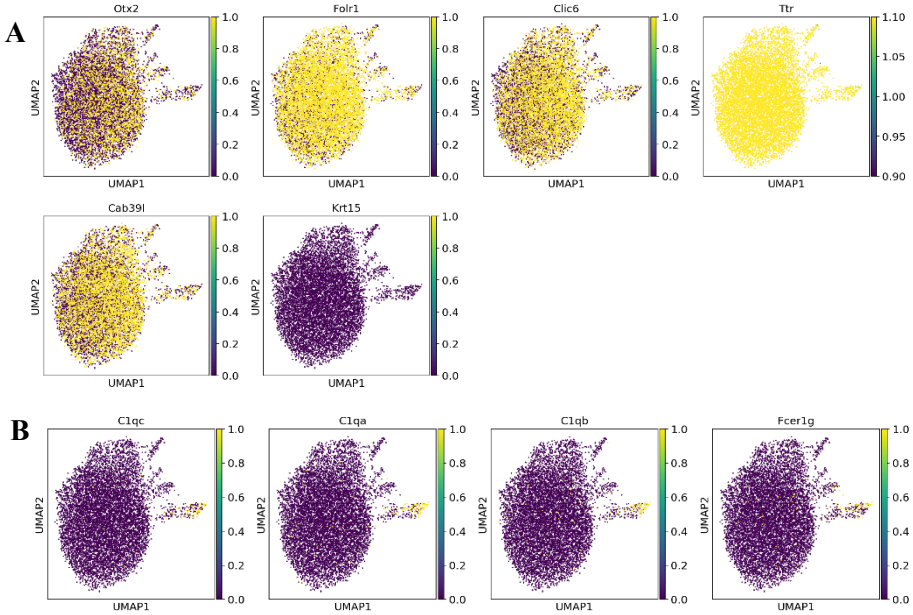

Figure 10: *UMAP representation of cell distribution and gene expression. A – Epithelial Cells; B - Immune Cells*

Visualization of Cell Types and Markers

A Dotplot was generated to visualize the most differentially expressed genes between identified cell clusters, aiding in the characterization of distinct cell types present in the dataset and their potential roles in disease progression. Additionally, a UMAP plot colored by identified cell types provided a visual representation of the distribution of cell types within the dataset.

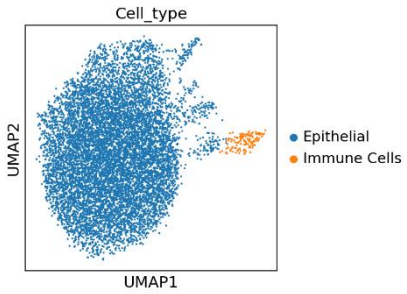

Figure 11: *Cell type classification.*

Cell Type, and Experimental Condition Distribution

Separate bar charts were generated to illustrate the distribution of cells across different cell types, and both experimental conditions. These charts provided a comprehensive overview of sample characteristics and experimental design considerations,

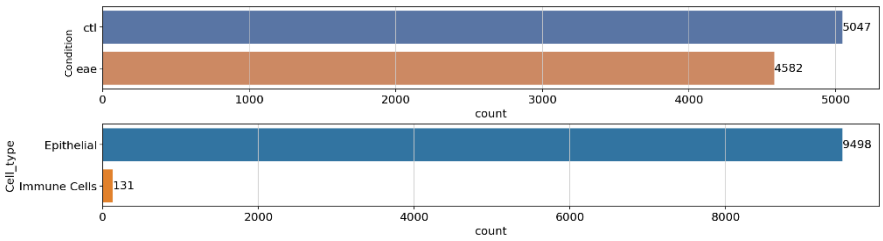

Figure 12: *Visualization of cell distribution among both condtitions and both cell types, respectively.*

4. Discussion

The expected outcome was to identify distinct cell types such as epithelial, mesenchymal, endothelial, immune, and neuronal/glial cells within the CP. However, the analysis revealed a more focused cellular landscape than initially hypothesized.

The scRNAseq data analysis identified two major cell types within the CP: epithelial cells and immune cells. Epithelial cells constituted most of the cell population in the CP tissue. These cells were characterized by the expression of well-known epithelial markers such as Folr1, Clic6 and Ttr. The second significant population identified was immune cells. These cells expressed markers associated with various immune cell types, such as C1qa C1qb, C1qc, indicating a mixture of lymphocytes and macrophages/microglia. The presence of these cells supports the role of the CP as an immunological interface, particularly relevant in the context of MS where immune system dysregulation is a sure sign.

The presence of immune cells in the CP aligns with previous statements implicating the CP as a gateway for immune cell entry into the CNS during neuroinflammatory conditions like MS. This finding points to the CP's significant role in disease pathogenesis by facilitating immune surveillance and potential infiltration of inflammatory cells into the CNS, contributing to demyelination and neuronal damage characteristic of MS.

4.1. Limitations and Considerations

The discrepancy between the expected and observed cell types highlights several potential insights into the CP's role in MS. The absence of mesenchymal, endothelial, and neuronal/glial cell types suggests either a limitation of the scRNAseq methodology used or a specific biological state of the CP in the context of MS. It is essential to consider that scRNAseq provides an unbiased view of gene expression profiles but may not capture all cell types due to technical limitations or

the specific conditions of the experimental model. The technique might have missed certain cell populations due to low abundance or inadequate capture efficiency, the depth of sequencing and the resolution of cell type identification may impact the completeness of our findings. Additionally, the preprocessing steps, such as quality control, might have excluded some rare cell types, affecting the overall cell type distribution.

To address the limitations and validate our findings, future studies should consider using complementary techniques to confirm the presence of the missing cell types. Additionally, examining CP tissue at different stages of MS progression or using different MS models might provide a more comprehensive understanding of its cellular composition.

5. Conclusion

In conclusion, our study highlights the power of scRNAseq in revealing the cellular and molecular foundations of MS within the choroid plexus. By diving into the complex network of gene expression patterns and cell types involved in disease progression, we contribute to a deeper understanding of the CP's cellular architecture and its involvement in MS. The identification of epithelial and immune cell populations highlights the CP's dual role in CNS protection and immune regulation. This research captions the importance of continued research efforts aimed at breaking down the complexities of neuroinflammatory diseases and advancing therapeutic strategies.

References

Dugast, E., Shah, S., & Laplaud, D. A. (2022). Single-Cell Analysis to Better Understand the Mechanisms Involved in MS. *International journal of molecular sciences*, *23*(20), 12142. https://doi.org/10.3390/ijms232012142

Cheng, C., Chen, W., Jin, H., & Chen, X. (2023). A Review of Single-Cell RNA-Seq Annotation, Integration, and Cell-Cell Communication. Cells, 12(15), 1970. https://doi.org/10.3390/cells12151970

Traka, M., Arasi, K., Avila, R. L., Podojil, J. R., Christakos, A., Miller, S. D., Soliven, B., & Popko, B. (2010). A genetic mouse model of adult-onset, pervasive central nervous system demyelination with robust remyelination. Brain : a journal of neurology, 133(10), 3017–3029. https://doi.org/10.1093/brain/awq247

Dani, N., Herbst, R. H., McCabe, C., Green, G. S., Kaiser, K., Head, J. P., Cui, J., Shipley, F. B., Jang, A., Dionne, D., Nguyen, L., Rodman, C., Riesenfeld, S. J., Prochazka, J., Prochazkova, M., Sedlacek, R., Zhang, F., Bryja, V., Rozenblatt-Rosen, O., Habib, N., … Lehtinen, M. K. (2021). A cellular and spatial map of the choroid plexus across brain ventricles and ages. Cell, 184(11), 3056–3074.e21. https://doi.org/10.1016/j.cell.2021.04.003

Haque, A., Engel, J., Teichmann, S.A. et al. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. Genome Med 9, 75 (2017). https://doi.org/10.1186/s13073-017-0467-4

Hwang, B., Lee, J.H. & Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. Exp Mol Med 50, 1–14 (2018). https://doi.org/10.1038/s12276-018-0071-8

Ghersi-Egea, J. F., Strazielle, N., Catala, M., Silva-Vargas, V., Doetsch, F., & Engelhardt, B. (2018). Molecular anatomy and functions of the choroidal blood-cerebrospinal fluid barrier in health and disease. Acta neuropathologica, 135(3), 337–361. https://doi.org/10.1007/s00401-018-1807-1

Muthuraman, M., Oshaghi, M., Fleischer, V., Ciolac, D., Othman, A., Meuth, S. G., Gonzalez-Escamilla, G., & Groppa, S. (2023). Choroid plexus imaging to track neuroinflammation - a translational model for mouse and human studies. Neural regeneration research, 18(3), 521–522. https://doi.org/10.4103/1673-5374.346471

Rodríguez-Lorenzo, S., Ferreira Francisco, D.M., Vos, R. et al. Altered secretory and neuroprotective function of the choroid plexus in progressive multiple sclerosis. acta neuropathol commun 8, 35 (2020). https://doi.org/10.1186/s40478-020-00903-y

Gharagozloo M, Mace JW and Calabresi PA (2022) Animal models to investigate the effects of inflammation on remyelination in multiple sclerosis. Front. Mol. Neurosci. 15:995477. doi: 10.3389/fnmol.2022.995477

Vercellino, M., Votta, B., Condello, C., Piacentino, C., Romagnolo, A., Merola, A., Capello, E., Mancardi, G., Mutani, R., Giordana, M. T., & Cavalla, P. (2008, August 1). Involvement of the choroid plexus in multiple sclerosis autoimmune inflammation: A neuropathological study. Journal of Neuroimmunology. https://doi.org/10.1016/j.jneuroim.2008.04.035

Dixon, G. A., & Pérez, C. A. (2020, February 1). Multiple Sclerosis and the Choroid Plexus: Emerging Concepts of Disease Immunopathophysiology. Pediatric Neurology. https://doi.org/10.1016/j.pediatrneurol.2019.08.007

Klistorner S, Barnett MH, Wang C, Parratt J, Yiannikas C, Klistorner A. Longitudinal enlargement of choroid plexus is associated with chronic lesion expansion and neurodegeneration in RRMS patients. Multiple Sclerosis Journal. 2024;30(4-5):496-504. doi:10.1177/13524585241228423