

Integration of Multi-omics Data and Single-cell Analysis for Alzheimer's Disease

Alzheimer's disease (AD) is the most common type of dementia, affecting nearly 45 million people worldwide [1]. Patients typically have a survival time of 3-10 years post-diagnosis and suffer from severe cognitive impairments [2,3]. Current treatments focus on symptom management rather than addressing the underlying causes[3]. While bulk multi-omics analyses have linked microglial cells to inflammation and synapse pruning, with implications for memory and cognition in AD, single-cell transcriptomic studies have identified AD-associated microglial subsets [4,5,6]. However, the molecular signatures and mechanisms associated with healthy and AD-associated microglial subsets are not yet comprehensively understood. Gaining insights into these signatures and mechanisms is vital for improving disease understanding, monitoring and treatment.

The ultimate goal of this study is to identify novel biomarkers and therapeutic targets focused on microglia in AD using single-cell multi-omics analysis. Specifically, we will consider epigenomics (scATAC-seq to be filtered for microglia from *AD Knowledge Portal*) [7], transcriptomics (scRNA-seq from *Gene Expression Omnibus*, Series GSE219284) [8], and proteomics (spectral data from *PRIDE*, *PRoteomics IDentification*) [9] data due to roles in differential enrichment of certain transcription factors [10], existence of different microglial subsets [11] and microglia-associated changes in energy metabolism respectively [12]. Specifically, we will use Multi-Omics Factor Analysis v2 (MOFA+) for its capability of handling multi-modal and missing data, large-scale data integration and facilitation of downstream analysis. Thus, microglia-specific multi-omics analysis is expected to capture signatures associated with healthy and AD-associated microglial subsets. Our analysis could provide holistic insights into overall changes in microglial subsets in AD pathogenesis, resulting in the identification of new therapies and biomarkers focused on driving microglial fate in AD.

Aim 1: Integration of multi-omics datasets for microglial analysis in AD. Data from microglial specific mono-omics datasets, specifically the epigenomics, transcriptomics and proteomics datasets mentioned above, will be gathered for both AD-affected and healthy controls and necessary preprocessing, quality control and feature selection will be undertaken to make it suitable for data integration. The output will be a comprehensive and standardized repository with columns representing different cells and rows representing accessible chromatin region, gene, and protein, correspondingly, consistent for downstream MOFA+ analysis. This dataset would also be independently useful for other researchers in the field trying to study the role of microglia in AD since this data set would be the first of its kind.

Aim 2: Identification of molecular signatures associated with different microglial subsets in AD. Apply MOFA+ on the integrated complex dataset from aim 1. We will derive N latent factors and feature weight matrices that elucidate the primary axes of variation across these diverse data types. While some factors could highlight known variations in AD-associated genes (such as *TREM2*, *ABCA7*, and *CD33*), DNA methylation and proteomic, and metabolomic profiles, others could potentially uncover novel, cross-layer changes that provide a more integrated understanding of molecular signatures associated with the emergence of AD-associated microglial subsets [6,13] (evidence: [14]). Success of this aim will be determined by the ability to consistently detect these molecular signatures for different subsets across multiple datasets.

Aim 3: Discovery of novel molecular biomarkers associated with microglia in AD. Supervised machine learning algorithms such as extreme gradient boosting (XGBoost) will be used to train a classifier by using features with higher weights from each latent factor due to its interpretability, ability to capture non-linear relationships and robustness in handling missing data and heterogenous data types [15]. The most important features are identified as candidate biomarkers since they encapsulate significant variations observed in the microglial subsets. Success of this aim will be determined by the consistent identification and preliminary validation of these candidate biomarkers in independent datasets and patient-derived clinical samples, laying the foundation for their use in detection and therapeutic interventions in AD.

This study aims to investigate microglia-specific molecular signatures in AD and identify novel biomarkers by employing existing techniques such as single-cell multi-omics analysis (MOFA+) on novel datasets and using supervised machine learning algorithms (XGBoost). Ultimately, this research holds the potential to revolutionize our understanding of AD and could pave the way for the development of targeted therapies and biomarkers for early detection and therapeutic interventions.

References

- [1] Li, X., Feng, X., Sun, X., Hou, N., Han, F., & Liu, Y. (2022). Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2019. *Frontiers in Aging Neuroscience*, 14, 937486.
- [2] Zanetti, O., Solerte, S. B., & Cantoni, F. (2009). Life expectancy in Alzheimer's disease (AD). *Archives of gerontology and geriatrics*, 49, 237-243.
- [3] Cleveland Clinic. (2024, June 3). *Alzheimer's Disease*. Cleveland Clinic. <https://my.clevelandclinic.org/health/diseases/9164-alzheimers-disease>
- [4] Schwabe, T., Srinivasan, K., & Rhinn, H. (2020). Shifting paradigms: The central role of microglia in Alzheimer's disease. *Neurobiology of disease*, 143, 104962.
- [5] Kodam, P., Sai Swaroop, R., Pradhan, S. S., Sivaramakrishnan, V., & Vadrevu, R. (2023). Integrated multi-omics analysis of Alzheimer's disease shows molecular signatures associated with disease progression and potential therapeutic targets. *Scientific Reports*, 13(1), 3695.
- [6] Gao, C., Shen, X., Tan, Y., & Chen, S. (2022). Pathogenesis, therapeutic strategies and biomarker development based on “omics” analysis related to microglia in Alzheimer's disease. *Journal of Neuroinflammation*, 19(1), 215.
- [7] Sage Bionetworks, info@sagebase.org (no date) *SnATAC-Seq*, *Synapse*. Available at: <https://www.synapse.org/Synapse:syn22130833> (Accessed: 04 June 2024).
- [8] *Geo accession viewer* (no date) *National Center for Biotechnology Information*. Available at: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE219284> (Accessed: 04 June 2024).
- [9] EMBL-EBI (no date) *Pride - Proteomics Identification Database*. Available at: <https://www.ebi.ac.uk/pride/archive/projects/PXD045001> (Accessed: 04 June 2024).
- [10] Scholz, R., Brösamle, D., Yuan, X., Beyer, M., & Neher, J. J. (2024). Epigenetic control of microglial immune responses. *Immunological Reviews*.
- [11] Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J. Z., ... & Tsai, L. H. (2019). Single-cell transcriptomic analysis of Alzheimer's disease. *Nature*, 570(7761), 332-337.
- [12] Johnson, E. C., Dammer, E. B., Duong, D. M., Ping, L., Zhou, M., Yin, L., ... & Seyfried, N. T. (2020). Large-scale proteomic analysis of Alzheimer's disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation. *Nature medicine*, 26(5), 769-780.
- [13] Yang, H.-S. *et al.* Cell-type-specific Alzheimer's disease polygenic risk scores are associated with distinct disease processes in Alzheimer's disease. *Nat. Commun.* **14**, 1–13 (2023).
- [14] Argelaguet, R., Arnol, D., Bredikhin, D., Deloro, Y., Velten, B., Marioni, J. C., & Stegle, O. (2020). MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome biology*, 21, 1-17.
- [15] Ng, S., Masarone, S., Watson, D., & Barnes, M. R. (2023). The benefits and pitfalls of machine learning for biomarker discovery. *Cell and Tissue Research*, 394(1), 17-31.