BIOMEDIN 217
TRANSLATIONAL BIOINFORMATICS

INTEGRATION OF MULTI-OMICS DATA AND SINGLE-CELL ANALYSIS FOR ALZHEIMER'S DISEASE

Daniella Morales, Gowri Subedar,
James Guo and Nolawit Elias

INTRODUCTION

7 million

people in the United States have Alzheimer's Disease

Every 65 seconds, someone in the US will develop AD

PROBLEM STATEMENT

- Patients diagnosed with AD generally have a survival time of 8-10 years, endure severe cognitive impairments, and require extensive care
- Current treatments focus primarily on symptom management rather than addressing the underlying causes, resulting in limited efficacy
- While bulk multi-omics analyses have implicated microglial cells in inflammation and synapse pruning with potential impacts on memory and cognition in AD, single-cell transcriptomic studies have identified specific AD-associated microglial subsets

However, the molecular signatures and mechanisms distinguishing healthy from AD-associated microglial subsets remain incompletely understood. Unraveling these signatures and mechanisms is crucial for advancing our understanding, monitoring, and treatment of the disease.

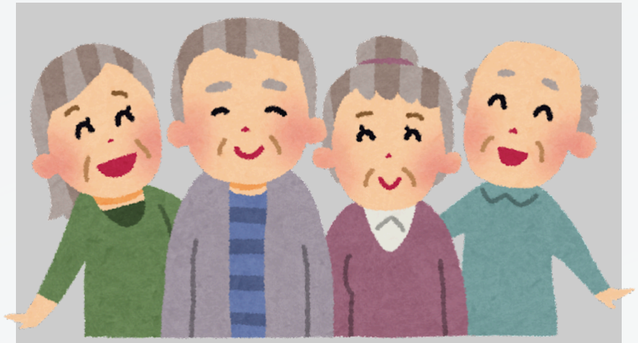
PROBLEM STATEMENT: LONG TERM

- The long-term objective of this study is to identify novel biomarkers and therapeutic targets focusing on microglia in AD through single-cell multi-omics analysis.
- MOFA+ has previously demonstrated success in revealing molecular signatures across different developmental stages in mammalian embryogenesis , indicating its potential for uncovering microglial-specific signatures in AD.
- By employing this approach, we aim to capture holistic insights into the overall changes in microglial subsets during AD pathogenesis, which could lead to the identification of new therapeutic strategies and biomarkers aimed at steering microglial fate towards a healthy subset.

RATIONALE

- The significance of this study lies in its potential to revolutionize our understanding and treatment of Alzheimer's Disease (AD) by focusing on microglial cells, which are crucial for brain health and implicated in AD pathogenesis
- Traditional treatments have largely failed to address the root causes of AD, focusing instead on symptomatic relief with limited success
- By integrating multi-omics data (epigenomics, transcriptomics, proteomics) with single-cell analysis, this research aims to uncover detailed molecular signatures of microglial subsets, distinguishing healthy from AD-associated cells, potentially leading to the identification of novel biomarkers for early diagnosis and new therapeutic targets for more effective and precise treatments.
- Such advancements hold the promise of not only improving patient outcomes but also significantly advancing the broader field of neurodegenerative disease research by providing a richer, more nuanced understanding of microglial roles in brain health and disease

RATIONALE

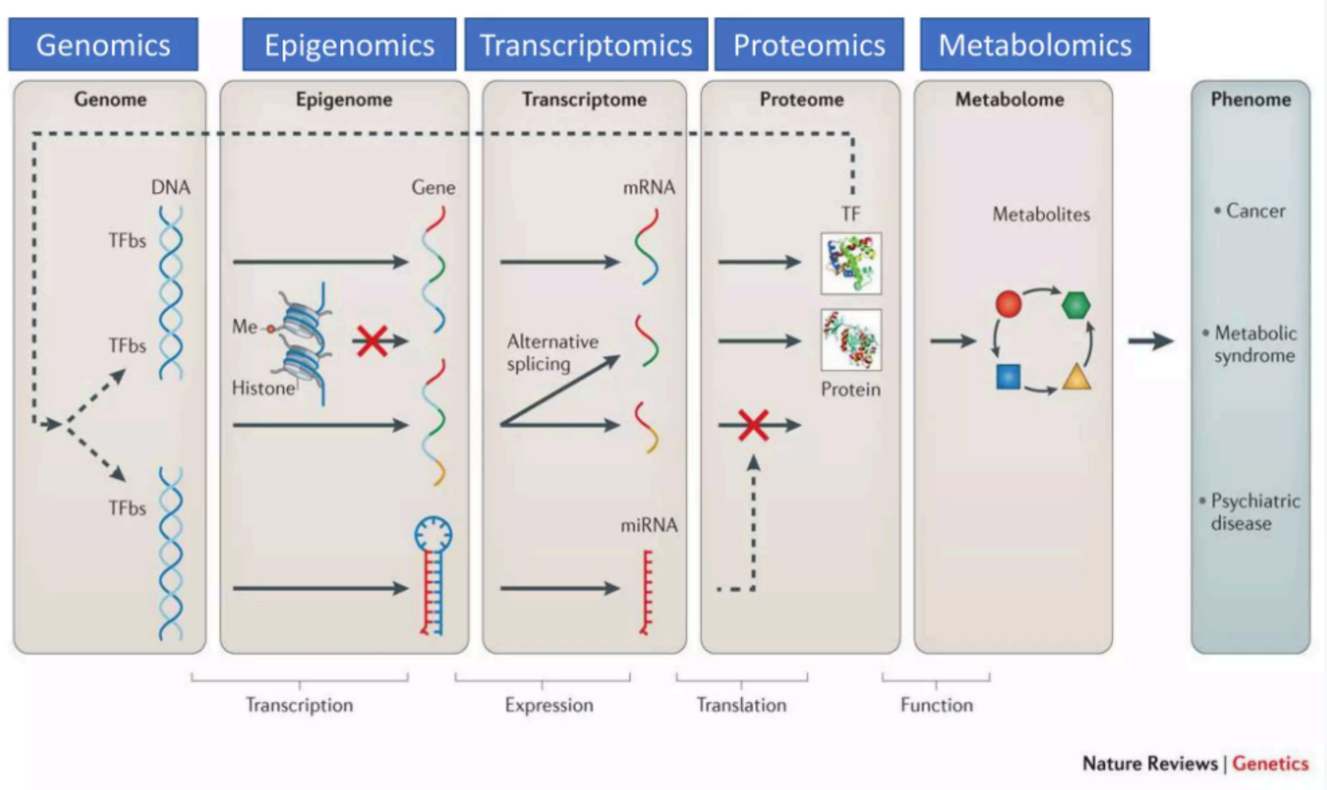


- Early Diagnosis and Intervention: The identification of novel biomarkers for AD, enabled by this research, could revolutionize early diagnosis. Clinicians may be able to detect AD-associated changes in microglial subsets before clinical symptoms manifest, allowing for early intervention and personalized treatment strategies.

- Broader Impact: The integration of multi-omics data and single-cell analysis in microglial-related Alzheimer’s Disease research could catalyze profound changes worldwide. It will lead to earlier and more accurate diagnosis, personalized treatments tailored to individual patients, and improved outcomes. This advancement could accelerate progress in understanding and managing other neurodegenerative diseases

OVERALL RESEARCH STRATEGY

Ultimate goal: identify novel biomarkers and therapeutic targets in microglial cells for AD by integrating epigenomics, transcriptomics, and proteomics.



OVERALL RESEARCH STRATEGY

Specific Aim 1

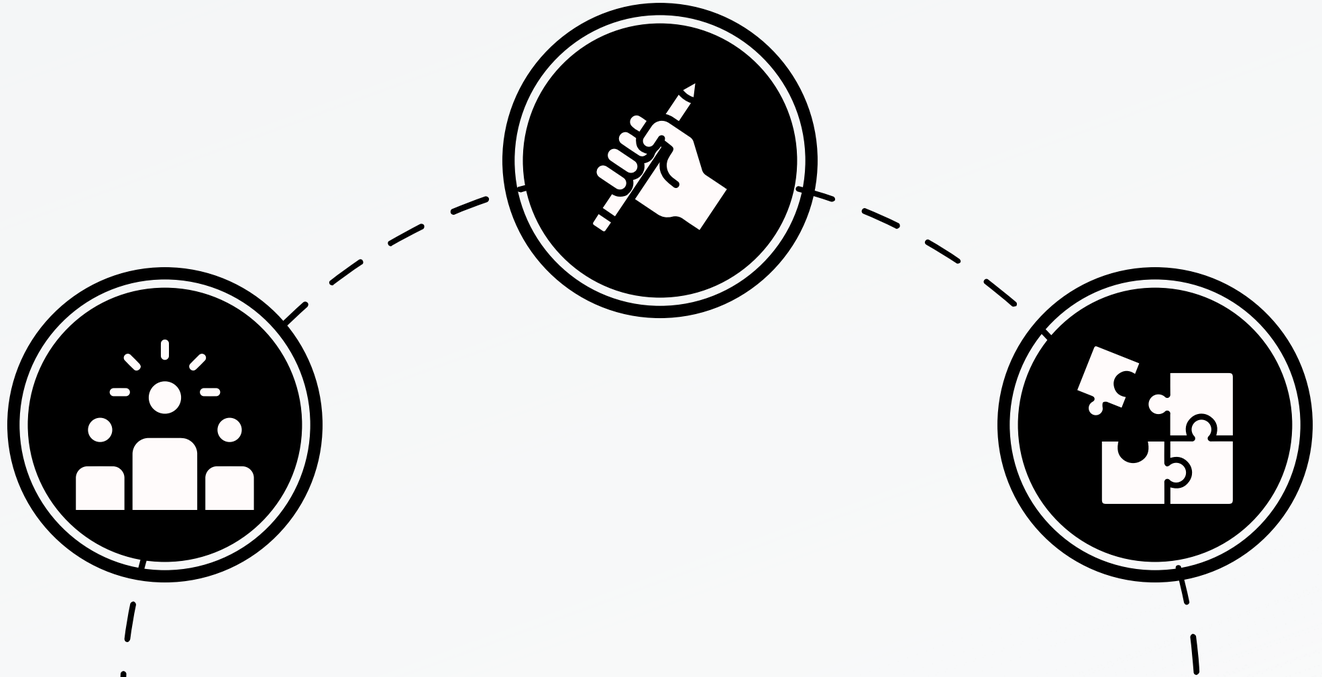
Integration of multi-omics datasets for microglial analysis in Alzheimer's Disease.

Specific Aim 2

Identification of molecular signatures and pathways associated with microglial dysfunction.

Specific Aim 3

Discovery of molecular biomarkers for microglial activation in Alzheimer's Disease.



SPECIFIC AIM 1

Integration of epigenenomic (scATAC-seq), transcriptomic (scRNA-seq), proteomic datasets (sc-MS) for microglial cell analysis in Alzheimer's Disease. scATAC-seq and scRNA-seq datasets are count matrix, and we need to perform quality control and feature selection before using MOFA+.

12330 entries and 641729 columns (will need to filter out microglial cells from meta data)

data source: AD Knowledge Portal

scATAC-seq

68,836 entries and 30 columns

data source: Gene Expression Omnibus (GSE219284)

scRNA-seq

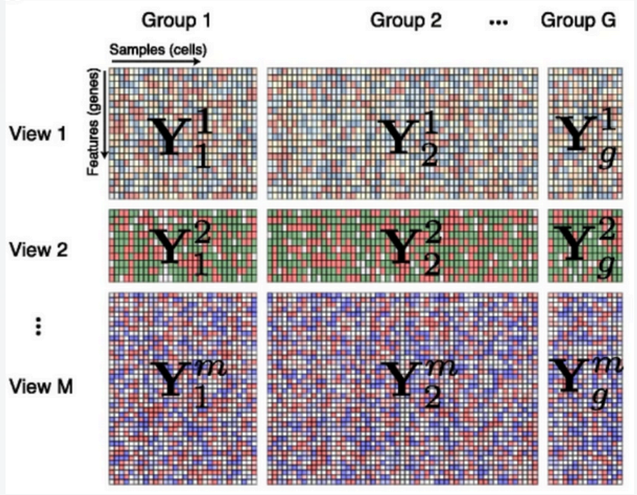
a matrix that stores spectral data that measure mass-to-charge ratio to measure the molecular weight of proteins

data source: PRIDE Archive (PRoteomcis IDEntification)

sc-MS

SPECIFIC AIM 1

MOFA+



- A statistical framework for comprehensive integration of multi-modal single-cell data
- The inputs are multiple datasets where features have been aggregated into non-overlapping sets of modalities (also called views) and where cells have been aggregated into non-overlapping sets of groups, and groups have been aggregated to different experiments, batches, or conditions.

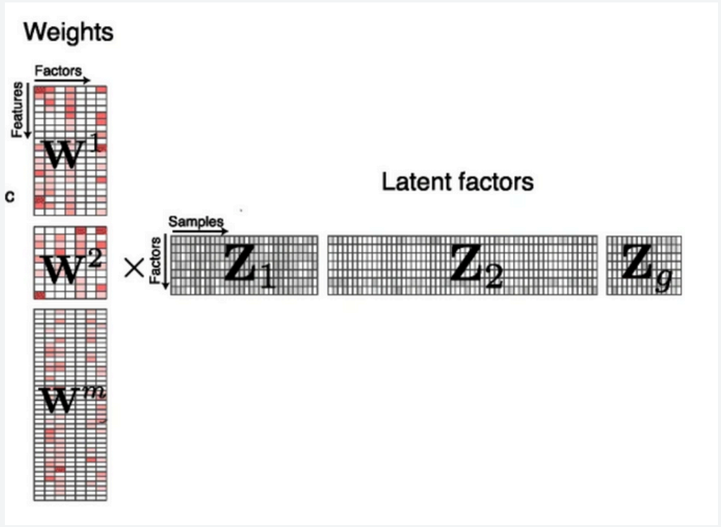
- Excellent in handling missing data, inherent assay noise, and the scale of modern single-cell datasets, which can potentially span millions of cells
- Dismiss the dependencies across cells and in particular those accounting for side information about the structure between cells, e.g., sample groups, such as batch, donors, or experimental conditions.

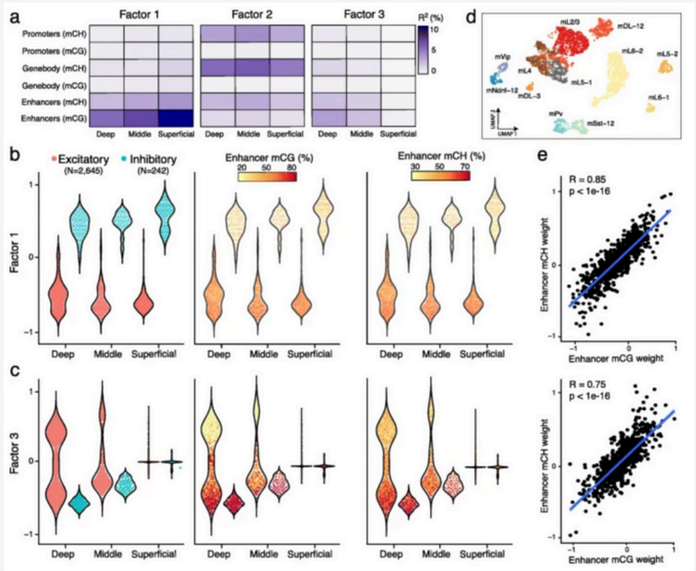
Advantages

SPECIFIC AIM 1

OUTPUT

- a comprehensive, standardized repository of latent factors, weights and variance, and also visualization with different modalities and batches across microglial omics layers consistent for MOFA+ analysis
- not only be useful for our study but also for downstream analysis





SPECIFIC AIM 2

- **Identification of molecular signatures associated with different microglial subsets in AD**
- Apply MOFA+ on the integrated dataset from aim 1
- Will derive N latent factors and feature weight matrices that elucidate the primary axes of variation across data types

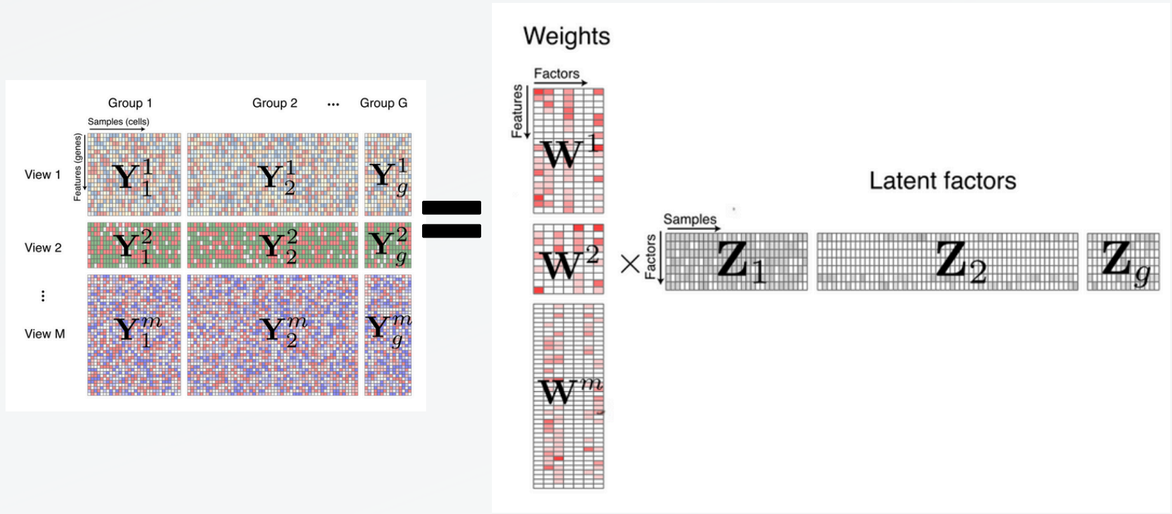
- While results may highlight known variations in AD-associated genes, DNA methylation, and proteomic and metabolomic profiles, others could potentially uncover novel, cross-layer changes that provide a more integrated understanding
- Success determined by ability to consistenly detect these signatures for different subsets of data

SPECIFIC AIM 2

- Method used will follow MOFA+ paper (Argelaguet, et al), which applied MOFA+ to single-cell datasets of different scales and designs with results validated

SPECIFIC AIM 3 — HOW

Discovery of novel molecular biomarkers associated with microglia in AD.

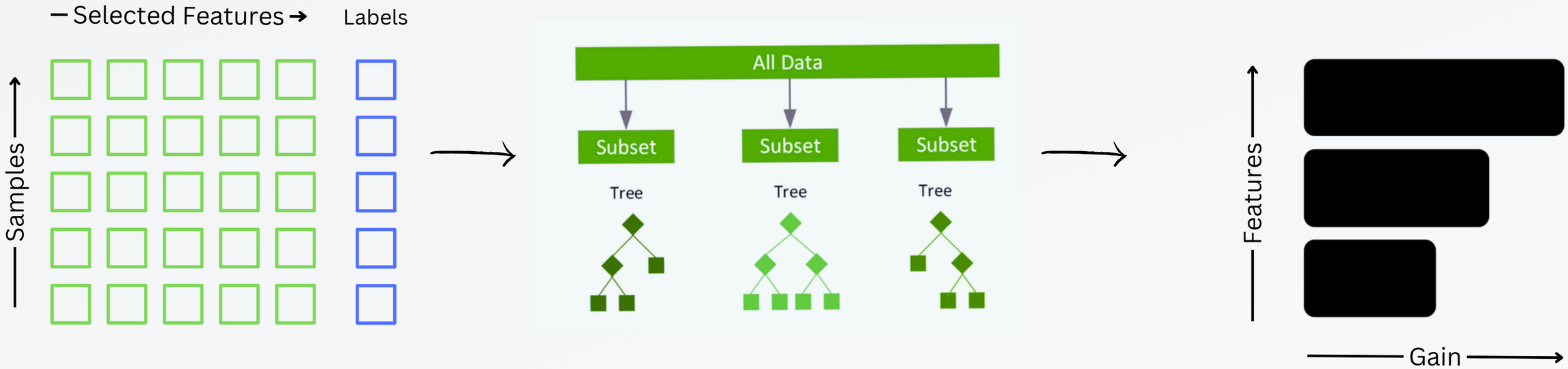


Weight Matrix & Factors

Feature matrix & Standardization

SPECIFIC AIM 3 — HOW

Discovery of novel molecular biomarkers associated with microglia in AD.



- Non-linear relationships
- Missing data
- Heterogenous data types

- Prevents overfitting
- k-fold cross-validation
- Hyperparameter tuning

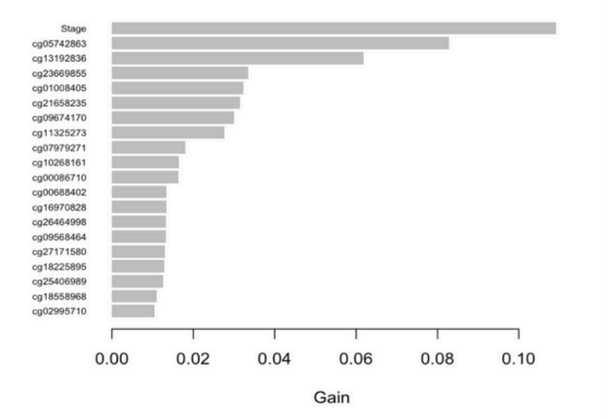
- Feature importance
- Candidate Biomarkers

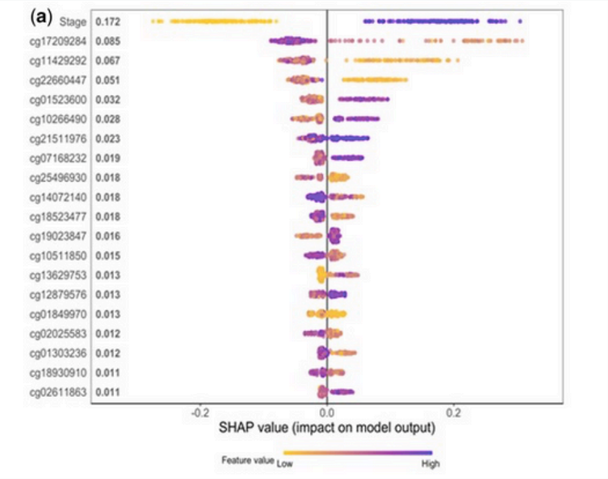
chart by Syaeful Amri from Noun Project (CC BY 3.0)

SPECIFIC AIM 3 — EVIDENCE

Prognostic biomarker discovery using a melanoma methylation dataset

- Matched methylation data and gene expression.
- cg13629753 (in gene GBP2) and cg17209284 (in gene SP140L) most selected.
- SP140L might play a key regulator role in silencing genes that establish immune cell identity and function
- Another notable biomarker - cg11429292 in LAG3, a well-known immune checkpoint regulator.
- Top prognostic CpGs identified in genes more pronounced in specific immune cells



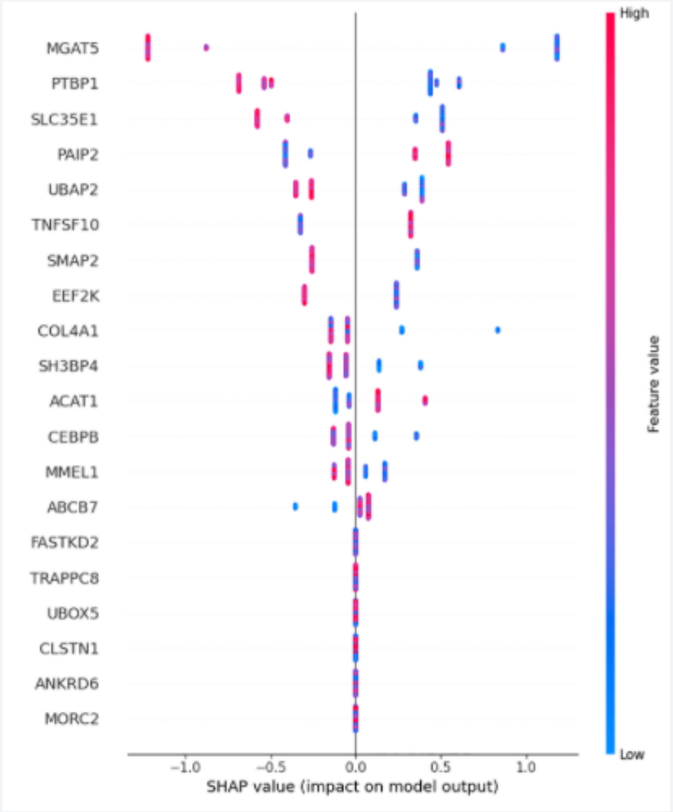


Li, K., Yao, S., Zhang, Z., Cao, B., Wilson, C. M., Kalos, D., ... & Wang, X. (2022). Efficient gradient boosting for prognostic biomarker discovery. Bioinformatics, 38(6), 1631-1638.

SPECIFIC AIM 3 — EVIDENCE

Biomarker discovery in rheumatoid arthritis using transcriptome expression

- MGAT5 - fine-tune inflammation
- TNFSF10 showed differential expression in RA
- ACAT1 identified in genetic association studies of RA
- CEBPB - a role in chronic inflammation of the synovium in RA
- Membrane metallo-endopeptidase-like (MMEL1 1) - implicated in a range on autoimmune disease, including RA



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.

SPECIFIC AIM 3 — SUCCESS

Evaluation and Validation

- Performance Metrics: accuracy, precision, recall, and area under the receiver operating characteristic curve (AUC-ROC).
- Independent Dataset Validation: Validate candidate biomarkers to assess robustness and generalizability.
- Clinical Sample Validation: Conduct preliminary validation to evaluate their potential clinical relevance.

Metrics of Success

- Identify atleast 6 candidate biomarkers.
- Validate 4 in independent datasets
- Validate atleast 2 in clinical samples
- Biological interpretation and functional analysis

FINALLY ...

INNOVATION

Use single cell multi-omics analysis, (MOFA+) to comprehensively understand the role of microglia in AD.

EXPECTED OUTCOMES

- Identify molecular signatures associated with healthy and AD microglial subsets across molecular subsets.
- Discover novel microglia-associated candidate biomarkers.

IMPACT

- Development of targeted therapies and biomarkers.
- Early detection and therapeutic interventions.