Clustering-Based Stratification of Mild Cognitive Impairment: Insights from Blood Transcriptomic Data

Laura Antonelli, Claudia Di Napoli, Lucia Maddalena, Giovanni Paragliola, Patrizia Ribino, Luca Serino, Ilaria Granata

BBCC 2024
Bioinformatics and Computational Biology Conference
27-29 November 2024

Introduction

Alzheimer's disease (AD), a leading cause of cognitive decline and mortality, progresses through stages, including mild cognitive impairment (MCI), further categorized into early MCI (EMCI) and late MCI (LMCI), both of which are key risk factors for dementia. Current diagnostic methods heavily rely on clinical assessments and invasive procedures (i.e., lumbar puncture), highlighting **the need for less invasive biomarkers to improve early detection**.

This study leverages **transcriptomic data from blood samples** of **381 patients** provided by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (https://adni.loni.usc.edu/) **to stratify MCI patients into distinct molecular subgroups**.

A suitable **AI-based pipeline** is built to process **~20.000** gene expression values performing:

- ❖ *SUP and UNSUP combined gene selection*
- ❑ *Deep learning encoding to extract higher-level features from selected genes*
- ➢ *K-means clustering to group patients*
- ○ *Data analysis of clusters*

The findings provide significant differences in molecular, clinical, and pathophysiological data under investigation.

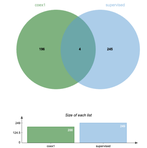
References

[1] Emon, Mohammad Asif, *et al.* SCIENTIFIC REPORTS, ***2020***
[2] Castellanos, Francisco, *et al.* NUCLEIC ACIDS RESEARCH, ***2024***
[3] Piñero, Janet, *et al.* NAR, ***2020***
[4] Milacic, Marija, *et al.* NAR, ***2024***
[5] Kanehisa, Minoru, *et al.* NAR, ***2024***
[6] Sügis, Elena, *et al.* SCIENTIFIC DATA, ***2019***
[7] Caponio *et al.* AGING BRAIN, ***2022***

Methods

❖ FEATURE SELECTION

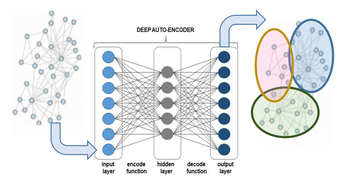
- SUPervised: **249 genes** involved in several neurodegenerative diseases including AD collected from several sources and investigations ([1]; HPO [2]; DisGeNET [3]; Reactome [4]; KEGG [5]; identification of hubs in HENA AD network [6])
- UNSUPervised: 1% data using the z-normalization step followed by co-expression filtering (CoEx1), so providing **200 genes**
- Combined SUP and UNSUP genes resulting in *445 genes*



445 genes

❑ DEEP LEARNING ENCODING

Autoencoder is a neural network that automatically learns data features from a large number of samples and acts as a dimensionality reduction method providing an encoded representation of the selected genes



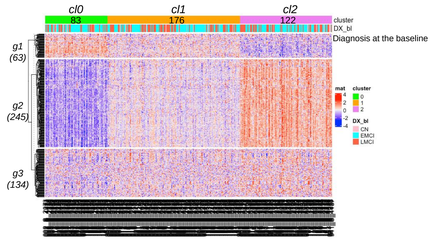
Autoencor + clustering

➢ K-MEANS CLUSTERING

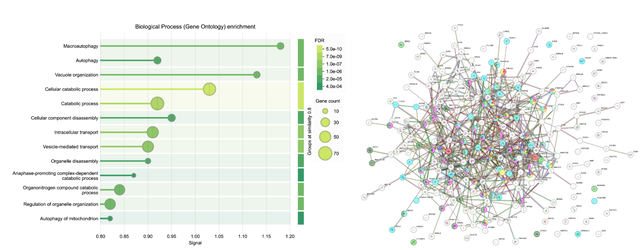
Clustering is performed by grouping patients from two to six classes, where three-classes clustering exhibited the best performance values (*silhouette* = 0.53, *Davies Bouldin* = 0.57).

○ DATA ANALYSIS

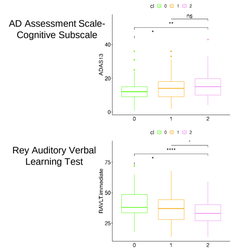
- The heatmap points out meaningful patterns in the complex gene expression data of the three MCI patient groups
- Specific groups of genes are over- or under-expressed in complementary ways for patient clusters 0 and 2 (i.e., *cl0* and *cl2*)
- In particular, two clusters of genes (i.e., *g1* and *g2*) have been identified showing an opposite trend
- The enrichment analysis of the gene cluster *g2* (245 genes) has highlighted the involvement of these genes in autophagy and mitophagy which are reported as related to the memory loss and clinical features of AD [7]
- The three MCI clusters exhibit also significant differences in some cognitive features



Genes Heatmap of clustered MCI

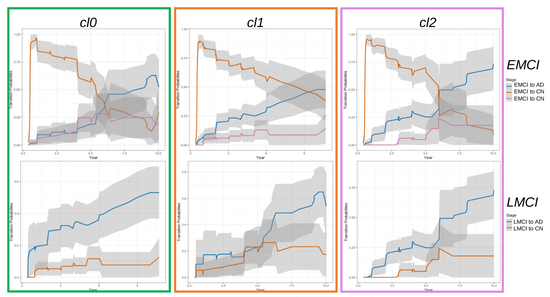


Enrichment Analysis of the gene cluster 2



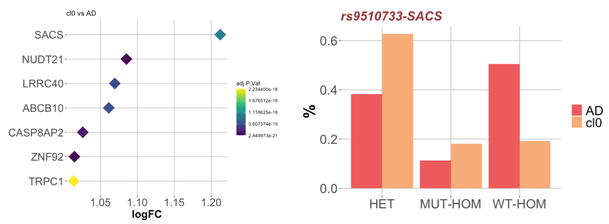
Cognitive test of clustered MCI

The transition from Early or Late MCI to worse or better states of the three clusters of MCI patients reveals that while *cl2* identifies patients with a greater disposition towards worsening, *cl0* includes patients showing a much more unstable status with a high probability of worsening but also a higher probability of getting reversion.



Transition stages of MCI patients

The cluster *cl0* showed differential gene expression compared both to AD and Control subjects (*CN*).
In the first case, it was worth noticing that the SACS gene was affected by an SNP found significantly associated with *cl0* compared to AD patients (Maximum test).



Analysis of Gene expression and SNP association

Results

- ✓ Based on blood transcriptomic data, **MCI patients were stratified** into distinct clusters, showing significant molecular, clinical, and pathophysiological differences between subgroups
- ✓ MCI clusters showed apparent differences compared to AD patients and Control samples (*CN*), **indicating the heterogeneity of MCI and its progression risk**
- ✓ Blood transcriptomic profiles could be used as **less invasive biomarkers**, offering an alternative to cerebrospinal fluid and imaging-based biomarkers.
- ✓ Other results… on future works!

Conclusions

The results promote the investigation of the **blood transcriptomic data** for the **molecular stratification of MCI patients**.
These findings support standard cognitive-based tests that can often be influenced by contextual factors that are difficult to isolate.

Acknowledgements

PNRR PE8 Mission 4,C2, Intervention 1.3 - **Age.it**

PRIN 2022 PNRR - P2022PMEN2 – **MATH4CHARM**