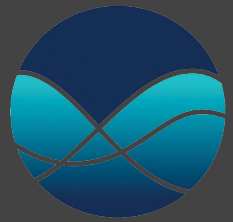
Robustness of Emergent Clusters from Multi-Protein Serum Data

Presented at Consortium of Multiple Sclerosis Centers (CMSC) Annual Meeting

2023

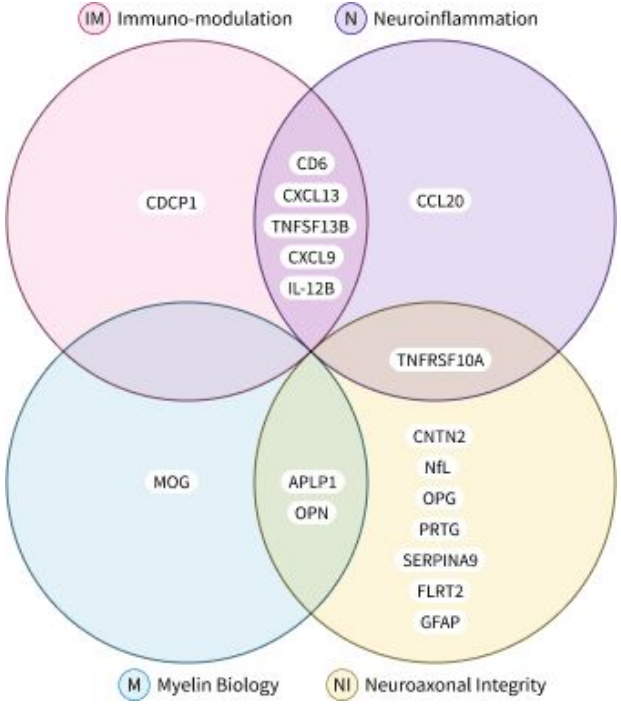
Ati Ghoreyshi, Elisa Sheng, Ferhan Qureshi, Tammy Hoyt, John Foley



ROCKY MOUNTAIN
Multiple Sclerosis Clinic

Octave MSDA Test Background

- Advances in serum proteomics have provided more precise tools for the characterization of Multiple Sclerosis (MS) and enabled its enhanced clinical management.
- The Octave MSDA (Multiple Sclerosis Disease Activity) Test, a mutivariate proteomic blood test, measures 18 serum biomarkers. It reports 4 Disease Pathway scores and an overall Disease Activity score.
- It is strongly associated with the following endpoints: Gd+ lesions, N/E T2 lesions and active/stable status.
- **The Octave MSDA Test has been analytically and clinically validated.**
- **Multi-protein signatures and patterns derived from the 18 proteins may be associated with key clinical variables of interest and additional MS endpoints.**



Phenoclustering

Introduction & Objectives

- Due to the complexity of treatment and heterogeneity in MS pathophysiology, identifying emergent distinct biomarker profiles using machine learning techniques can enable proteomic-based MS subtyping and deepen the clinical interpretability of Octave MSDA results.
- We have shown preliminary results using this technique (ECTRIMS 2022)
- **To examine the clinical utility and robustness of emergent groups further, we have examined:**

Alignment of clusters with clinical expectations

Robustness of clusters

Indication of subclinical signal

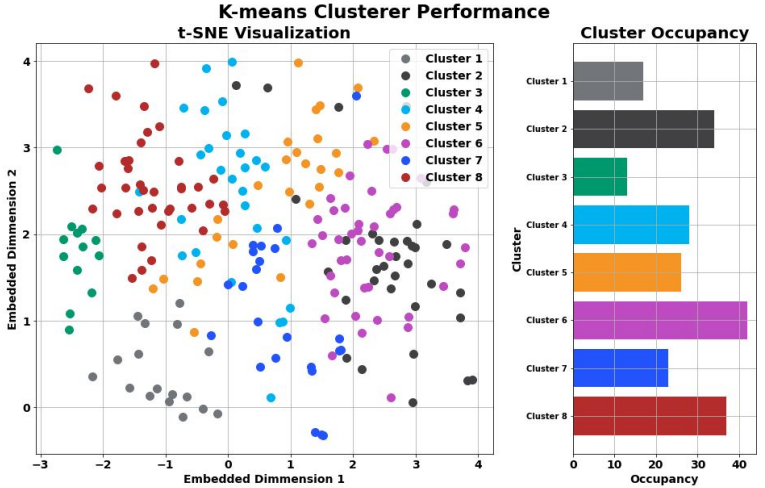
Alignment of Clusters with Clinical Expectations

Phenoclustering

Methods

- 220 patient samples from a study at Rocky Mountain Multiple Sclerosis Clinic were assayed in the MSDA test (Training data).
- The unadjusted proteomics data (18 proteins) was grouped into emergent clusters identified using unsupervised learning (K-means).
- We chose 8 clusters to be able to see clinically relevant distinctions between clusters, while keeping cluster occupancy reasonably high.
- We examined the distribution of DA scores and other demographic and clinical variables for each of the groups.

Clustering on Training Data Results

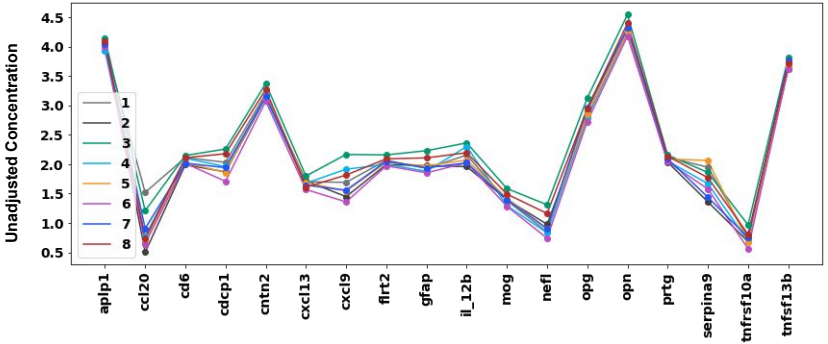


Clustering on Training Data

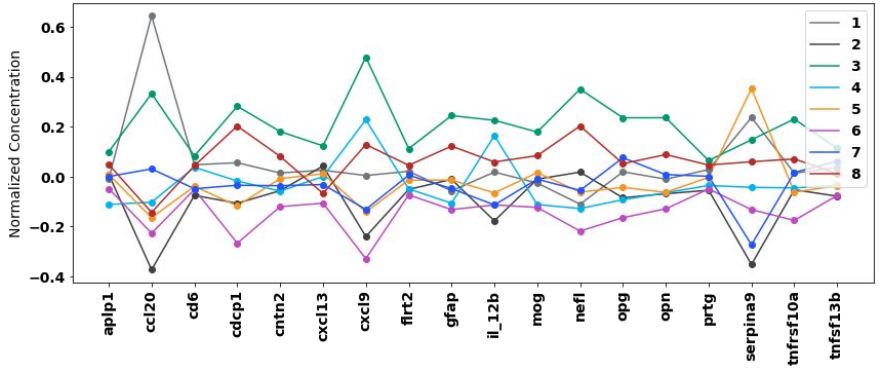
Protein Signatures of Clusters

Clusters have distinct protein signatures

Cluster centers



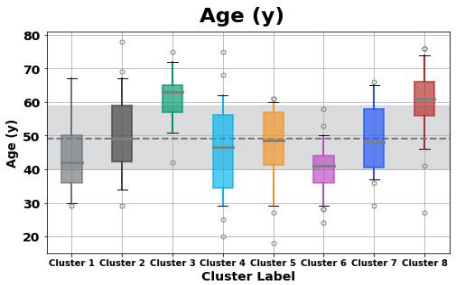
Cluster centers normalized with respect to the mean

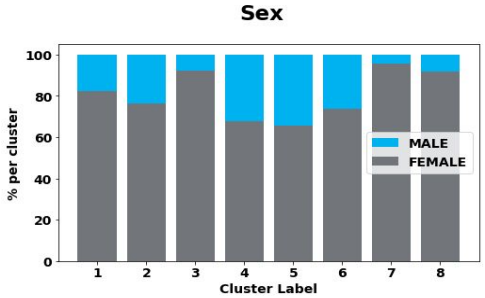


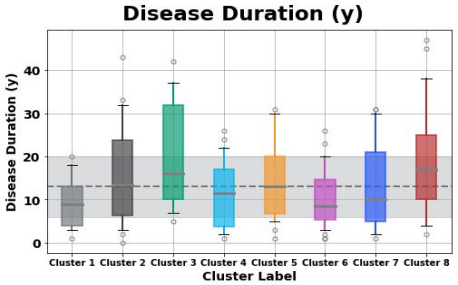
Clustering on Training data

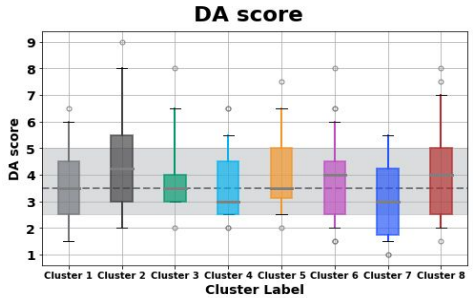
Breakdown of clinical and demographic variables

Clusters have distinct demographic and clinical characteristics







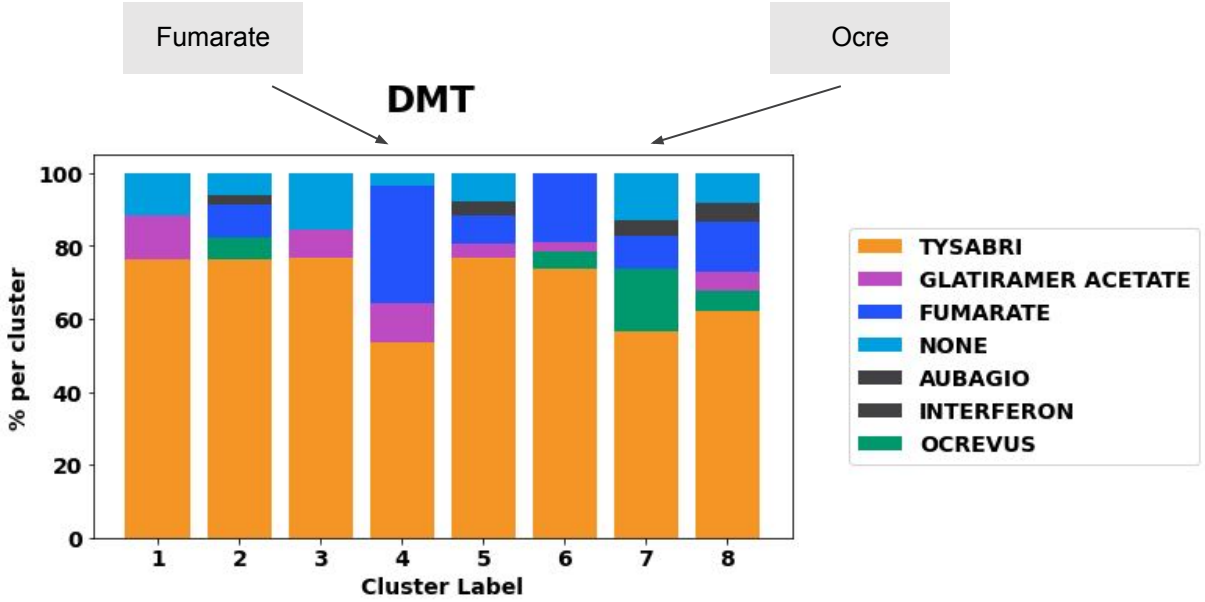


Clustering on Training Data

Do these groups contain meaningful information?

We see patients on Ocre mostly emerge in cluster 7 and patients on fumarates in cluster 4

Note: This cohort is very Tysabri heavy

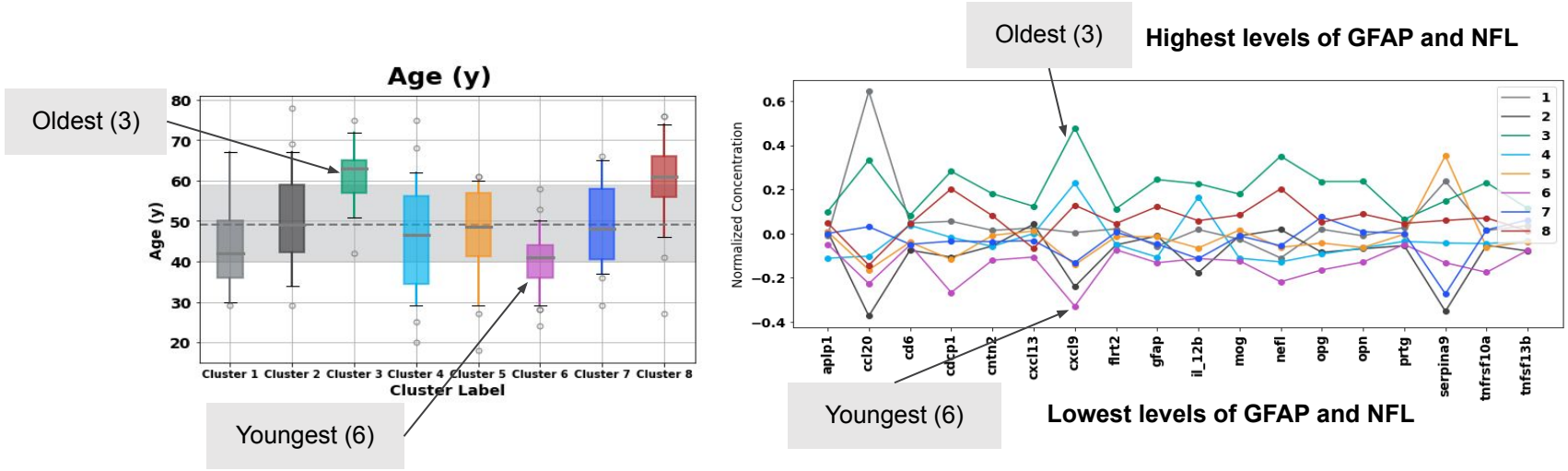


Clustering on Training Data

Do we see replication of expected results?

Results align with clinical expectations

- Cluster 3 contains patients with highest median age, and cluster 6 patients with lowest median age.
- For older patients, we expect nfl and gfap to be high. We expect the inverse for younger patients.



Robustness of Clusters:

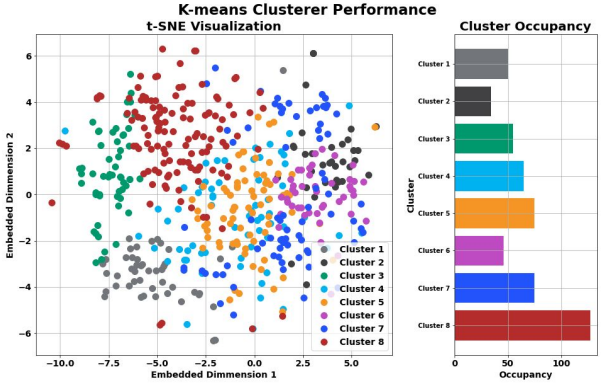
- Similarity in independent datasets

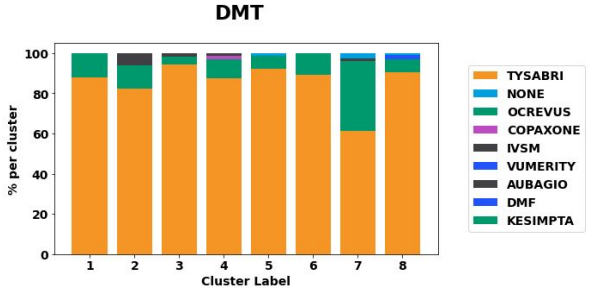
Robustness of Clusters

Methods

- We applied the grouping model learned on Training data to 527 samples from an independent cohort, unseen by the model (Test data).
- We assessed the distribution of the same variables in the new dataset.
- We also further assessed stability across cohorts using clustering quality metrics.

Clustering on Test Data Results



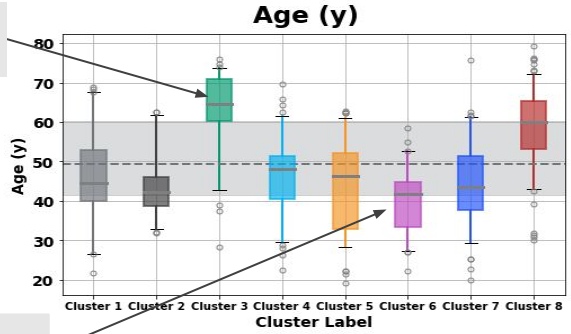


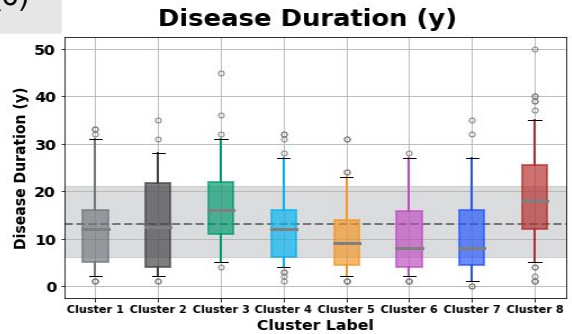
Clustering on Test data

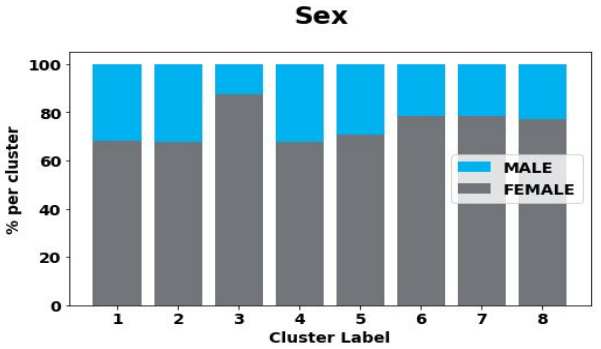
Clusters have distinct demographic and clinical characteristics similar to the training results

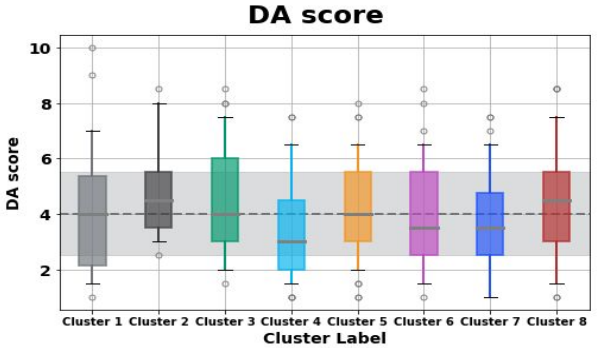
Oldest (3)

Youngest (6)





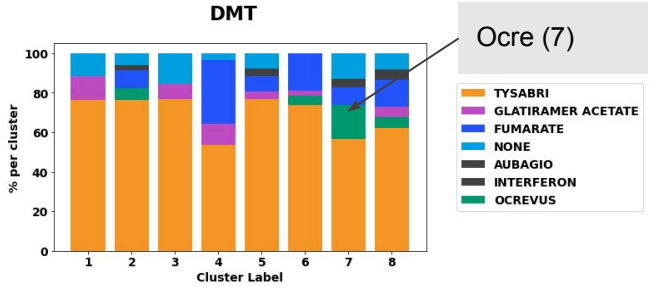


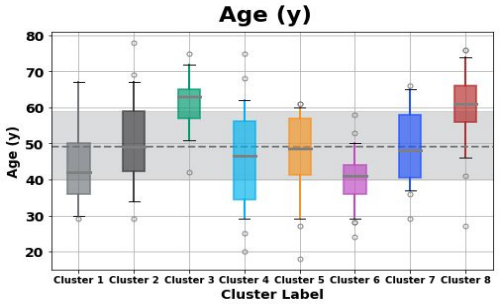


Robustness Results - Qualitative Evidence

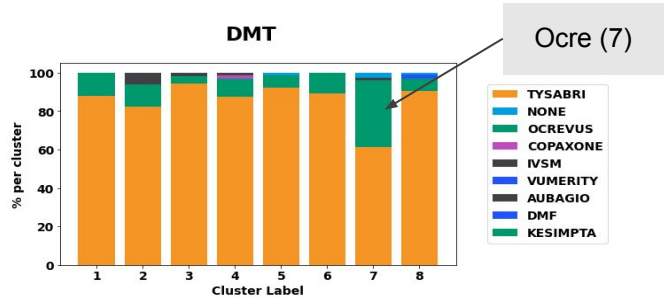
Training and test results are qualitatively similar

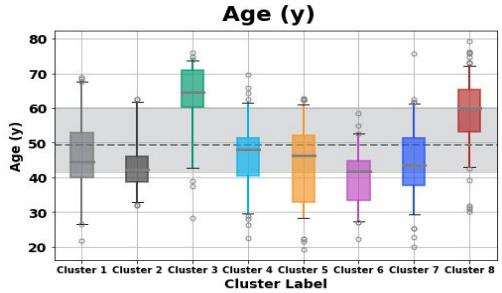
Training (220 Samples)





Test (527 Samples)

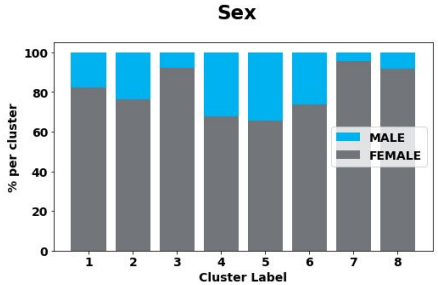


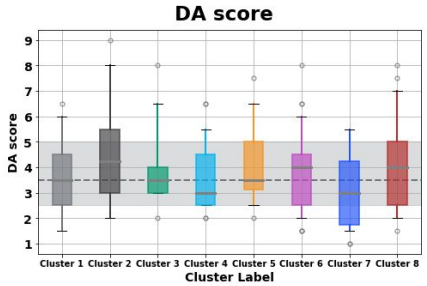


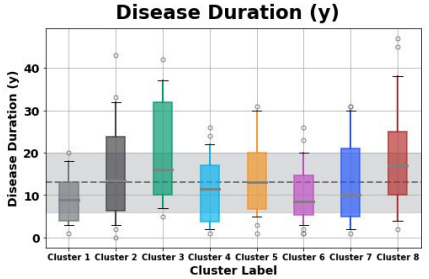
Robustness Results - Qualitative Evidence

Training and test results are qualitatively similar

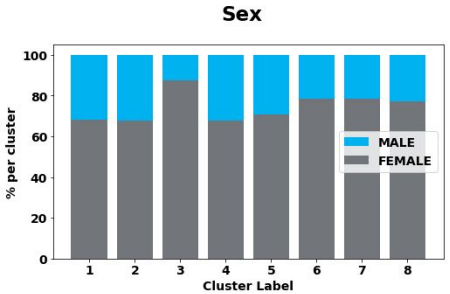
Training (220 Samples)

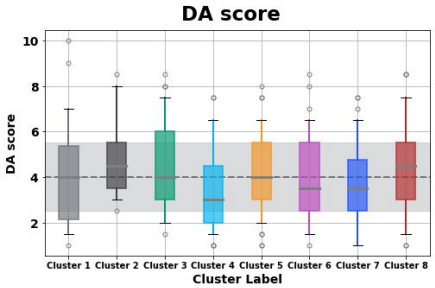


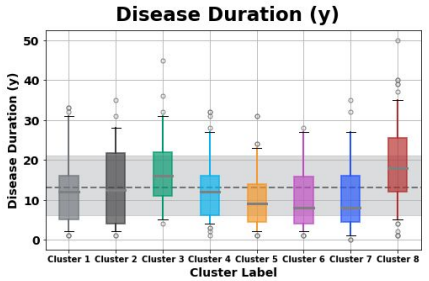




Test (527 Samples)



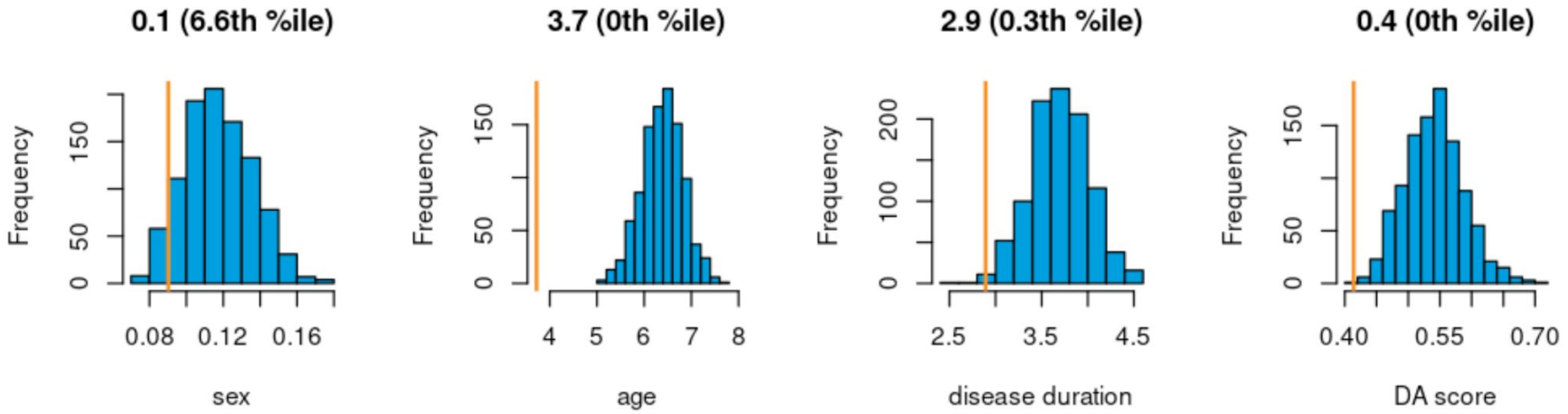




Robustness Results - Quantitative Evidence

Training and Test cluster distributions are extremely similar compared to random clusters.

These results complement the qualitative evidence for reproducing phenoclusters in unseen data.



Sex distributional distance metric = absolute difference in proportions between train and test clusters, mean of the 8 clusters is shown
Age, disease duration, DA score distributional distance metric = Wasserstein Distance between train and test clusters, mean of the 8 clusters is shown

Indication of Subclinical Signal

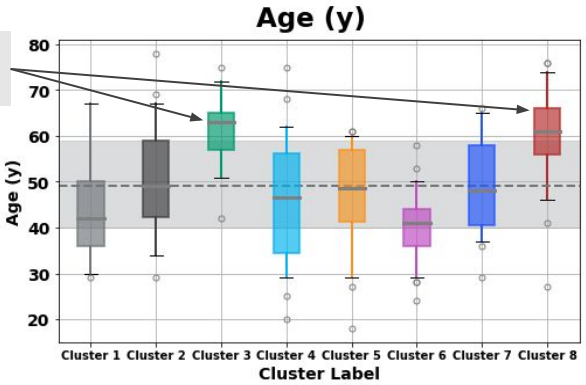
Clustering on Training Data

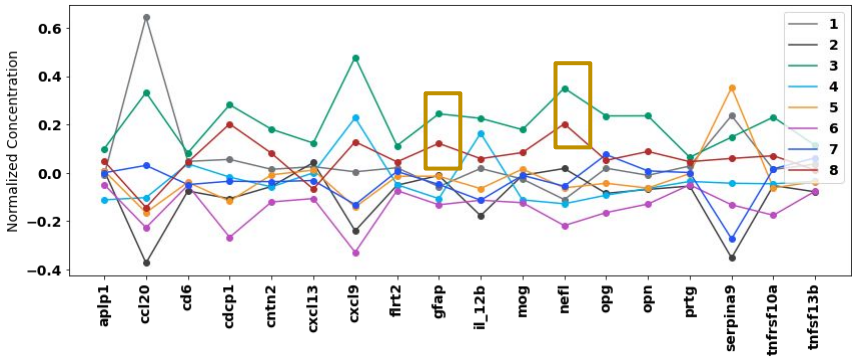
Is there subclinical signal contained in the clusters that could be clinically actionable?

Yes, some clusters are clinically and demographically similar, but have different protein signatures

- Clusters 3 & 8 contain oldest patients
- These groups have highest nfl and gfap levels as expected (2-sided t-test p-values < 10^{-12})
- In addition to nfl and gfap, clusters 3 & 8 have significantly higher levels of **cdcp1, cntn2, flrt2, mog, opn,** and **tnfrsf10a** (p-values ≤ 10^{-7})

Oldest (3,8)

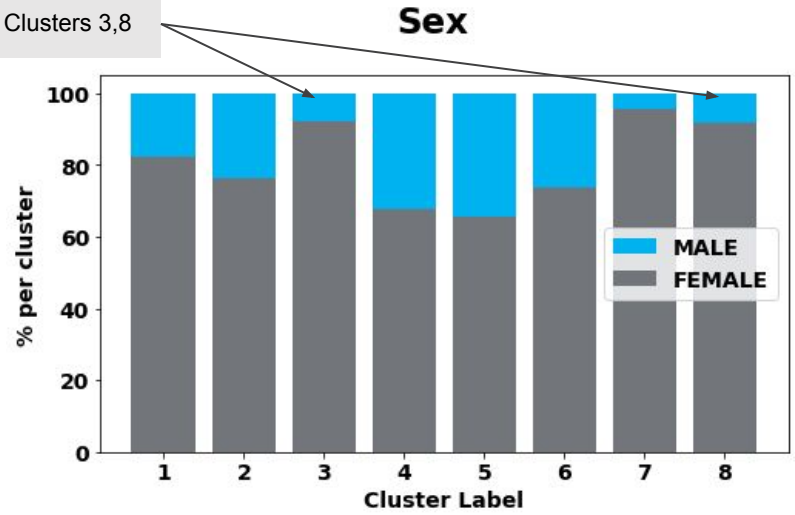




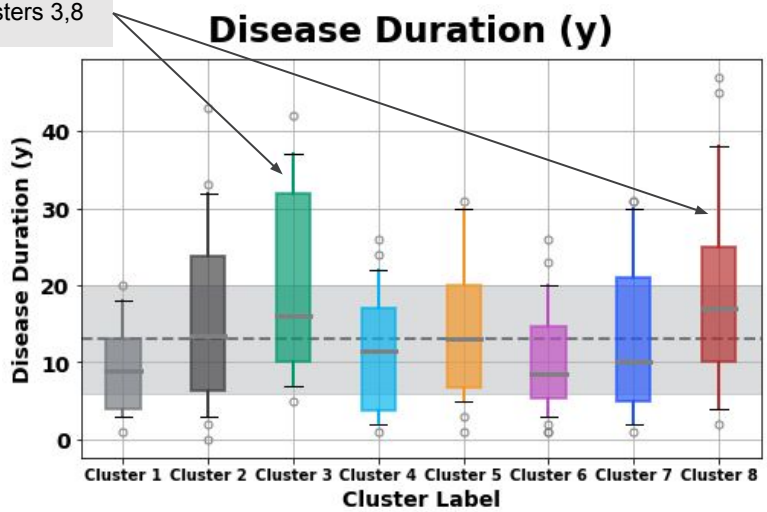
Clustering on Training Data

Clusters 3 and 8 are demographically similar and clinically mostly stable

Clusters 3,8



Clusters 3,8

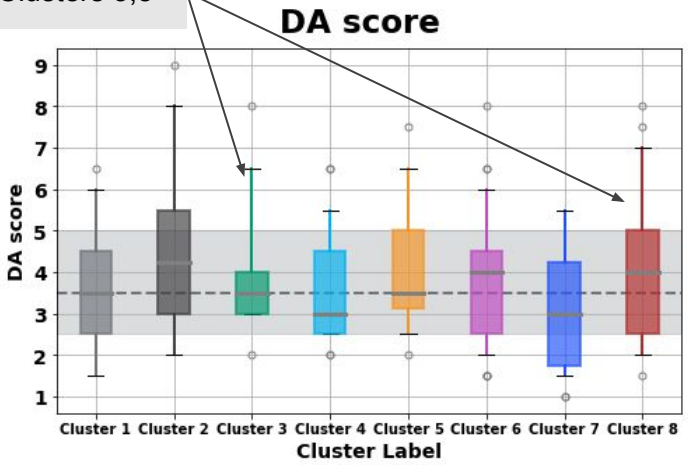


Clustering on Training Data

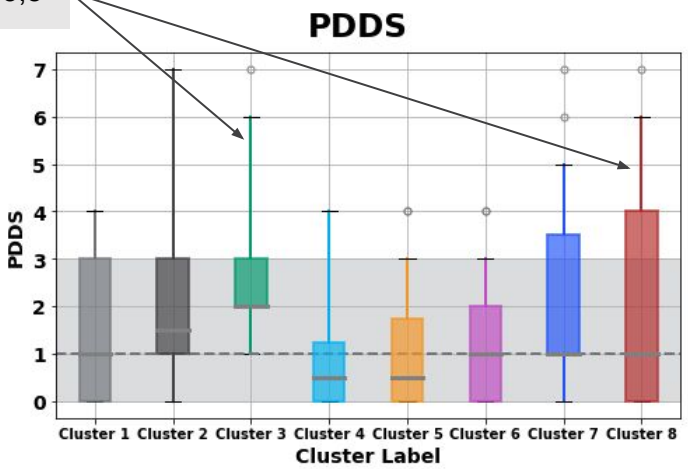
Is there an indication of subclinical differences?

Cluster 8 has higher median DA score and lower median PDDS than cluster 3

Clusters 3,8



Clusters 3,8



Summary & Future Work

Summary

- We described how we have used advanced machine learning and statistical techniques to identify emergent distinct biomarker profiles that deepen proteomic-based MS subtyping and the clinical interpretability of Octave MSDA results.
- We showed that the groupings resulted from this method
  - Align with clinical expectations
  - Are robust when tested on independent datasets
  - Could indicate subclinical signal

Future Work

- Expand the training and test sets to include more sites and a more balanced representation of DMTs
- Apply this technique to other data streams such as MRI metrics
- Apply this technique to data from multiple Octave dimensions, such as Biomarker + MRI metrics
- Explore clusters learned on isolated DMT data and indication of different response types

Appendix

Robustness of Clusters:

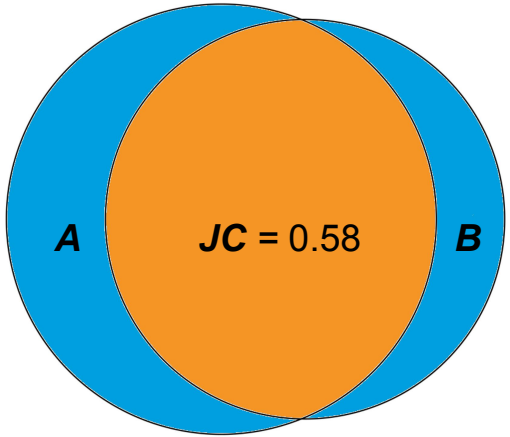
- Similarity in independent datasets
- Stability in repeat (resampled) data

The Jaccard Coefficient

A metric of stability in repeat samples

- The unsupervised clustering ***gold standard*** retrains clusters for each repeat sample; gold standard clusters are not meaningful across cohorts
- The ***Jaccard Coefficient*** measures overlap for analogous predicted and gold standard clusters in a given sample of patients
- JC > 0.6 reflects ***stability*** in two ways:
  1. Our method discovers consistent clusters in repeat samples
  2. Our method accurately predicts (1) when presented with new patient data

$$JC = \frac{intersection(A,B)}{union(A,B)}$$



A = predicted cluster
(clusters discovered in training data, predictions on test data)

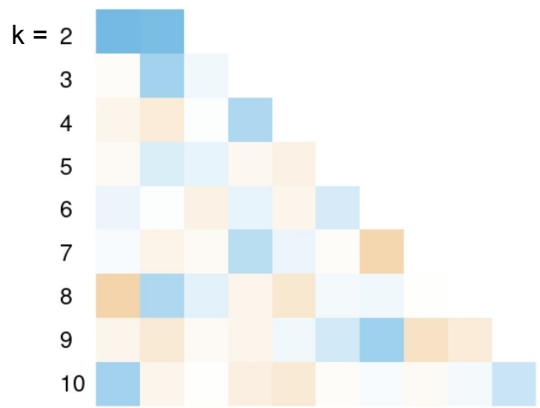
B = gold standard cluster
(clusters discovered in and predicted on test data)

Stability of Clusters

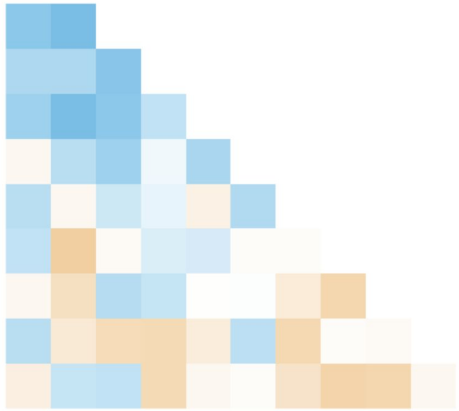
Phenoclusters are more stable with increased sample size

Clusters are consistently and accurately predicted in repeat samples These results give us confidence that the proof of concept methodology scales to other MS populations.

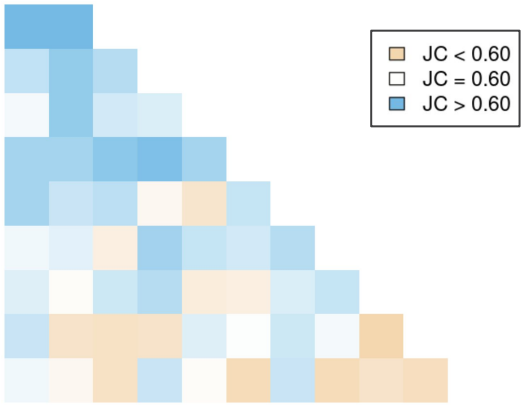
Training data (N=213 subjects)



Test data (N=325 subjects)



Octave cohort (N=1605 subjects)



JC = Jaccard coefficient, a metric that measures overlap of predicted and gold standard clusters to indicate evidence for stability. Values >= 0.60 indicate clusters are stable.