Genomic Subgrouping of ALS Patients to Investigate Difference in Clinical Progression

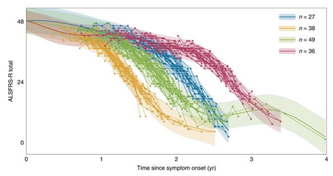
PCM Team Kiwi

Shaili Tripathi[1], Parimala Vedula[1], Yike Guo[1] ,Trinity Cheng[1], Ryder Li[1], Michael Potanin[1]
Advisors: Joseph Greenstein PhD[1], Casey Overby Taylor PhD[1], Alyssa Coyne PhD[2], Jeffrey Rothstein MD, PhD[2]

[1]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD | [2]Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD

Johns Hopkins Whiting School of Engineering

Introduction



Amyotrophic lateral sclerosis (ALS):
- Fatal and rapidly-progressing neurodegenerative disease
- 30,000 US cases, mean survival 2-5 y
- Clinical and genetic heterogeneity makes disease management and treatment development difficult

Fig. 1. ALS patients display distinct clustered progression[1]

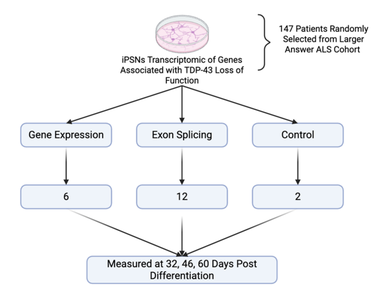
Given this heterogeneity, the ability to assign patients to meaningful subgroups offers the potential of an improved understanding of biological pathways, allowing for better-targeted interventions and opening up the possibility for future personalized treatments.

Aims

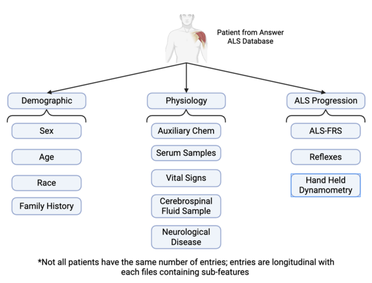
Hypothesis: clustering with TDP-43 data will show meaningful patient subgroups
- Create ALS patient subgroups through clustering of novel iPSC data focused on TDP-43 loss of function
- Identify clinical correlates of patient subgroups, with an emphasis on sporadic ALS patients

Data Overview

Genomic Data



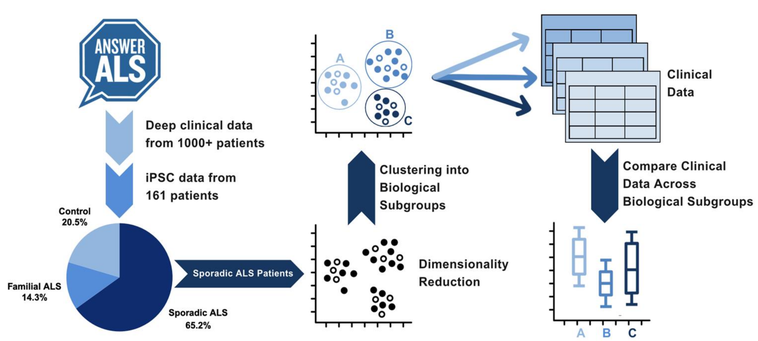
Clinical Data



*Not all patients have the same number of entries; entries are longitudinal with each files containing sub-features

Methods

iPSC Data

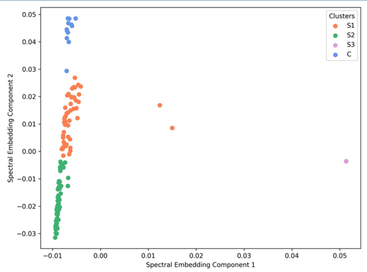


Clustering

Clinical Correlates

Results

Clustering



Cluster	# C9orf72	# sALS
S1	0	51
S2	0	60
S3	0	18
C	12	0

Fig. 2. Spectral embedding with K-means creates four clusters.
Table 1. The number of patients of each mutation in each cluster.

Genetic Features Grouped by Clustering

After generating the patient clusters visualized in Figure 2, we examined the underlying genetic composition of each cluster. We tested for differences in gene distribution across clusters and identified 14 genes with statistically significant variation ($p < 0.01$). These genes include: *myo18a ce, dnm1 ce, hdgfl2 ce, syt7 ce, camk2b ce, elavl3, rcan1, stmn2, actl6b ce, arhgap32 ce, selplg, pfkp, cdk7 ce, truncated stmn2*

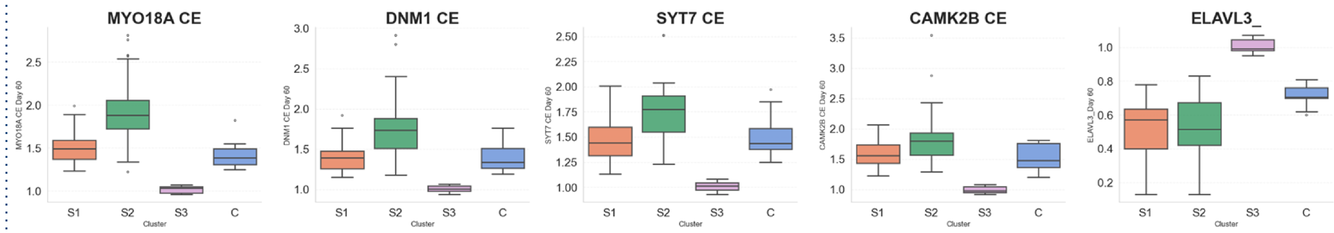


Fig. 3. Expression levels of five representative genes across patient clusters, illustrating distinct distribution patterns between clustered subgroups.

Clinical Features Grouped by Clustering

A linear mixed-effects model (LMM) was used for longitudinal analysis of clinical outcomes with significant inter-cluster differences. Cluster C was used as the reference group.Interaction terms between time (Days Since First Visit) and Cluster Label, allowing us to assess whether the rate of clinical change over time differed by cluster. A significant interaction term indicates that the slope of change (i.e., disease progression) varies between clusters.Tukey's HSD Test was used for post-hoc analysis. In total, 32 outcomes showed significant time × cluster interactions, highlighting distinct longitudinal progression patterns across patient subgroups.

ALSFRS Subscore	Interaction Effects (Slope)		
	S1	S2	S3
Speech	+0.003 (p<0.001**)	+0.003 (p<0.001**)	+0.002 (p=0.020*)
Salivation	+0.003 (p<0.001**)	+0.004 (p<0.001**)	+0.003 (p=0.001**)
Swallowing	+0.003 (p=0.001**)	+0.002 (p=0.010**)	+0.002 (p=0.042*)
Handwriting	+0.002 (p=0.027*)	+0.002 (p=0.019*)	+0.002 (p=0.032*)
Cutting w/o Gastrostomy	+0.004 (p<0.001**)	+0.004 (p<0.001**)	+0.003 (p=0.002**)
Walking	+0.003 (p=0.001**)	+0.004 (p=0.001**)	+0.003 (p=0.022*)
Climbing Stairs	+0.004 (p<0.001**)	+0.004 (p<0.001**)	+0.003 (p<0.001**)

Table 2. Significant ALSFRS subscores: measures functional ability, higher scores indicate better functioning.

Reflex Score	Interaction Effects (Slope)		
	S1	S2	S3
L Biceps	+0.004 (p=0.005**)	+0.004 (p=0.006**)	+0.004 (p=0.004**)
L Brachioradialis	+0.004 (p=0.018*)	+0.004 (p=0.017*)	+0.004 (p=0.014*)
L Hoffman Sign	+0.002 (p=0.011*)	+0.002 (p=0.018*)	+0.002 (p=0.017*)
L Triceps	+0.003 (p=0.073*)	+0.003 (p=0.083)	+0.002 (p=0.140)
R Biceps	+0.004 (p=0.008**)	+0.004 (p=0.003**)	+0.004 (p=0.008**)
R Brachioradialis	+0.003 (p=0.018*)	+0.003 (p=0.013*)	+0.003 (p=0.012*)
R Hoffman Sign	+0.003 (p=0.009**)	+0.003 (p=0.019*)	+0.003 (p=0.009**)
R Triceps	+0.003 (p=0.067*)	+0.003 (p=0.085)	+0.002 (p=0.120)

Table 3. Significant reflex scores. Higher scores indicate increased responsiveness.

Fig. 4. ALSFRS-Speech distribution by cluster.

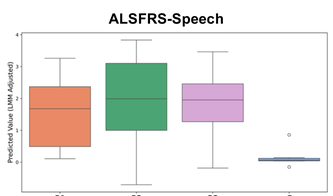
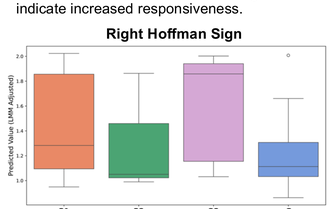


Fig. 5. Right Hoffman sign reflex score distribution by cluster.



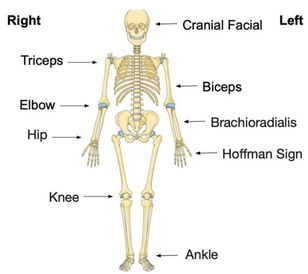
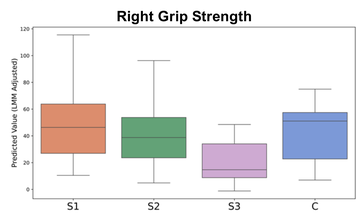


Fig. 6. (above) Labeled reflexes examined by tests[2].

Other	Interaction Effects (Slope)		
	S1	S2	S3
R Grip Strength	+0.003 (p=0.049*)	+0.003 (p=0.066)	+0.004 (p=0.018*)
Cognitive Behavioral Total	+0.009 (p=0.008**)	+0.009 (p=0.006**)	+0.003 (p=0.422)

Table 4. Other significant features including functional assessments and neurological tests.

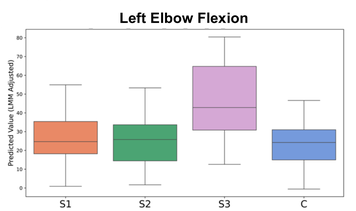
Fig. 7. Right grip strength distribution by cluster.



Muscle Strength	Interaction Effects (Slope)		
	S1	S2	S3
R Elbow Extension	+0.003 (p=0.049*)	+0.003 (p=0.066)	+0.004 (p=0.018*)
R Knee Extension	+0.057 (p=0.010*)	+0.054 (p=0.013*)	+0.043 (p=0.075)
R Ankle Dorsiflexion	+0.025 (p=0.185)	+0.023 (p=0.219)	+0.083 (p<0.001**)
L Elbow Extension	+0.003 (p=0.033*)	+0.003 (p=0.074)	+0.004 (p=0.015*)
L Knee Flexion	+0.022 (p=0.266)	+0.025 (p=0.196)	+0.054 (p=0.014*)
L Knee Extension	+0.038 (p=0.048*)	+0.036 (p=0.060)	+0.005 (p=0.817)

Table 5. Hand held dynamometry scores. Higher scores indicate greater muscle strength.

Fig. 8. Left elbow extension distribution by cluster.



Conclusions and Future Work

- Spectral embedding combined with K-means clustering was applied to iPSC gene expression data focused on TDP-43 markers, resulting in the identification of three distinct ALS patient subgroups that highlight differences that contribute to variations in clinical symptoms and progression, particularly in ALSFRS scores, muscle strength, and reflex responses.

Future work:
- Apply similar clustering methods to whole genome sequencing data to determine if comparable subgroup patterns are observed at the genomic level.
- Evaluate the consistency and robustness of these clusters across larger patient populations as additional iPSC-derived transcriptomic datasets become available.

References:

1. Ramamoorthy, D., Severson, K., Ghosh, S. *et al.* Identifying patterns in amyotrophic lateral sclerosis progression from sparse longitudinal data. *Nat Comput Sci* 2, 605–616 (2022). https://doi.org/10.1038/s43588-022-00299-w
2. Les Laboratoires Servier. (n.d.). *Servier Medical Art*. https://smart.servier.com/smart_image/skeleton-and-cartilage-face/