

Deep Learning of Plasma Proteomics Reveals Molecular Heterogeneity in Early-Stage Alzheimer’s Disease

Buddhiprabha Erabadda¹, Muhammad Ali², Hanjun Zhao¹, Carlos Cruchaga², Alejo Nevado-Holgado¹, Laura Winchester¹

¹Department of Psychiatry, University of Oxford, UK

²Department of Psychiatry, Washington University School of Medicine, St. Louis, USA

Contact: buddhiprabha.erabadda@psych.ox.ac.uk and laura.winchester@psych.ox.ac.uk

P1-001

In this study, we analysed plasma proteomics of early-mid stage AD cases ($15 \leq \text{MMSE} \leq 30$, age ≥ 65) cross-sectionally to identify molecular signatures that are not detectable with standard cognitive testing. For this, we used a graph-smoothed autoencoder with a STRING protein-protein interaction prior to derive deep-learning based latent representations followed by clustering. It revealed an Endolysosomal–Translation bipolar axis (a bidirectional axis of coordinated endolysosomal and translational protein changes) within a disease continuum. The proposed framework supports mechanism-based stratification for precision medicine and trial enrichment using minimally invasive methods.

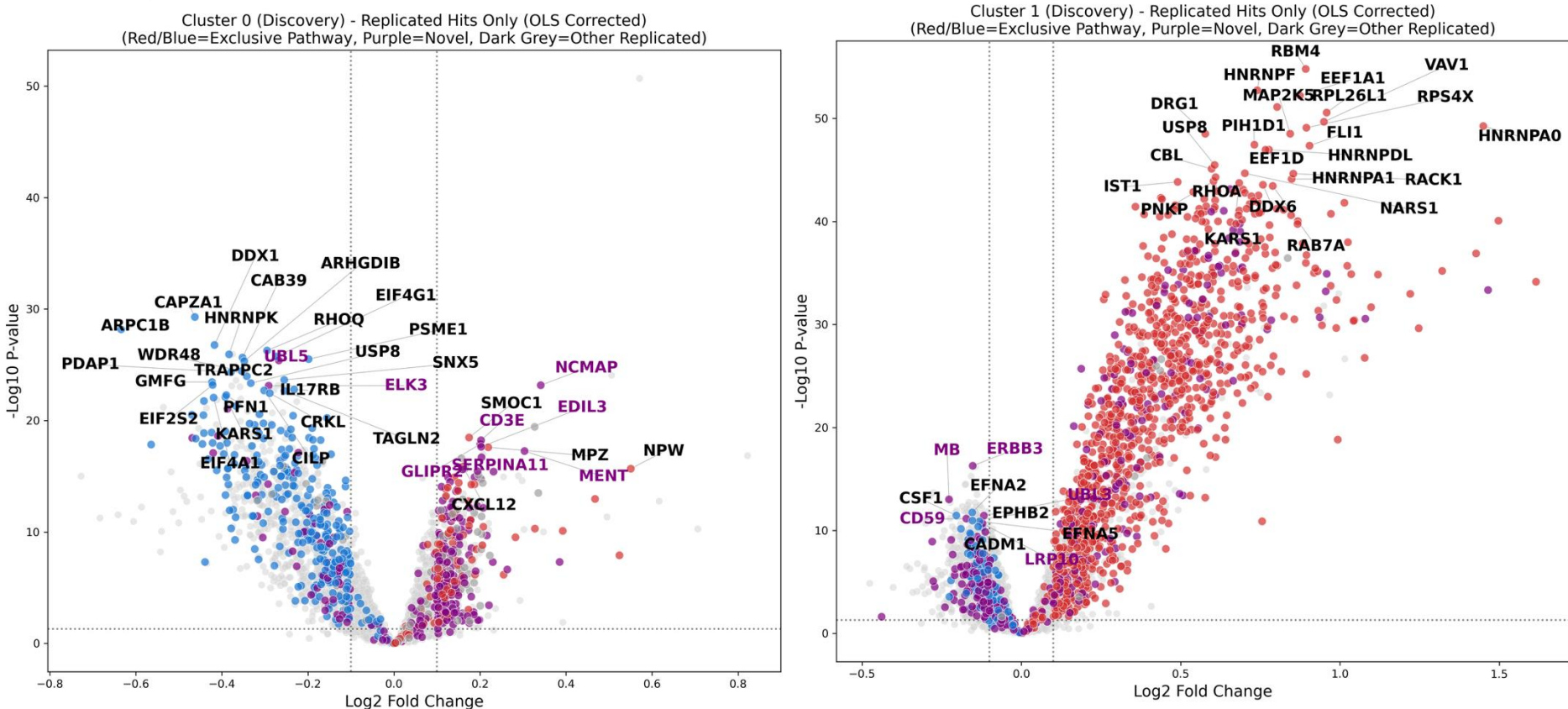
INTRODUCTION

Clinical staging of early-stage Alzheimer’s Disease (AD) relies heavily on cognitive scores (e.g., MMSE), which can mask underlying biological heterogeneity. While patients may present with similar mild impairment, their molecular drivers can vary, obscuring distinctions relevant to clinical trials and therapeutic targeting. We hypothesised that deep learning applied to plasma proteomics could resolve this molecular heterogeneity in early-stage disease.

Our analysis revealed a continuum of the disease with a distinct Endolysosomal-Translation based bipolar axis comprising two clusters that was robustly replicated in an independent cohort – both with the discovery-trained cluster centroids and independently trained AE + de novo clustering. Furthermore, we discovered a third cluster that exhibited a complement-coagulation/metabolic-suppression signature.

RESULTS

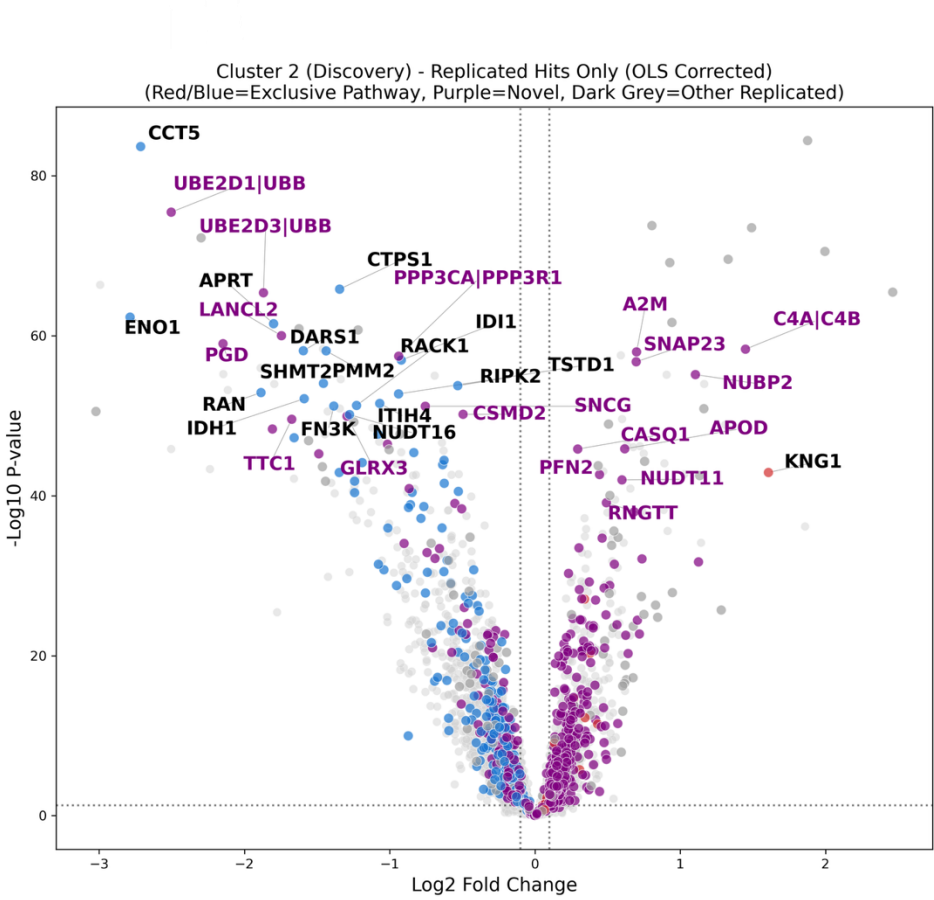
Volcano plots for the clusters in the Endolysosomal-Translation based bipolar axis



Cluster 1 exhibited coordinated upregulation across endolysosomal-trafficking and translation programmes (554 enriched pathways; FDR < 0.05), dominated by Endocytosis (KEGG adj-p = 2.6×10^{-12}), Translation (Reactome adj-p = 6.5×10^{-12}), Cellular Responses to Stimuli (adj-p = 1.1×10^{-11}), Membrane Trafficking and Signaling by Rho GTPases (adj-p = 1.1×10^{-9}).

Cluster 0 exhibited 9 upregulated and 479 downregulated pathways, with the suppressed pathways closely mirroring cluster 1’s upregulation (per-gene fold-change $r = -0.705$).

Volcano plot for the complement-coagulation/metabolic-suppression

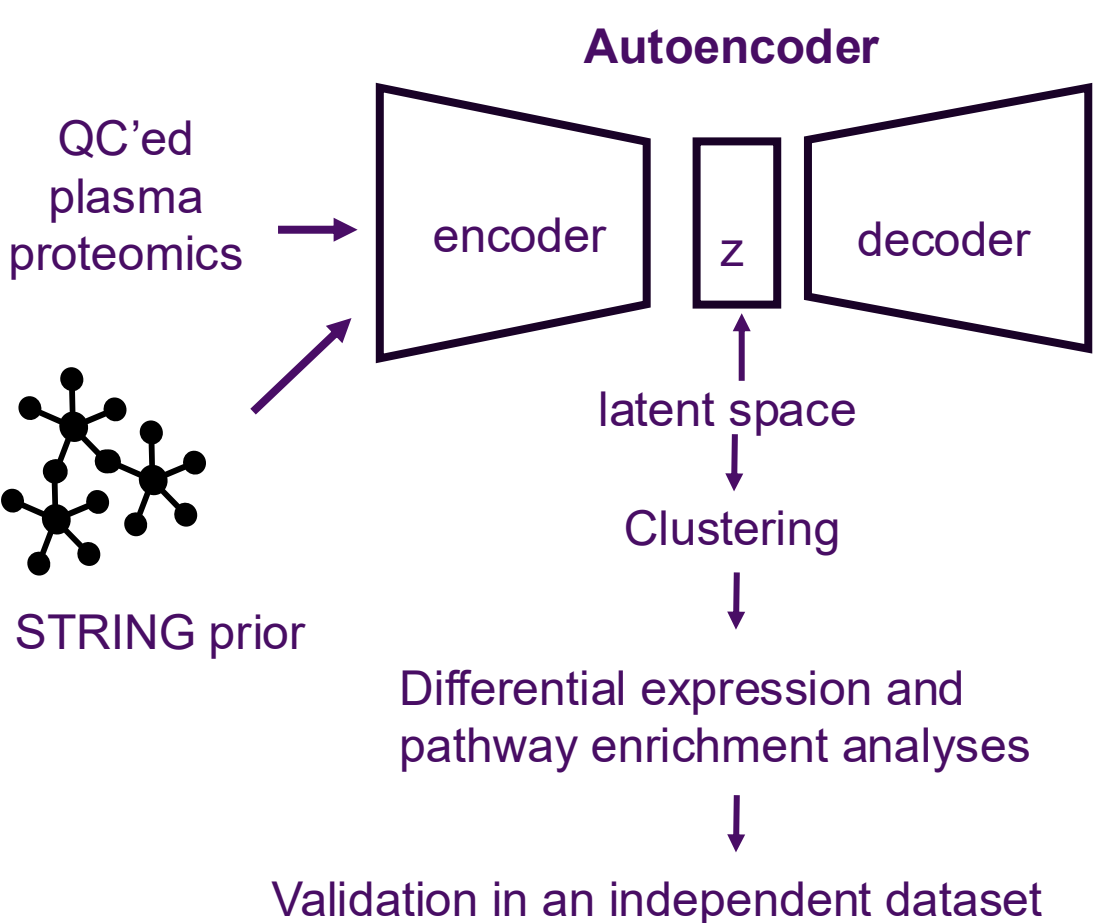


This cluster did not align with the primary bidirectional axis, instead representing a second, orthogonal dimension. It was defined by 0 upregulated and 145 downregulated pathways, with the downregulated pathways centred on Metabolism (Reactome adj-p = 1.9×10^{-7}), Translation (adj-p = 6.6×10^{-4}), Eukaryotic Translation Elongation (adj-p = 4.1×10^{-5}), Pyruvate Metabolism (KEGG adj-p = 4.1×10^{-5}), and Cellular Responses to Stimuli (adj-p = 5.4×10^{-5}) alongside upregulation of complement and coagulation components (F2, KNG1, MASP1). In the independent validation cohort, this appears as a bidirectional axis (independent-validation cluster 0 vs cluster 2, per-gene $r = -0.90$; aptamer-level $r = -0.89$). This second axis was only partially resolved in discovery at K=3, consistent with the continuous, low-silhouette structure of the latent space.

CONCLUSIONS

- Plasma proteomics + deep learning reveals reproducible molecular structure in early-stage AD — invisible to cognitive testing.
- Identified an **Endolysosomal–Translation bipolar axis**, replicated in an independent cohort and an independently trained model.
- Plus a distinct third state: **complement-coagulation activation with metabolic suppression**.
- The axis poles move in opposite directions — an intervention on an axis protein could help one pole while being neutral or harmful at the other, relevant to trial stratification.

METHODS



- Two AD cohorts (discovery/validation, n=898/954) from the Global Neurodegeneration Proteomics Consortium (GNPC), 8 contributors total.
- Graph-smoothed autoencoder with a STRING protein–protein interaction prior derives a 128-D latent space from 7K SomaScan plasma proteins.
- k-means clustering (K=3) on the latent space; differential expression and pathway enrichment vs. age- and sex-matched AD-negative controls.

LIMITATIONS

- We observed a continuous axis, not discrete subtypes. The labels show regions on the axis.
- The axis is largely independent of clinical variables, so progression and treatment-response require large longitudinal data.

CONTACT



Scan the
QR code
to get in
touch