Plasma Proteomic Landscape of Alzheimer's Disease: An 1800-Sample Cohort Study

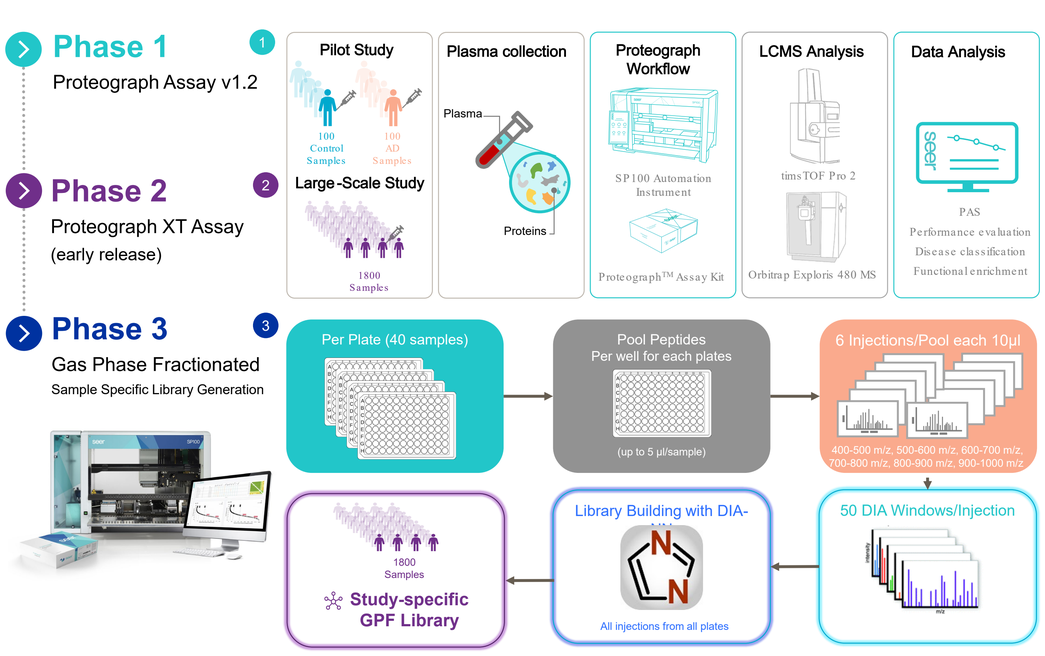
Khatereh Motamedchaboki*[1], *Benjamin Lacar*[1], *Amir Alavi*[1], *Guhan R Venkataraman*[1], *Harendra Guturu*[1], *Daniel Hornburg*[1], *Matthijs de Geus*[2], *Sudeshna Das*[2], *Pia Kivisakk*[2], *Serafim Batzoglou*[1], *Steven E. Arnold*[2] *and Asim Siddiqui*[1]

Deep and unbiased plasma proteomics for a large AD cohort

Biofluids, like serum or plasma are a rich source of protein biomarkers for mechanistic understanding and early detection of disease, but the wide dynamic range of protein concentrations in plasma often require complex proteomics workflows and trade-offs between protein coverage and analysis throughput. Meanwhile, the mechanisms underlying the development and progression of Alzheimer's Disease (AD) remain poorly understood. Here, we use an unbiased, deep, and rapid proteome interrogation approach to investigate AD. This approach leverages physicochemically distinct nanoparticles to provide broad coverage of the plasma proteome at scale.

We previously reported on a 200-plasma sample study of 100 AD-affected and 100 healthy age- and sex- matched controls (HC), which were analyzed with the Proteograph™ Product Suite (Seer Inc.) and liquid-chromatography mass-spectrometry analysis. Here, we report an extended study of 1800 samples and longitudinal samples from the same individual. We plan to investigate a variety of aspects related to Alzheimer's disease including a) a better classifier for neurocognitive disorders of aging including Alzheimer's disease, frontotemporal dementias, Lewy body disorders and healthy cognitive aging; and b) Identifying potential biomarkers and pathophysiological pathways associated with heterogeneity in Alzheimer's disease including, age of onset and rate of progression.

Study Design



Results

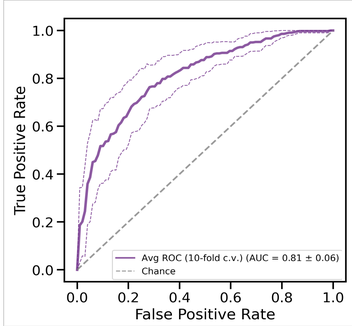
Figure 2. Technical and biological CV in the large AD cohort. Technical CV measured using sample replicates on each plate. Technical CV well below CV of biological samples indicating strong biological signal in the dataset. The GPF library trends above the other two libraries indicating that GPF library has higher quality calls from the perspective of precision.

Proteograph XT Workflow Performance

Figure 1. Number of quantified protein groups. Raw LC-Ms data were analyzed with DIA-NN in library-free search mode. We identified over 5,000 protein groups across the cohort (on average >2,500 per sample) with the majority being supported by multiple peptides and found in multiple samples.

Disease classification and biomarker characterization of cognitive disorder cohort

A Classification of healthy vs. dementia — AD Study



N = 470 Normal controls

N = 392 Dementia patients

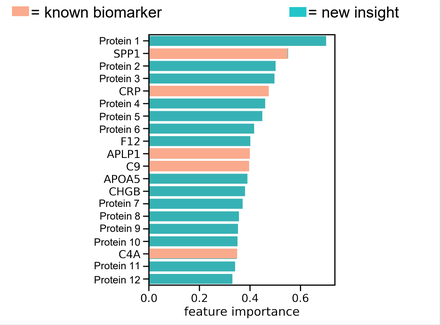
Greedy ensemble selection results in more powerful final model

Classification performance: AUC = 0.81

Model selection and tuning using a structured Automated ML pipeline (with Bayesian Optimization)

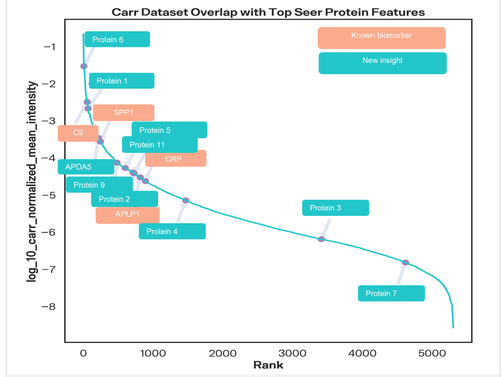
Model reveals distinct differences between normal and dementia patients with good performance.

A Top 20 classifier features



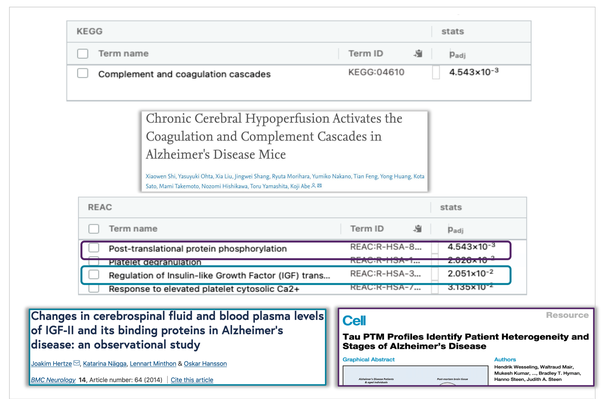
Top classification features include known dementia biomarkers and potentially novel insights/biomarkers

C Depth of top 20 protein features



Identified low abundant, potentially novel biomarkers, suggesting scale/depth achieved with Proteograph XT workflow supports enhanced biological insights

D Functional enrichment of top features



Top 100 features are enriched for dementia-associated functional pathways

Figure 3: Deep plasma proteomics analysis and AD classification with Proteograph workflow. (A) ROC curve of healthy vs dementia, **(B)**Top 20 classifier features highlighting known biomarkers **(Orange)** and novel target identification **(Teal)**, **(C)** Distribution of top 20 features against HPPP standard intensities, **(D)** Functional enrichment of top 20 features identifying dementia associated pathways.

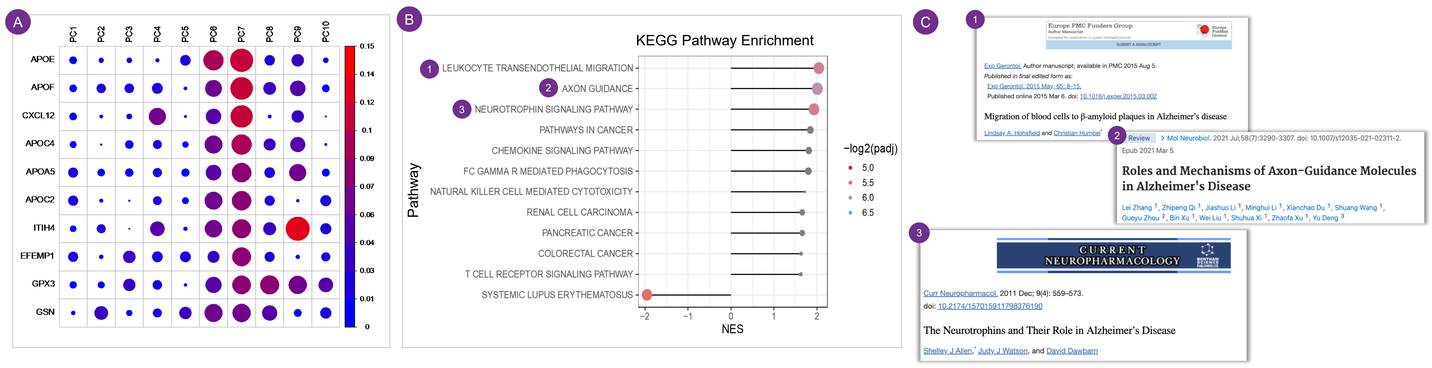
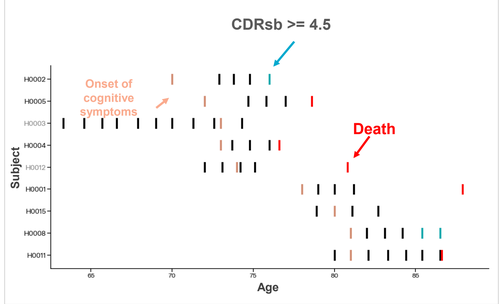
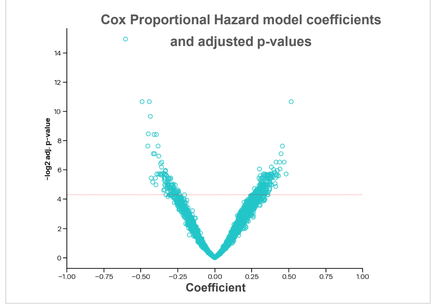


Figure 4. Differential proteomics analysis of the AD cohort study: (A) PCA unsupervised learning reveals features associated with dementia demonstrated by **(B)** pathway enrichment and **(C)** confirming pathways previously reported to be implicated in AD in addition to new targets.

A Clinical progression of dementia



B Proteins associated with differences in dementia progression



C Dose response of selected protein with time to CDRsb >= 4.5

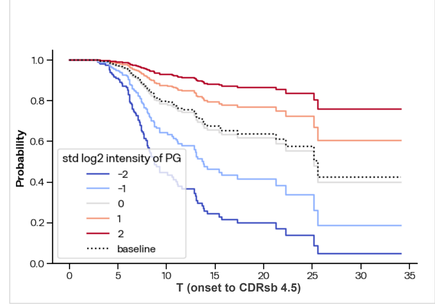


Figure 5. Biomarkers distinguishing the fast progression of dementia. Clinical Dementia Rating Sum of Boxes (CDRsb) Scores for Dementia Subjects **(A)** Raster plot example to illustrate Time-to-Event analysis. Black y-axis labels are subjects that were diagnosed with AD. "Time" is defined from the onset of symptoms and subject progressing to CDRsb>=4.5 (E=1) or their last documented record (E=0). **(B)** Volcano plot of Cox Proportional Hazard (CPH) model protein coefficients. Red dashed line indicates p-value threshold of 0.05 with BH correction. **(C)** Example time-to-event curves showing the partial effect contribution of selected protein which has a negative CPH model coefficient. Higher levels of this protein are associated with decreased dementia progression.

Conclusion

- 1,800 samples run in 11 weeks on a single MS instruments generating over 5,000 protein groups across the cohort
- AutoML framework generates a classification curve of dementia vs controls with an AUC-ROC 0.81.
- Enrichment analysis of top features reveal known pathways.

References

[1] Blume et al. *Nat. Comm.* (2020)
[2] Demichev et al. *Nat Comm.* (2020)
[3] Raudvere, Kolberg et al. *Nucleic Acids Res.* (2019)

Publications

Funded by SBIR grant

[1]Seer, Inc., Redwood City, CA 94065, USA | *ssiddiqui@seer.bio
[2]Massachusetts General Hospital, Boston, MA 02114 , USA