Large-scale, deep plasma proteomics: An 1800 sample study of Alzheimer's disease.

Asim Siddiqui[1], Harendra Guturu[1], Matthijs de Geus[2], Sudeshna Das[2], Pia Kivisakk[2], Serafim Batzoglou[1] and Steven E. Arnold[2]

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.

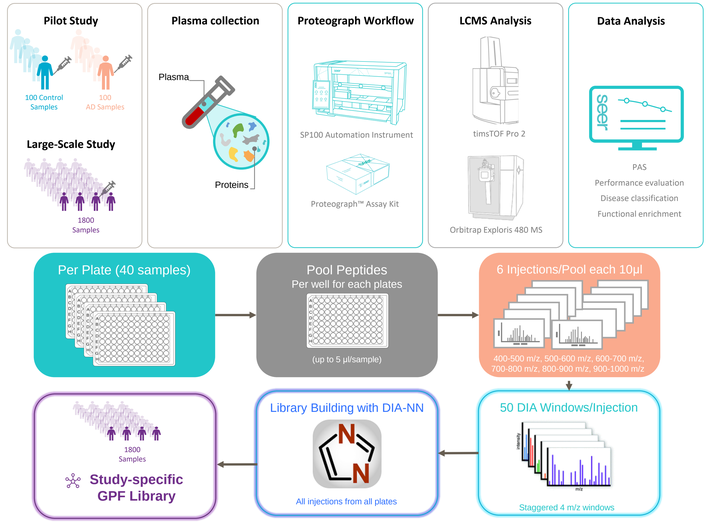
We previously reported on a 200-plasma sample study of 100 AD-affected and 100 healthy age- and sex- matched controls (HC), which were analyzed with the Proteograph™ Product Suite (Seer Inc.) and liquid-chromatography mass-spectrometry analysis. Here, we report an extended study of 1800 samples and longitudinal samples from the same individuals . 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.

Phase 1
Proteograph Assay v1.2

Phase 2
Proteograph XT Assay
(early release)

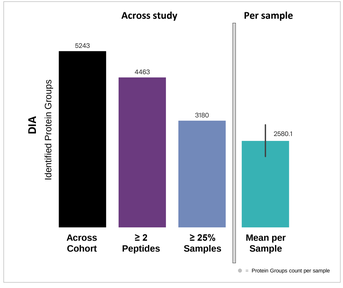
Phase 3
Gas Phase Fractionated (GPF)
(Sample Specific Library Generation)



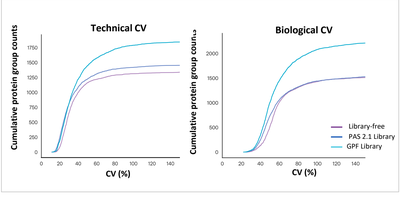


Deep proteomics enhances disease classification and biomarker characterization of cognitive disorder cohort

Results

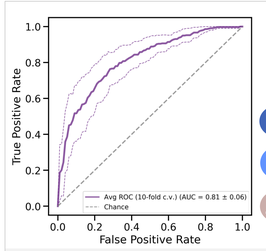


Number of quantified protein groups. Using DIA-NN, we identified over 5,000 protein groups across the cohort (on average >2,500 per sample) with the majority being supported by multiple peptides and found in multiple samples.



Technical and Biological CV in the 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 that library has higher quality calls from the perspective of precision.

Classification of healthy vs. dementia



N = 470
Normal controls

N = 392
Dementia patients

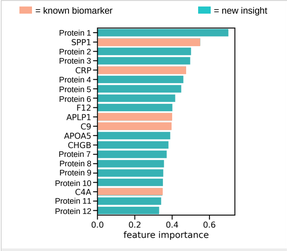
Greedy ensemble selection results in more powerful final model

Classification performance: AUC = 0.81

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

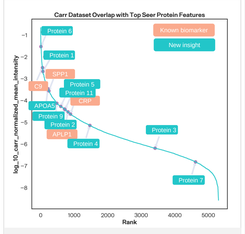
Model reveals distinct differences between normal and dementia patients with good performance.

Top 20 classifier features ranked by importance



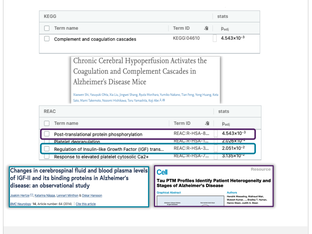
Top classification features include known dementia biomarkers and potentially novel insights/biomarkers

Depth of top 20 features

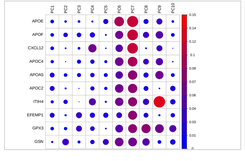


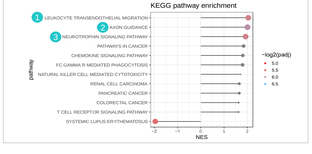
Identified low abundant, potentially novel biomarkers, suggesting scale/depth achieved with Proteograph XT workflow supports enhanced biological insights

Functional enrichment of top features



Top 100 features are enriched for dementia-associated functional pathways





Top Left: ROC curve of healthy vs dementia

Top Right: Top 20 classifier features highlighting known biomarkers

Far Left: Distribution of top 20 features against HPPP standard intensities

Near Left: Functional enrichment of top 20 features identifying dementia associated pathways.

Bottom Left: PCA unsupervised learning reveals features associated with dementia demonstrated by pathway enrichment (**Bottom middle** and **Bottom right**)



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

SBIR·STTR America's Seed Fund™ POWERED BY SBA
Funded by SBIR grant

Copyright Seer, Inc 2023

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