Deep plasma protein profiling in Alzheimer's subjects with a novel unbiased and scalable proteogenomics approach

***Asim Siddiqui*[1*]**, *Harendra Guturu[1], Mahdi Zamanighomi[1], Jian Wang[1], Ting Huang[1], Ryan Benz[1], Matthijs B. De Geus[2], Sudeshna Das[2], Pia Kivisäkk[2], Steven E. Arnold[2], and Serafim Batzoglou[1]*

Deep and unbiased plasma proteomics for disease cohort studies at scale

Plasma is a rich source of protein biomarkers for early detection of diseases, but the large dynamic range of protein concentrations necessitates complex workflow and trade-offs between throughput, coverage, and precision. We have analyzed plasma samples from 100 Alzheimer and 100 healthy matched controls, using a multi-nanoparticle approach combined with LC-MS analysis. Across samples, LC-MS analysis in data-dependent acquisition (DDA) mode yielded 36,496 peptides and 4,706 proteins and Data-independent acquisition (DIA) mode yielded 39,699 peptides and 5,060 proteins.

In summary, we identified a combination of known and potential novel plasma protein markers, demonstrating the utility of Proteograph™ Product Suite workflow for an unbiased, deep, and rapid interrogation of the plasma proteome, enabling large-scale studies to detect novel biological insights.

Proteograph Product Suite

Proteograph Product Suite provides untargeted, deep, and rapid proteomics at scale

From sample to peptides, ready for analysis on most LC-MS instruments with a variety of proteomics methods

Methods

Plasma samples from 200 subjects comprising 100 AD, and 100 healthy controls were analyzed using Proteograph plasma protein profiling platform[1].

Using 5 injections per sample, proteins were quantified using two liquid-chromatography mass-spectrometry (LC-MS) methods on Bruker timsTOF Pro:

1. 60-minute DDA method runs split over two LC-MS systems
2. 30-minute DIA method runs on single LC-MS

Protein intensities were used to develop models for disease state prediction. Feature selection for the classifiers was done using protein level t-test. Test/train separation was maintained end to tend.

Proteograph Workflow

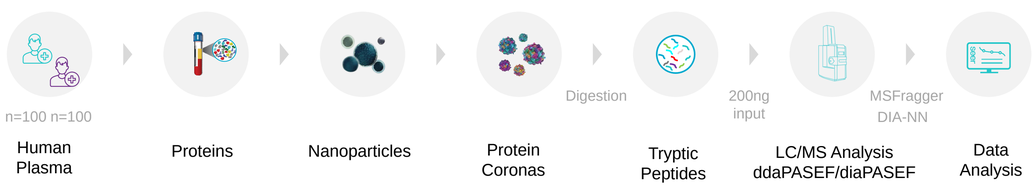
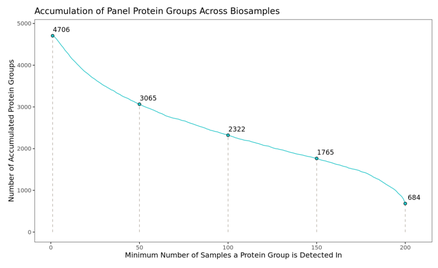


Figure 1. Workflow used to acquire cohort data using DIA and DDA analysis

In the workflow, n=100 healthy human plasma and n= 100 AD plasma samples were processed using the Proteograph workflow. Following processing, LC-MS analysis of digested peptides were performed with a microflow ddaPASEF/diaPASEF on a Bruker's timsTOF Pro mass spectrometer. DDA data was analyzed with MSFragger, and DIA data were analyzed with DIA-NN, in both cases applying 1% FDR cutoff at the protein and peptide levels.

Data Overview

A **DDA Data:**
4,706 protein groups are found across all 200 subjects and **3,065** in **25%** of all the subjects.



B **DIA Data:**
5,060 protein groups are found across all 200 subjects and **3,819** in **25%** of all the subjects.

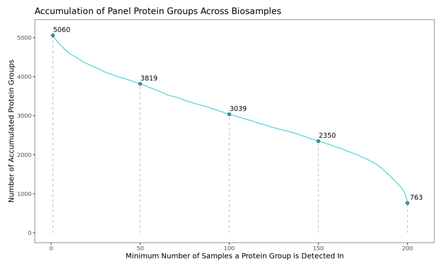


Figure 2. Protein groups identification across two LC-MS acquisition modes.

Number of protein groups detected across various fractions of the 200 initial in DDA **(A)** and DIA **(B)** mode.

Deep proteomics data enhances disease classification and biomarker characterization of AD cohort

Overlap of protein groups and peptides between two LC-MS acquisition modes

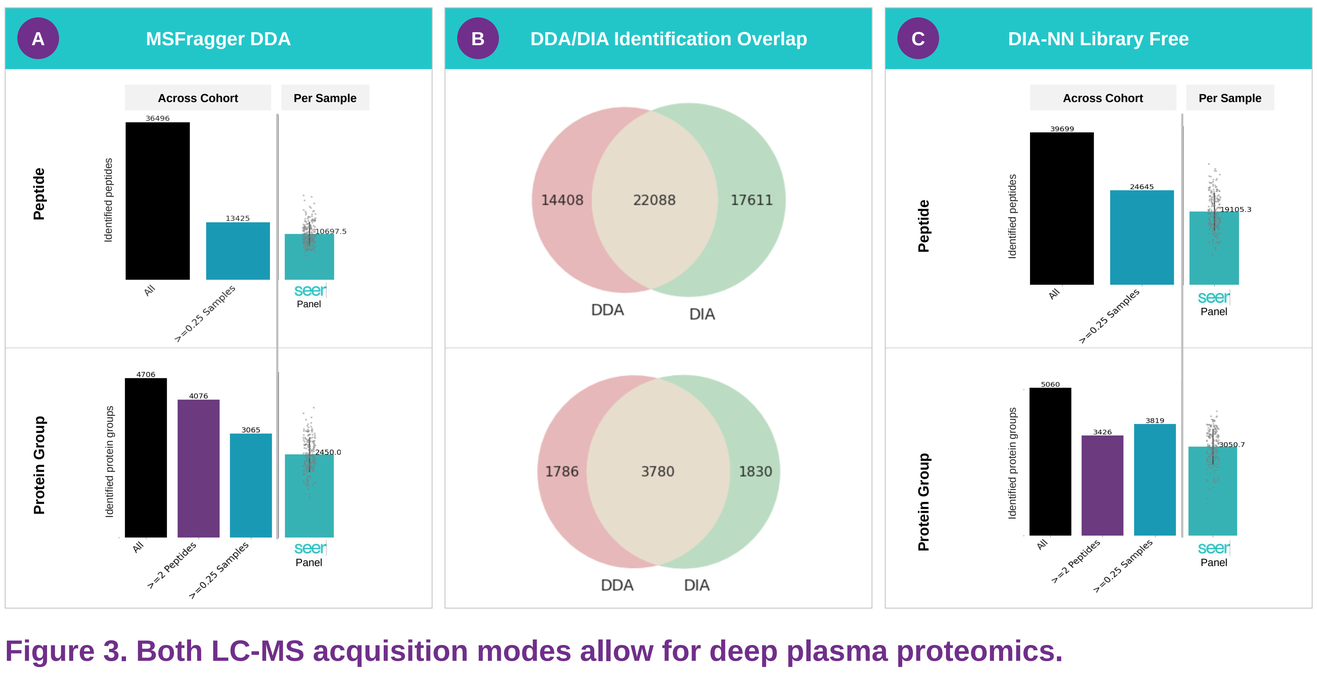


Figure 3. Both LC-MS acquisition modes allow for deep plasma proteomics.

The number of identified peptides and protein groups by **A)** MSFragger searching the DDA data and **C)** DIA-NN searching the DIA data. **(B)** The shared and unique peptide/protein identifications from both LC-MS acquisition modes. Both acquisition modes identify comparable, but unique content. DIA acquisition (2.5 hours/sample ~ 21 days/cohort) has twice the throughput as DDA acquisition (5 hours/sample ~ 42 days/cohort).

Plasma protein markers can classify Alzheimer's from healthy controls

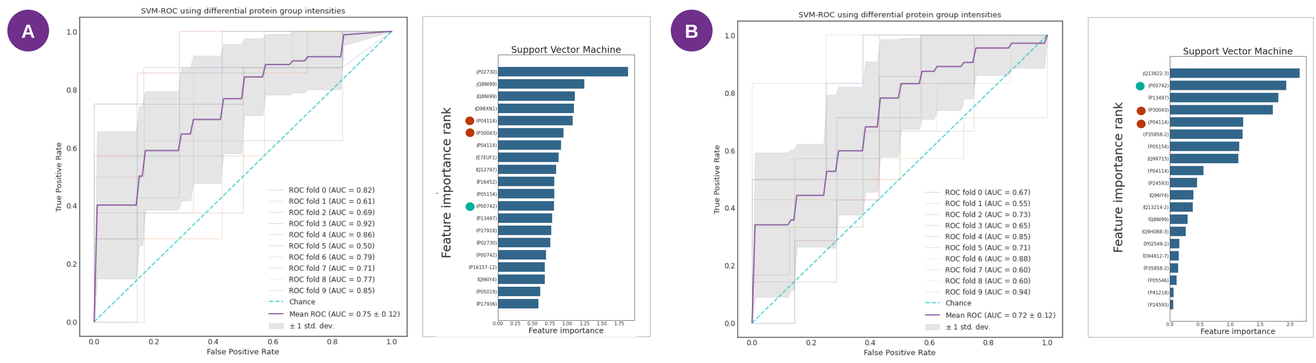


Figure 4. Classification performance for predicting Alzheimer's status with plasma markers.

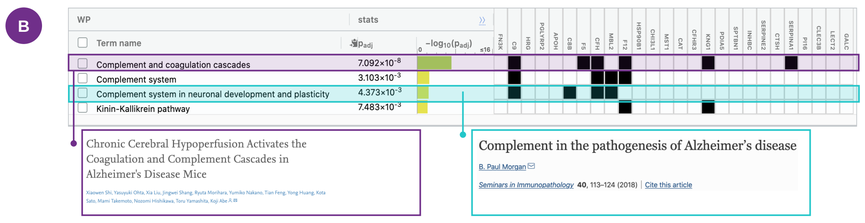
ROC curve average of 10-fold cross validation using a support-vector machines (SVM) classifier and the top 20 most important features on **A)** DDA and **B)** DIA data. Both classifiers yield similar classification performance. Proteins with red dot supported by Mueller *et al.*, 2010 and Picard *et al.*, 2021 and green dot supported by Begic *et al.*, 2020 among other new features.

Both LC-MS acquisition modes allow identification of pQTLs

A

	DDA	DIA
Possible associations (NP:PG)	21,253	22,453
Processed associations	10,612	14,092
trans-pQTLs	1,527	3,288
PGs	**294**	**397**
SNPs	1,269	2,802
cis-pQTLs	73	59
PGs	**20**	**20**
SNPs	41	38

B



Various studies to connect the cis genes to AD

APOH
Plasma Apolipoprotein Levels Are Associated with Cognitive Status and Decline in a Community Cohort of Older Individuals
Fei Song, Anne Poljak, John Crawford, Nicole A. Kochan, Wei Wen, Barbara Cameron, Ora Lux, Henry Brodaty, Karen Mather, George A. Smythe, Perminder S. Sachdev
Published: June 11, 2012 • https://doi.org/10.1371/journal.pone.0034078

SERPINA1
ANNALS of Clinical and Translational Neurology
CSF SerpinA1 in Creutzfeldt–Jakob disease and frontotemporal lobar degeneration
Samir Abu-Rumeileh, Steffen Halbgebauer, Petra Steinacker, Sarah Anderl-Straub, Barbara Polischi

CFH
CFH Variants Affect Structural and Functional Brain Changes and Genetic Risk of Alzheimer's Disease
Deng-Feng Zhang, Jin Li, Huan Wu, Yue Cui, Rui Bi, He-Jiang Zhou, Hui-Zhen Wang, Chen Zhang, Dong Wang, Alzheimer's Disease Neuroimaging Initiative (ADNI), Qing-Peng Kong, Tao Li, Yiru Fang, Tianzi Jiang & Yong-Gang Yao
Neuropsychopharmacology 41, 1034–1045 (2016) | Cite this article

Figure 5. Identified pQTLs show relevant functional enrichment for Alzheimer's.

A) Summary table of discovered cis-/trans-pQTLs across both DDA and DIA data. pQTLs were identified using BOLT-LMM[2] with FDR < 0.01 using 20 shuffled cohorts. **B)** The functional enrichment of proteins from the union of cis-pQTLs from both DDA and DIA obtained via gProfiler[3].

Conclusion

200 samples

1000 injections

5000 protein groups

~21 days LC-MS time

- The 200-sample DDA and DIA study was processed in approx. 4 weeks, generating unbiased proteomics data of over 5,000 proteins.
- The data from both acquisition schemes was equally powerful for both classification of disease state and identifying pQTLs with DIA providing 2X higher sample analysis throughput.
- Proteograph workflow is enabling deep, broad, and rapid processing of samples enabling larger and more powerful studies per unit time and resource.

References
[1] Blume *et al*. Nat. Comm. (2020)
[2] Loh *et al*. Nat Genet. (2015)
[3] Raudvere *et al*. Nucleic Acids Res. (2019)

Acknowledgements
Becky Carlyle[1], Bradley T. Hyman[1], Clemens R. Scherzer[2]
[1] Massachusetts General Hospital
[2] Brigham and Women's Hospital

PAS Proteogenomics

Publications

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

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