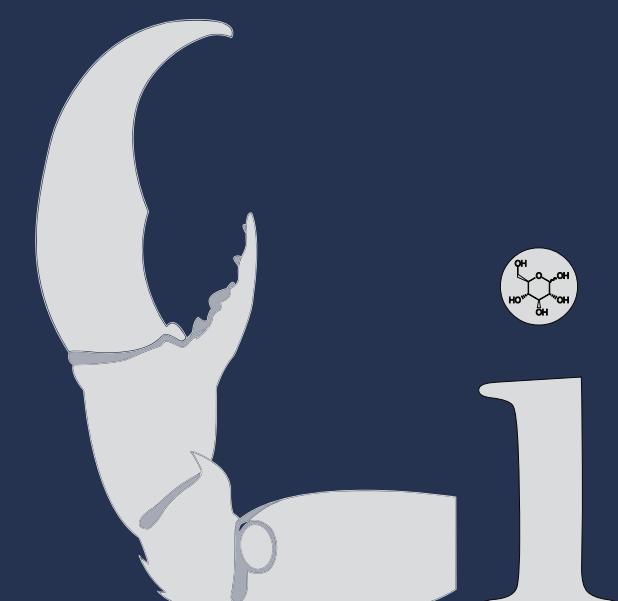


Sample Agnostic Spectral Libraries Enhance Proteomic Coverage in Sample-Limited Data Independent Analyses

Daniel G. Delafield^{1*}, Xiaofang Zhong^{2*}, Qinying Yu², Chris Sauer¹, Henrik Zetterberg³ Lingjun Li^{1,2}

¹Department of Chemistry, ²Division of Pharmaceutical Sciences, University of Wisconsin-Madison, ³Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden



Overview

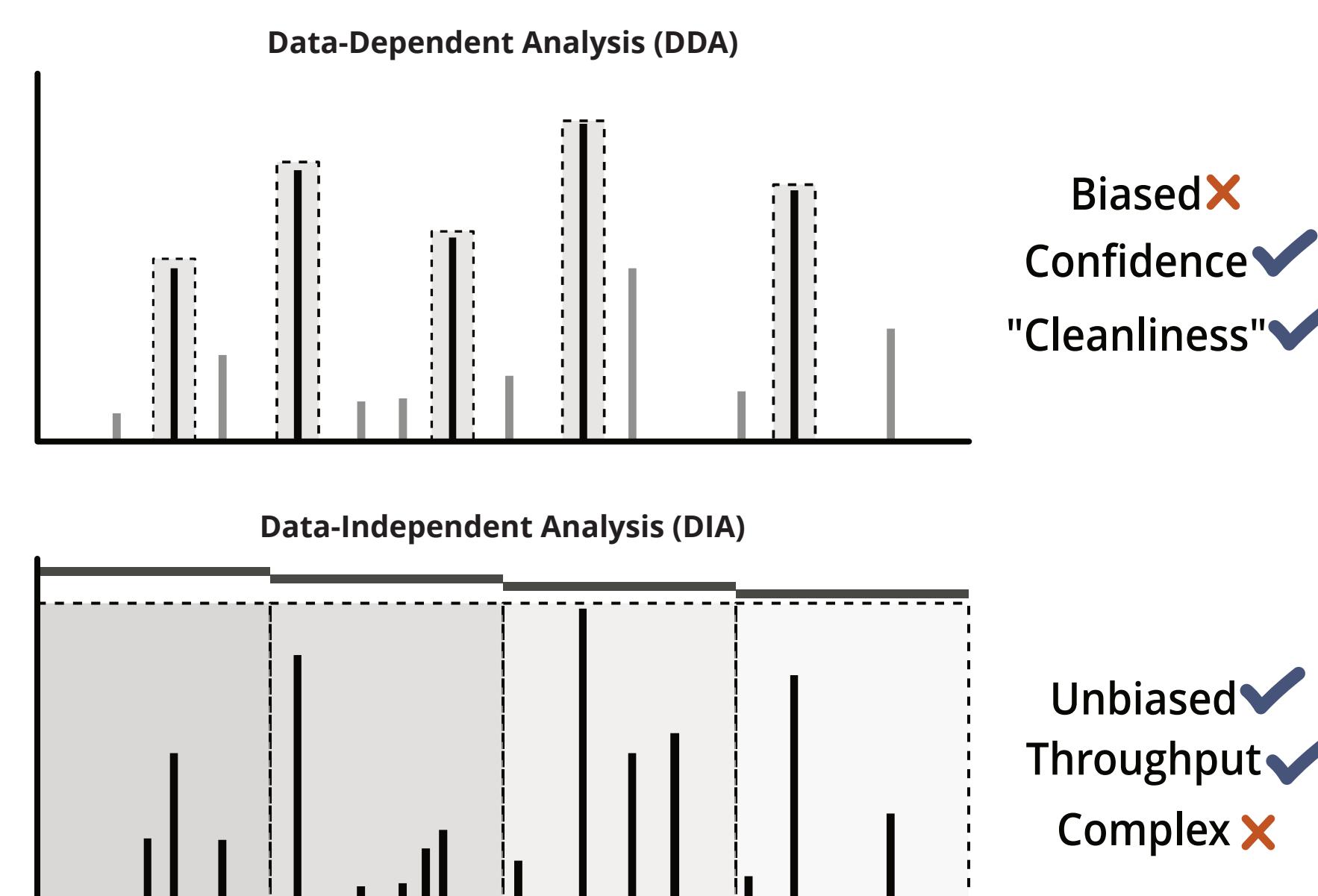
This work provides an open workflow to creating comprehensive *a priori* spectral libraries. Using extensible neural network retention time prediction, we demonstrate the ability to calibrate third-party libraries to a new experiment. In tandem with data-independent acquisition mass spectrometry, this sample agnostic library resulted in 14-fold increase in protein identifications, with 9,313 quantified and 1,642 dysregulated proteins found to be consistent with previous literature.

Introduction

Mass spectrometry is the method-of-choice for modern-day biomolecular interrogation. Typical data-dependent acquisition (DDA) experiments are facile but highly biased and irreproducible. Data-independent acquisition (DIA) promises greater profiling depth and experimental accuracy but peptide identification is traditionally reliant on the availability of user-constructed spectral libraries.

Given the ubiquity of proteomic measurement and the availability of high-quality experimental data, we posture that machine learning may be used to 'calibrate' unseen spectral libraries to a new experiment. Using this strategy, we demonstrate agnostic libraries outperform traditional methods while retaining the accuracy and completeness of evidence-based libraries. Quantifying 9,313 proteins from human cerebrospinal fluid (CSF), we identified 1,642 protein signatures known to be associated with neurological decline and Alzheimer's Disease.

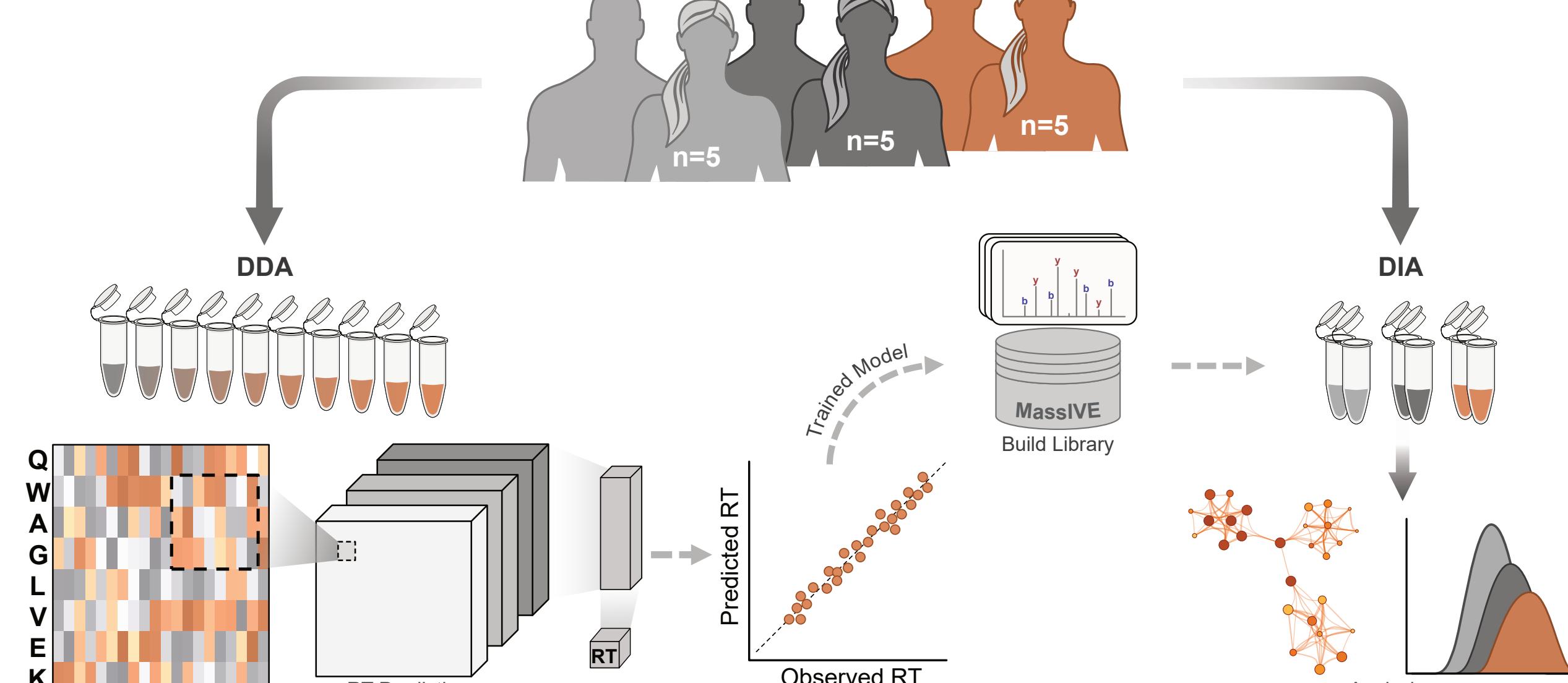
MS Acquisition Modes



DDA analysis, the more traditional approach requires little user input and provides stochastic sampling of abundant precursors within the mixture. This process results in clean, easily processed data but is biased towards highly abundant analytes.

DIA analysis works by fragmenting all precursors within a defined mass window. This improves profiling depth and removes bias but precursor assignment is complex and dependent on curated spectral libraries.

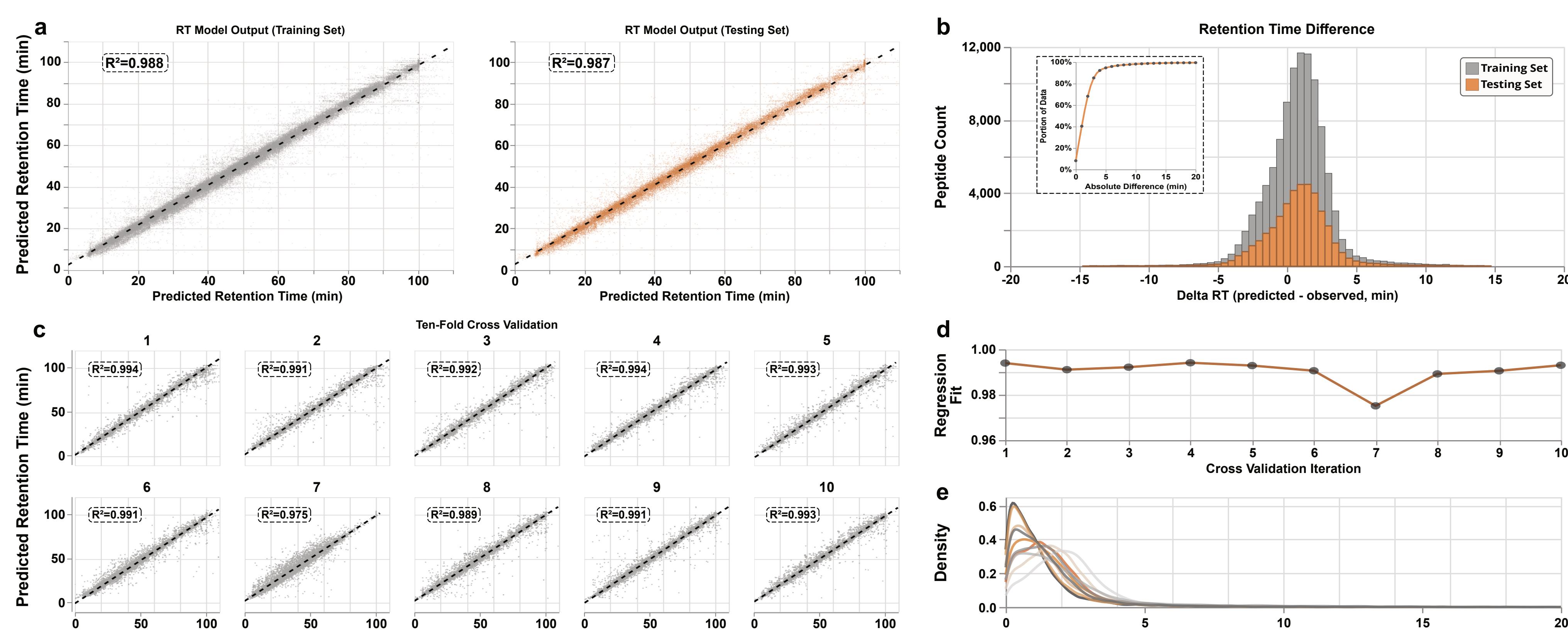
Workflow



Healthy control, mild cognitive impairment (MCI) and Alzheimer's Disease (AD) cohorts ($N=3$, $n=5$) were pooled for fractionation and analyzed through DDA-MS. Using these results to train a retention time prediction model¹, all sequences in the MassIVE spectral library were calibrated to our experiment.

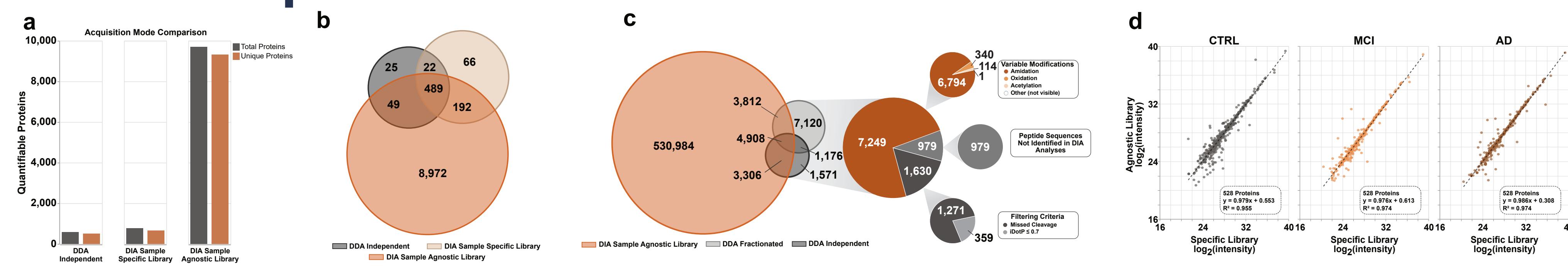
Control, MCI and AD cohorts were then analyzed separately in technical duplicate. Precursors were assigned in Skyline using the curated spectral library. Results were filtered for confidence prior to analysis.

Model Validation



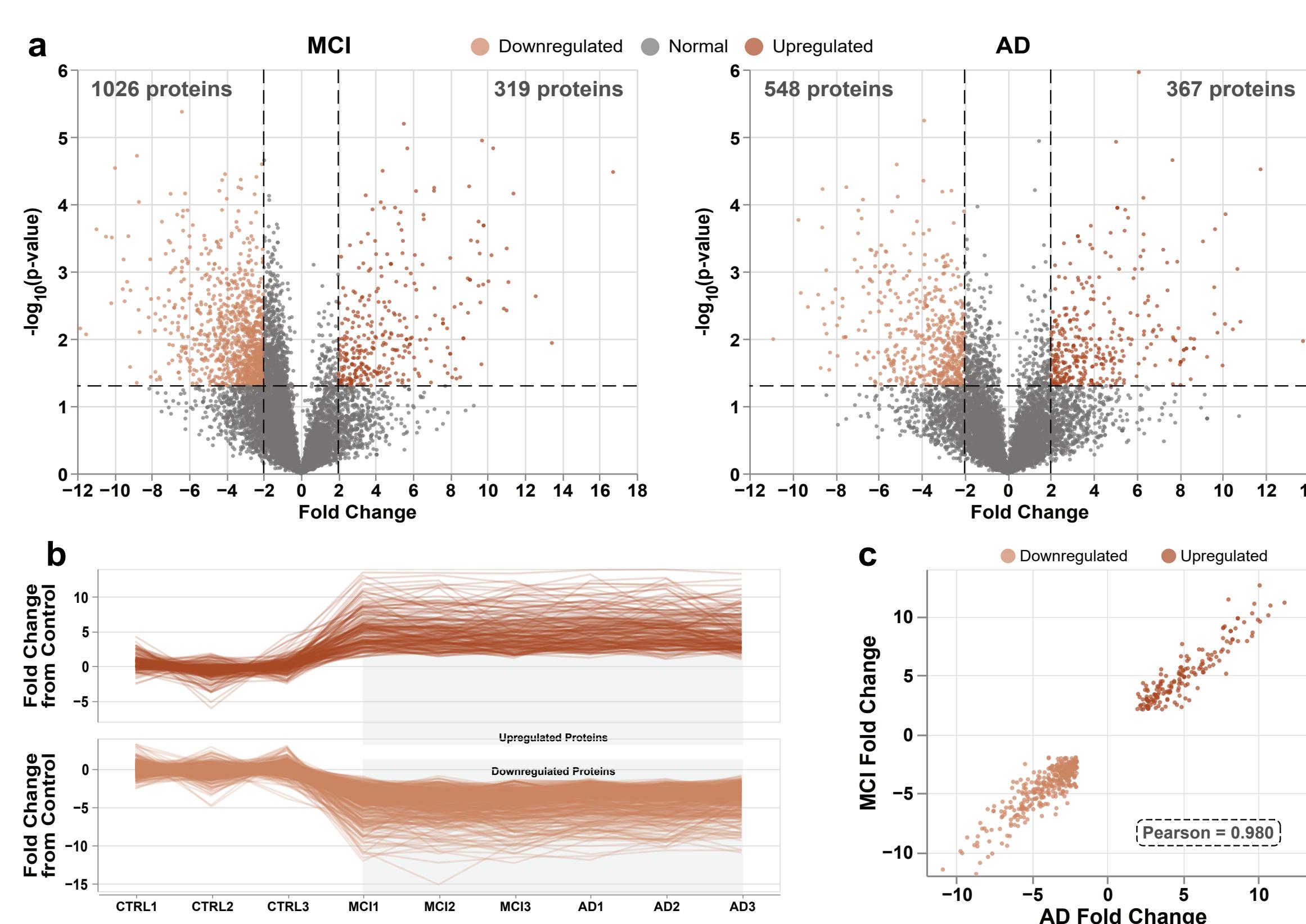
Using >160,000 precursor identifications, data were split 70:30 to train the global prediction model (a), which displayed an average difference of 0.8 minutes against empirical values. 94.7% of all predictions deviated <5 minutes (b). Ten-fold cross validation reveals low out-of-sample error with all models showing high correlation (c, d) and low variance (e).

Data Completeness



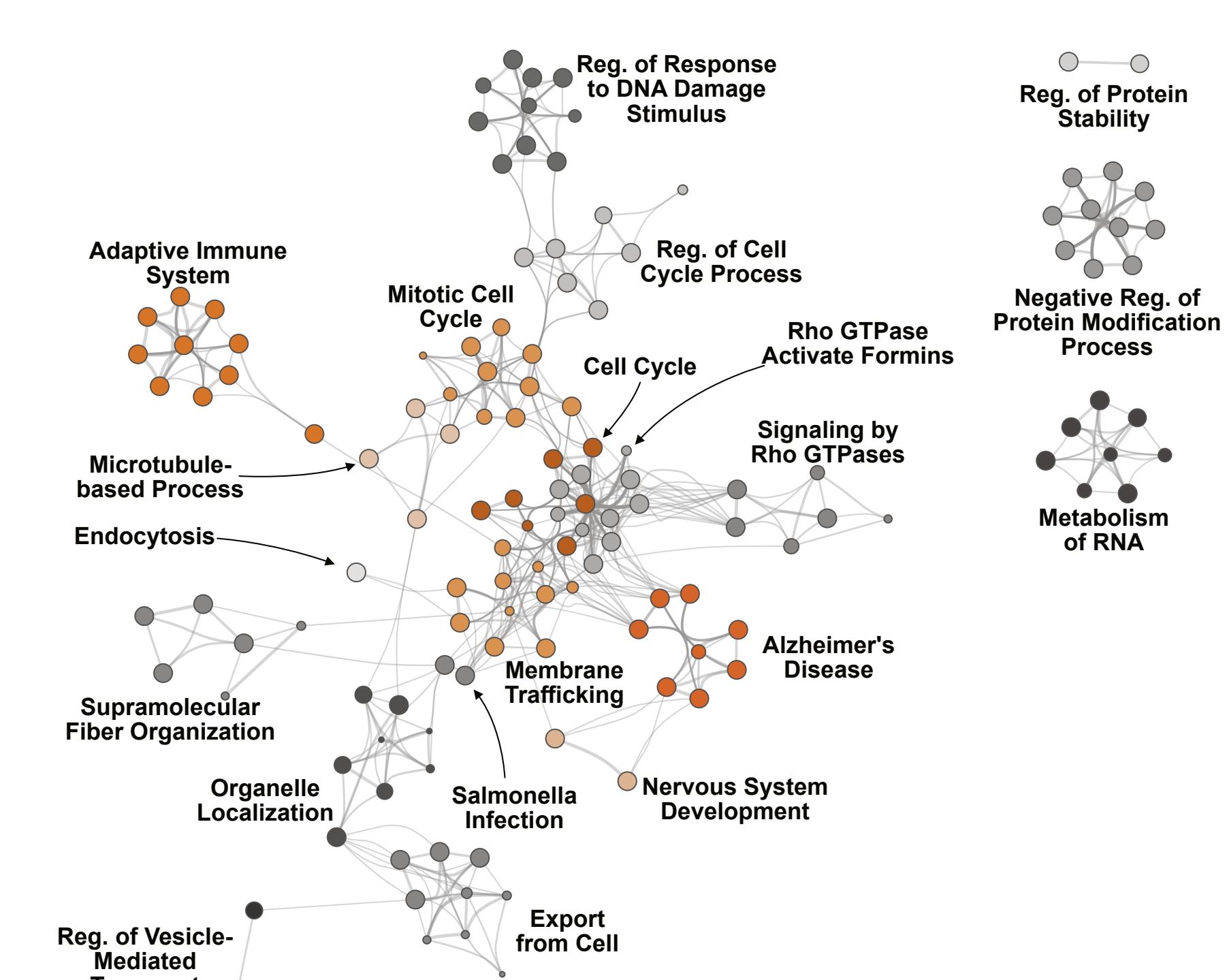
Agnostic libraries outperformed DDA and sample specific libraries ~14-fold, quantifying 9,313 protein groups (a). The majority of quantified proteins were shared among all data sets, though unique species were quantified in DDA experiments due to presence of unique peptides (b, c). Between specific and agnostic libraries, 528 proteins could be quantified on the same peptide sequences and were found to have correlational accuracy between both datasets (d) with slight discrepancy.

Disease Stratification



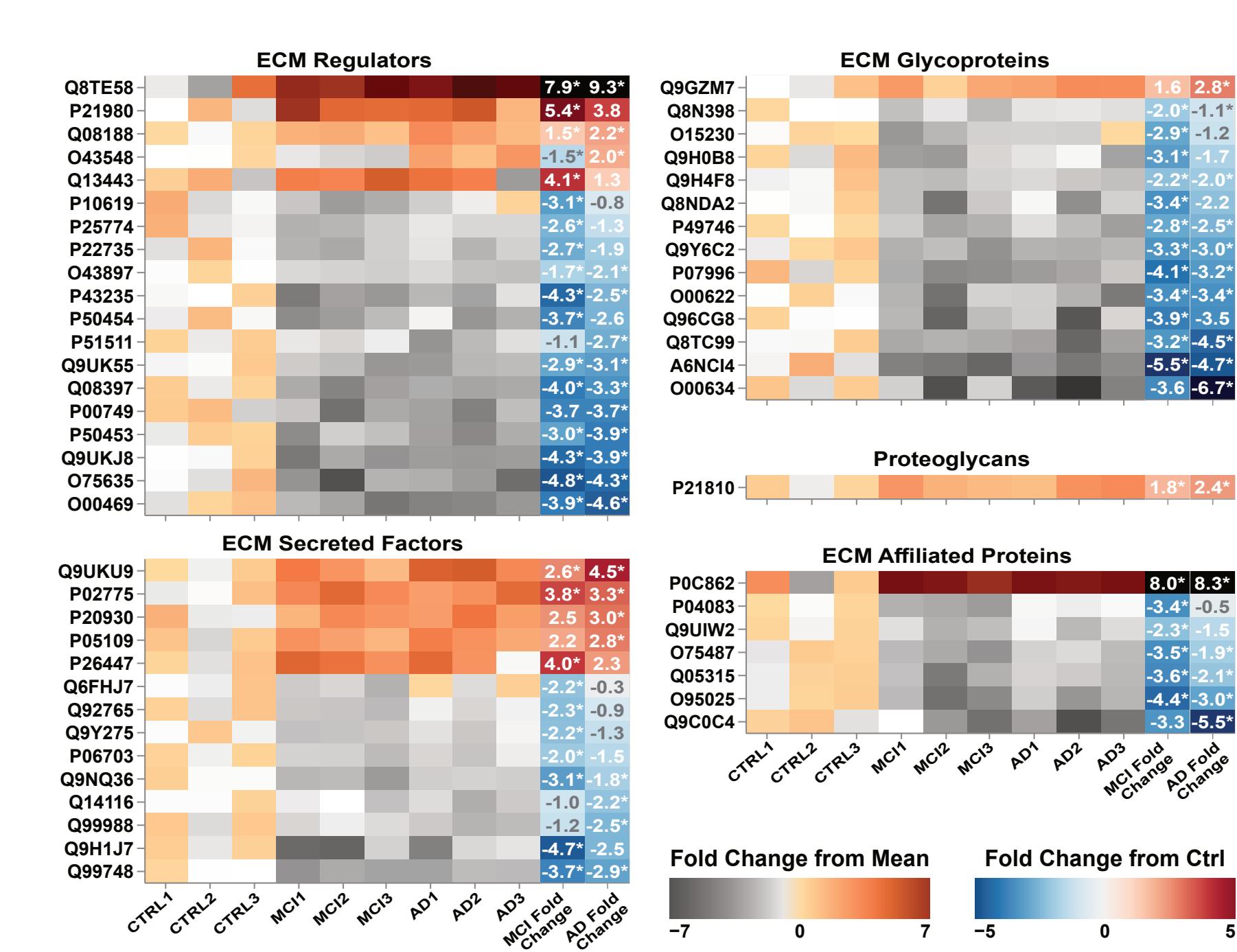
1,642 proteins were dysregulated in MCI and AD cohorts, with a significant population in MCI (a). Dysregulated proteins clustered together, showing correlation between disease groups (b, c). These protein profiles accurately stratified disease states through hierarchical clustering and PCA (d, e).

Dysregulated Pathways



Gene Ontology analysis² of downregulated proteins displayed enrichment of numerous disease relevant terms such as Alzheimer's Disease, adaptive immune system, and membrane trafficking. This enrichment further indicates the capacity to discern disease-relevant information when using sample agnostic spectral libraries.

ECM Reorganization



The extracellular matrix (ECM)³ has gained significant attention due to its known relationship to AD pathology. Our data display significant ECM reorganization in disease groups with metalloproteinases, laminin subunits, cadherins, and proteasome subunits comprising targets of interest.

Conclusion

Sample agnostic spectral libraries represent a facile approach towards deep proteomic profiling and quantitative investigations. Here we have demonstrated an open, modular framework for how these libraries may be constructed and demonstrate their utility in uncovering disease relevant information in sample-limited scenarios.

Based on these results, we advocate for greater accessibility and utilization of proteomic datasets and for greater attention on how they may be effectively employed. A community-driven approach is critical to rapid advances in human health-related research.

Contact

Graham Delafield
Ph.D. Candidate, University of Wisconsin-Madison
delafield@wisc.edu
grahamdelafield.com

Acknowledgments

This work was funded in part by the National Institutes of Health (NIH) grants RF1 AG052324, U01CA231081, and R01 DK071801. LL would like to acknowledge NIH grant support R21AG065728, NCRNS10RR025531, and S10OD025084, a Pancreas Cancer Pilot grant from the University of Wisconsin Carbone Cancer Center (233-AA19632), as well as a Vilas Distinguished Achievement Professorship and Charles Melbourne Johnson Distinguished Chair Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy.

References

- Chunwei Ma, et al. Anal. Chem. 2018
- Zhou Y, et al. Nat Commun. 2019 Apr 3;10(1):1523.
- Xinhao Shao, Isra N Taha, et al. Nucleic Acids Research, 2020.