

Sources of Variability in Quantitative LC-MS Proteomics Analysis of Human Brain Tissue Samples: Towards a Large-Scale Population Proteomics Study

Paul D. Piehowski¹, Fang Xie¹, Daniel J. Orton¹, Manuel Ramirez², Vladislav A. Petyuk¹, David G. Camp II¹, Amanda J. Myers², and Richard D. Smith¹

¹Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA; ²Division of Neuroscience, University of Miami, Miller School of Medicine, Miami, FL



Pacific Northwest
NATIONAL LABORATORY

Overview

- Pilot study of a label-free, high-throughput proteomics pipeline for analyzing human brain tissue.
- Novel study design was used to assess the variability contribution of four critical components of the technical platform.
- The platform demonstrated to be stable and suitable for high-throughput operation.
- Technical sources of variability: homogenization 76.8%, LC-MS 13.6%, proteomics processing 5.6%; the remainder attributed to temporal drift.

Introduction

- An understanding of both the inherent heterogeneity of the human population and the technical variability of the proteomics platform is needed to design a study for confident identification of disease-specific protein markers.¹
- Additionally, to obtain sufficient statistical power, it is critical that the proteomics platform is capable of analyzing large numbers of clinical samples in a high-throughput and reproducible fashion.¹
- To address these issues, we have constructed a proteomics platform that utilizes robotic sample handling, 4-column LC system, Orbitrap MS, and the AMT tag approach for the analysis of ~1000 clinical samples from human brain tissue.^{2,3}
- Pooled samples were created at different stages of the processing to isolate the variability associated with individual technical procedures and determine their contributions to the overall variability.
- The technical pipeline was divided into four categories: homogenization, proteomics processing, LC-MS analysis, and temporal variability.
- A simple protein spike-in was added to the design to evaluate this as a potential normalization strategy to control for temporal drift of the platform.

Experimental

Design

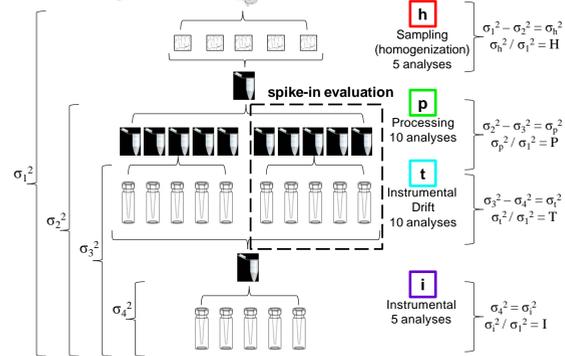


Figure 1. Experimental design for analysis of variability for technical components. The variances denoted on the left of the schematic represent the experimentally measured quantities. The right demonstrates utilization of these quantities to isolate variance for each component. This variance is then divided by the total experimental variance, σ_h^2 , to determine the contribution to the total variability.

- **Sampling (homogenization):** 5 samples were taken from the same brain region of a single patient and homogenized in individual wells to create sampling (homogenization) replicates.
- **Processing:** Aliquots from replicates were used to create a pooled sample that was distributed across 3 plates prior to proteomics processing, to create the processing replicates.
- **Temporal:** Temporal analysis replicates were created by making 5 injections from a single preparation replicate.
- **Instrumental:** Instrument replicates were created by pooling preparation replicates and making back-to-back injections on a single LC-column

Methods

- Frozen human brain tissues obtained from the University of Michigan Alzheimer's Disease Research Center's brain bank were homogenized using a Retsch Mixer Mill MM 400 at 20 Hz for 2 min.
- Proteomics processing (i.e., sample randomization, denaturation, alkylation, digestion, SPE, and normalization) was carried out using a Biomek FX (Beckman Coulter) liquid-handling robot.
- MS analysis was performed using an Exactive Orbitrap mass spectrometer (Thermo Scientific, San Jose, CA) outfitted with a custom electrospray ionization interface.
- Peptide Identification and quantification was accomplished using the accurate mass and time (AMT) tag approach.

Results

Platform stability

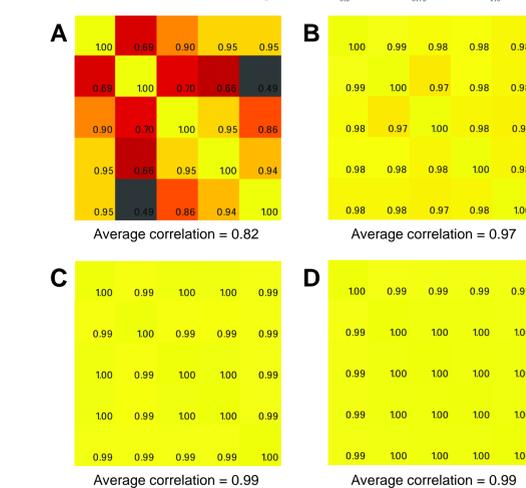


Figure 2. Pearson correlation among replicate sets for each technical component. A) Sampling (homogenization) replicates, B) proteomics processing replicates, C) temporal replicates, D) instrument replicates. Samples are sorted by run order.

Normalization to a protein spike-in

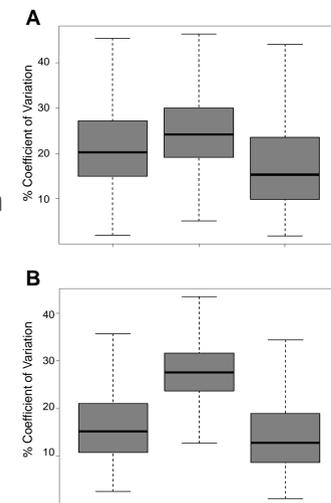


Figure 3. Comparison of spike-in normalization with central tendency normalization for A) processing replicates and B) temporal replicates. Boxplots show left: distribution of peptide coefficient of variations (CVs) using raw data; center: peptide CVs normalized using spike-in peptide intensities; right: peptide CVs after central tendency normalization.

Technical component variability

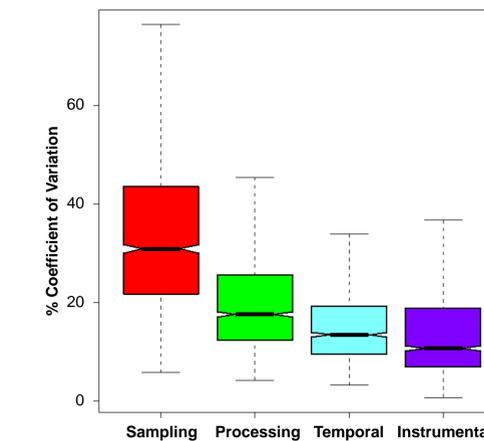


Figure 4. Distributions of peptide-level CVs after central tendency normalization for technical components.

Prospective power analysis

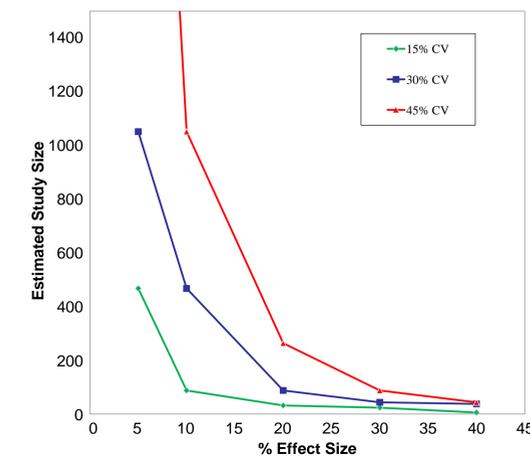


Figure 5. Estimated study size necessary to make a statistically significant measurement vs. % effect size to be detected. Sample sizes were estimated using Cohen's d and a study power of 0.90.⁴

Contribution of technical components

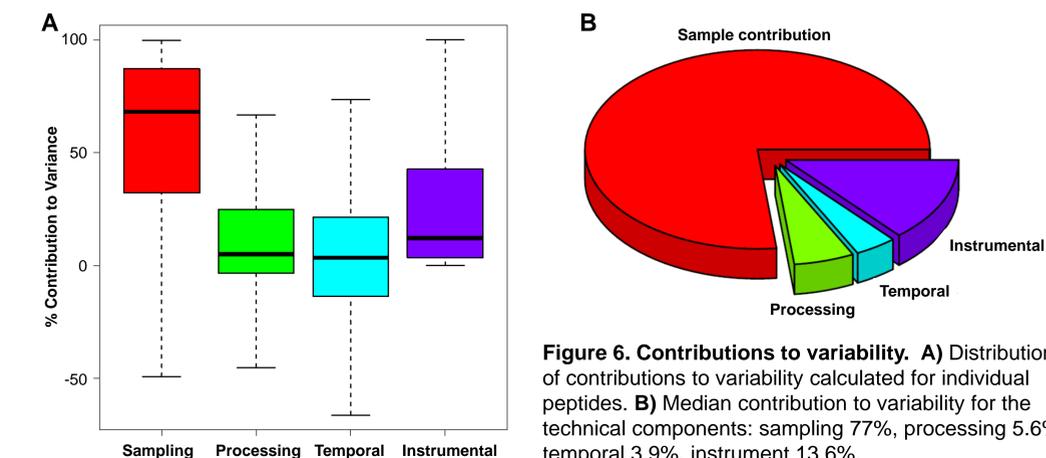


Figure 6. Contributions to variability. A) Distribution of contributions to variability calculated for individual peptides. B) Median contribution to variability for the technical components: sampling 77%, processing 5.6%, temporal 3.9%, instrument 13.6%.

Conclusions

- We have assembled a high-throughput proteomics pipeline that utilizes automated sample handling, a 4-column LC system, and the AMT tag approach for population level proteomics measurements.
- Pilot testing of our platform has demonstrated its stability and suitability for analysis of large sample sets.
- While a simple protein spike-in strategy was ineffective for normalization; however, the platform demonstrated good long term temporal stability as evidenced by strong correlation throughout the study.
- We successfully quantified the technical sources of variability in our pipeline in the context of population-scale, human-tissue studies.
- The relatively small contribution derived from the sample processing and temporal components demonstrated that automated sample handling and the 4-column LC system are well suited for high-throughput operation.
- The findings of this study are invaluable for informing study design as well as for guiding further improvements to the analytical pipeline.

Acknowledgements

This work was funded by NIH grant R01AG0345404-01 (AJM and RDS). Samples were analyzed using capabilities developed under the support of the NIH (NCRR 5P41RR018522-10 and NIGMS 8 P41 GM103493-10) and the U.S. Department of Energy Biological and Environmental Research (DOE/BER). Significant portions of the work were performed in the Environmental Molecular Science Laboratory, a DOE/BER national scientific user facility at Pacific Northwest National Laboratory (PNNL) in Richland, Washington. PNNL is operated for the DOE by Battelle under contract DE-AC05-76RLO-1830.

References

1. Qian W-J, Jacobs JM, Liu T, et al. *Mol Cell Proteomics* 5:1727-1744 (2006).
2. Zimmer JS, Monroe ME, Qian WJ, Smith RD. *Mass Spectrom Rev* 25:450-482 (2006).
3. Livesay EA, Tang KQ, Taylor BK, et al. *Anal Chem* 80:294-302 (2008).
4. *Statistical Power Analysis for the Behavioral Sciences 2nd edition*, Cohen J (Ed), Lawrence Erlbaum Associates, Hillsdale, NJ (1988).

CONTACT: Paul Piehowski, Ph.D.
Biological Sciences Division, K8-98
Pacific Northwest National Laboratory
P.O. Box 999, Richland, WA 99352
E-mail: paul.piehowski@pnnl.gov