Published in final edited form as:

J Proteome Res. 2013 May 3; 12(5): 2128–2137. doi:10.1021/pr301146m.

Sources of Technical Variability in Quantitative LC-MS Proteomics: Human Brain Tissue Sample Analysis

Paul D. Piehowski¹, Vladislav A. Petyuk¹, Daniel J. Orton¹, Fang Xie¹, Manuel Ramirez-Restrepo, Anzhelika Engel, Andrew P. Lieberman^{2,3}, Roger L. Albin^{3,4}, David G. Camp¹, Richard D. Smith¹, and Amanda J. Myers^{6,7,8}

¹Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, USA

²Department of Pathology, University of Michigan, Ann Arbor, MI, USA

³Michigan Alzheimer's Disease Research Center, Ann Arbor, MI, USA

⁴Department of Neurology, University of Michigan, Ann Arbor, MI, USA

⁵Geriatrics Research, Education, and Clinical Center, VAAHS, Ann Arbor, MI, USA

⁶Department of Psychiatry and Behavioral Sciences, University of Miami Miller School of Medicine, Miami, FL, USA

⁷Division of Neuroscience, University of Miami Miller School of Medicine, Miami, FL, USA

⁸Department of Human Genetics and Genomics, University of Miami Miller School of Medicine, Miami, FL, USA

Abstract

To design a robust quantitative proteomics study, an understanding of both the inherent heterogeneity of the biological samples being studied as well as the technical variability of the proteomics methods and platform is needed. Additionally, accurately identifying the technical steps associated with the largest variability would provide valuable information for the improvement and design of future processing pipelines. We present an experimental strategy that allows for a detailed examination of the variability of the quantitative LC-MS proteomics measurements. By replicating analyses at different stages of processing, various technical components can be estimated and their individual contribution to technical variability can be dissected. This design can be easily adapted to other quantitative proteomics pipelines. Herein, we applied this methodology to our label-free workflow for the processing of human brain tissue. For this application, the pipeline was divided into four critical components: Tissue dissection and homogenization (extraction), protein denaturation followed by trypsin digestion and SPE clean-up (digestion), short-term run-to-run instrumental response fluctuation (instrumental variance), and long-term drift of the quantitative response of the LC-MS/MS platform over the 2 week period of continuous analysis (instrumental stability). From this analysis, we found the following contributions to variability: extraction (72%) >> instrumental variance (16%) > instrumental stability (8.4%) > digestion (3.1%). Furthermore, the stability of the platform and its' suitability for discovery proteomics studies is demonstrated.

Keywords

Label-free quantification; technical variation; sample preparation; reproducibility; study design; tissue analysis

INTRODUCTION

MS-based proteomic technologies have become indispensable for the interrogation of the proteome.^{1–3} Broad discovery-based quantitative proteomics measurements, also known as “global” proteomics, is a powerful implementation whereby a broad survey of the proteome is conducted across multiple sample conditions with the purpose of identifying species with differential abundance.⁴ This analysis conventionally requires a number of processing steps including homogenization/extraction of proteins, denaturation, reduction of disulfide bonds, alkylation of cysteine residues, enzymatic digestion, LC separation and analysis by MS.^{5,6} This is followed by extensive informatics processing to identify and quantify peptide and protein abundances. Due to the complexity of these analyses, a thorough understanding of the pitfalls and potential sources of variability in the technical platform is critical to ensure adequate study design and identify bottlenecks in the pipeline to guide their further improvement.^{7–10}

For this analysis, the above described analytical pipeline was divided into 4 parts: extraction, digestion, instrumental variance, and instrumental stability. Extraction refers to the biological handling of the sample, and for this study consists of the dissection and homogenization of the tissue. The second step (digestion) is converting the extracted proteins into peptides and getting them in a buffer that is mass spectrometry compatible. For this study, we minimized the impact of “intermittent” human errors by carrying out all digestion steps using an automated liquid handler. When making replicate injections for LC-MS/MS analysis, there is some variation in the instrument response from run-to-run. This fluctuation is described by the instrumental variance component. Lastly, analyses can take place over the course of many days and even weeks. During this time, many small changes may take place in the LC-MS/MS platform: different batches of buffer, decay in performance of columns, electrospray tips, inlet capillaries and instrument components, drifts in calibration and tuning, etc. These changes can contribute to the observed technical error and are captured by the instrumental stability component.

Fluctuation in proteomic platform quantitative sensitivity and response is widely appreciated.⁸ As a result, approaches based on stable isotope labeling have been developed to specifically mitigate the effects of instrumental fluctuations. Examples of stable isotope labeling for quantification approaches include: enzymatic ¹⁸O-labeling,^{11–14} metabolic labeling (SILAC),^{15,16} isobaric mass tagging (e.g. iTRAQ),^{17–19} and mass-difference tagging (ICAT).^{20–23} While these methods have proven very useful for increasing measurement precision, they have drawbacks for discovery measurements. Specifically: additional reaction and cleanup steps utilizing expensive reagents, reduced limit of quantification (LOQ), and reduced peptide identifications.^{24–26} With the exception of metabolic labeling, these labeling protocols are aimed at minimizing the impacts of instrumental variances, but do not encompass the processing steps that occur prior to the addition of label. Therefore, labeling cannot control for all of the technical variability present in such measurements.

Here we present an experimental design aimed at detailed assessment of the technical variability present in a proteomics study. Additional attention was paid to including variability that would be present in a large-scale study: samples being processed across

multiple plates and days, effect of well position, utilization of multiple LC columns, and instrumental stability. This was done to provide an as close as possible characterization of the variability present in a real-world biomarker discovery study. Multiple replicates were analyzed at each stage of the pipeline making it possible to isolate the variability associated with individual technical procedures and determine their contributions to the overall variability. The objective of this investigation is to provide a more thorough understanding of the technical variability present in a quantitative proteomics experiment in order to better inform study design and guide future pipeline developments.

MATERIALS AND METHODS

Human brain tissue samples

Frozen human brain tissues were obtained from the University of Michigan Alzheimer's Disease Research Center's brain bank. Our criteria for inclusion were as follows: self-defined ethnicity of European descent, neuropathologically confirmed late onset Alzheimer's disease or no neuropathology present, male gender and age of death greater than 65. Additionally, the genotypes of all samples in the cohort were analyzed via the program STRUCTURE^{27,28} as described in^{29,30} and ethnic outliers (non-Caucasian) were removed from the study. Neuropathological diagnosis was defined by board-certified neuropathologists as per standard National Alzheimer's Coordinating Center protocols. We selected two samples for evaluation of our pipeline. The control patient was age 74 and post mortem interval (PMI) = 24, and the affected patient was age 76 and PMI = 4.5. Samples were de-identified before receipt, and the study met local human studies institutional review board and HIPPA regulations. This work is declared not human-subjects research and is IRB exempt under regulation 45 CFR 46.

Proteomic sample processing

Five samples each were collected from 2 individuals, resulting in ten tissue sections. Approximately 80 mg of tissue per section was dissected from the dorsolateral prefrontal cortex. The tissue was placed in a 2 mL 96-deepwell plate and 1mL of homogenization buffer (8M Urea, 10 mM DTT in 50 mM Tris-HCl) was added. Homogenization was performed using a Retsch Mixer Mill MM 400 at 20 Hz for 2 minutes. 100 μ L aliquots were taken from homogenized samples and filtered using 1.2 μ m low-protein binding filter plates (Pall Life Sciences) to remove particulates. Total protein concentration was determined by coomassie assay. Protein concentrations were then normalized and a subsequent 100 μ L aliquot was taken to provide 150 μ g of starting material per sample. Apomyoglobin (Sigma Aldrich) was spiked in at 1% by mass, to appropriate samples prior to the denaturation step. Samples were then incubated 1 hr at 37 °C for denaturation/reduction. Protein cysteinyl residues were alkylated with 40mM iodoacetamide (Sigma Aldrich) for 1 hr at 37 °C in the dark. Samples were diluted 4-fold with 50 mM NH_4HCO_3 buffer prior to digestion using mass spectrometry grade trypsin (Promega) with a ratio of 1:50 (w/w). Tryptic digests were desalted using C₁₈ SPEC tips (Varian). Peptides were eluted in 200 μ l 80 % ACN/ 0.1 % TFA and lyophilized. Desalted peptide concentrations were determined by BCA assay and adjusted to 0.5 mg/ml prior to storage at -80 °C until LC-MS analysis. Sample randomization, denaturation, alkylation, tryptic digestion, SPE, and normalization, collectively referred to as digestion, was carried out using a Biomek FX (Beckman Coulter) liquid-handling robot.

Instrumental analysis

To maximize LC-MS throughput, an in-house built 4-column LC system was employed, resulting in nearly 100% duty cycle³¹ and allowing analysis of up to 24 samples per day when employing a 60 min separation gradient. The LC system was custom built using two

Agilent 1200 nanoflow pumps and one 1200 capillary pump (Agilent Technologies, Santa Clara, CA), various Valco valves (Valco Instruments Co., Houston, TX), and a PAL autosampler (Leap Technologies, Carrboro, NC). Full automation was made possible by custom software that allows for parallel event coordination and therefore near 100% MS duty cycle through use of four analytical columns (Supplemental Figure 1). Reversed-phase columns were prepared in-house by slurry packing 3- μ m Jupiter C₁₈ (Phenomenex, Torrance, CA) into 35-cm $\times$ 360 μ m o.d. $\times$ 75 μ m i.d fused silica (Polymicro Technologies Inc., Phoenix, AZ) using a 1-cm sol-gel frit for media retention.³² Mobile phases consisted of 0.1% formic acid in water (A) and 0.1% formic acid in 100% acetonitrile (B) with a gradient profile as follows (min:%B/event); 0:0, 1.2:8, 12:12, 45:35, 58:60, 60:85. Sample injection occurred 20 min prior to beginning the gradient while data acquisition lagged the gradient start and end times by 10 min to account for column dead volume that allowed for the tightest overlap possible in multi-column operation. Multi-column operation also allowed for columns to be 'washed' (shortened gradients) and regenerated off-line without any cost to duty cycle.

MS analysis was performed using an Exactive Orbitrap mass spectrometer (Thermo Scientific, San Jose, CA) outfitted with a custom electrospray ionization (ESI) interface. Electrospray emitters were custom made by chemically etching 150 μ m o.d. $\times$ 20 μ m i.d. fused silica.³³ The heated capillary temperature and spray voltage were 250°C and 2.2 kV, respectively. MS spectra (AGC 1×10^6) were collected from 400–2000 m/z at a resolution of 50k. Instrument cleaning and any necessary maintenance is performed at 24 hr intervals to help control instrument drift.

Data analysis

Identification and quantification of peptides was performed using the accurate mass and time (AMT) tag approach.³⁴ The AMT tag database was populated using a pooled sample that was fractionated offline by high-pH reversed-phase LC (bRPLC) prior to LC/MS/MS analysis on an LTQ-Orbitrap (Thermo Scientific).³⁵ MS/MS data was searched by MS-GFDB.³⁶ Confident identifications were compiled along with the observed elution times to generate the AMT tag database. For individual study samples, Decon2LS was used for peak-picking and for determining isotopic distributions and charge states.³⁷ Deisotoped spectral information was loaded into VIPER to find and match LC-MS features (same monoisotopic mass present in a number of consecutive MS scans) to the peptide identifications in the AMT tag database.³⁸ VIPER provided an intensity report for all detected features, normalized LC elution times via alignment to the database, and feature identifications. The area under the curve (AUC) from extracted ion chromatograms was used as the measure of peptide abundance. The resultant peptide identifications were filtered using an UP score of 0.5 to remove ambiguous identifications of a single LC-MS feature, and an FDR of < 15%.³⁹ The false discovery was further limited by requiring that an LC-MS feature appear in all datasets and have a mass error < 6 ppm in each of them. This resulted in a list of 2083 peptides, which mapped to 778 proteins, to be used for downstream analysis. Peptide abundance crosstabs were exported to R for statistical analysis and plotting. The coefficient of variation (CV), which is calculated by dividing the standard deviation of peptide profiles by the mean, is reported as percentages. Global scaling normalization coefficients were calculated as the ratio of peptide abundance to the median peptide abundance measured across all samples.⁴⁰ All calculations of variance were carried out using peptide-level abundance measurements. This was done because peptides are the chemical entities being measured in a bottom-up proteomics experiment. Thus, it was decided that peptide abundance variability would best reflect variations in the performance of our platform. Furthermore, protein inference^{41,42} and peptide rollup to gene based profiles^{43,44} are active areas of research and prone to uncertainties which could potentially impact our results in

unanticipated ways. All .raw data files as well as tabulated AUC abundances from peak matching are available for download at omics.pnl.gov.

RESULTS

Modeling the Variability in Quantitative Proteomics Data

In a proteomics experiment, the measured peptide intensity, A , can be expressed as a multiplicative relationship:

$$A = A_0 \times E \times D \times T \times I \quad (\text{eq. 1})$$

Where A_0 is the mean intensity of the peptide, E , D , T , and I are the yields of multiplication factors associated with extraction and digestion, instrumental stability, and instrumental variance, respectively. Assuming yields and multiplicative factors follow a log-normal distribution with mean equal to 1, then after log-transform eq. 1 can be expressed as:

$$\log(A) = \log(A_0) + \mathcal{N}(0, \sigma_E^2) + \mathcal{N}(0, \sigma_D^2) + \mathcal{N}(0, \sigma_T^2) + \mathcal{N}(0, \sigma_I^2) \quad \log(A) = \log(A_0) + \mathcal{N}(0, \sigma_E^2 + \sigma_D^2 + \sigma_T^2 + \sigma_I^2) \quad (\text{eq. 2})$$

Thus, the log of the peptide intensity measurement can be expressed as equation 2, where errors following a normal distribution with composite variance and contributions from multiple steps in the pipeline (E , D , T and I). The latter form of equation assumes normal distribution of measurement error after log-transform of intensities is widely used in statistical analysis of LC-MS proteomics data.^{45,46} The mapping of this model to the experimental design allows us to quantitatively estimate the contribution of the individual steps in the pipeline into the composite variance as outlined in Figure 1.

Experimental design

A successful quantitative proteomics analysis depends on the accurate execution of a number of diverse technical processes. In this work, we chose to break the pipeline into 4 critical components that are commonly believed to contribute appreciably to variability: extraction, digestion, instrumental stability, and instrumental variance. A schematic of the design is shown in Figure 1.

The extraction component is composed of variability that derives from the precise location of tissue dissection as well as homogenization. To create extraction replicates, 5 samples were dissected from the same brain region of a single patient and homogenized in individual wells. Aliquots from extraction replicates were used to create a pooled sample that was distributed across the three 96-well plates prior to processing, to create the digestion replicates. Consequently, the digestion component contains variability associated with denaturation, alkylation, trypsin digest and SPE. Instrumental stability replicates were created by making 5 injections from a single preparation replicate spread throughout the 2 week study, therefore containing variability associated with LC column degradation and instrumental performance drift. To measure the variability in instrumental response, instrumental variance replicates were created by pooling preparation replicates and making back-to-back injections on a single LC-column. An identical design was utilized for both individuals.

The experimental design was constructed to closely reflect true operating conditions to more effectively capture variances that are present in a real-world, moderate-scale study. To achieve this end, samples were spread across three different 96-well plates, and placed in randomized well positions, to account for variance associated with plate-to-plate and well position differences. The plates were processed over the course of 1 week, to include

variability associated with batch processing. The analysis run order was designed in such a way that association between confounding factors, such as plate, disease state and LC column, was disrupted by corresponding randomizations. To achieve this, the samples were divided into 5 blocks. Each block contained only one sample from each combination of group (extraction, digestion, digestion + spike-in, instrumental stability, and instrumental stability + spike-in) and condition (control and disease), resulting in ten samples per block. Within each block there were 2 additional considerations: at least one sample from each group must be run on each LC column, and control and disease samples were paired to ensure even distribution of disease state analyses. The 5 replicate injections that comprise the instrumental variance group were completed at the end of the blocked study.

Normalization using a single protein spike-in

Instrumental response drift is a continual concern in label-free proteomics measurements. This is especially true when the analysis is carried out over a long period of time. We evaluated the utility of adding a single protein spike-in as an internal standard to control for global fluctuations in instrument response. For the spike-in experiments, 5 aliquots were taken from the pooled sample before digestion and spiked at 1% (m/m) with purified myoglobin. Because the spike-in was added prior to digestion, we were able to assess its effectiveness at the digestion and instrumental stability level. The results are summarized in Figure 2, and are compared to variability with no normalization, and a global scaling normalization where normalization coefficients were derived based on tentative identifications of the tryptic peptides originally present in the sample.⁴⁷ For both components the global scaling method outperformed the spike-in, in fact, using the spike-ins for normalization actually increased the CV. We postulate that this is because the small number of peptides observed from myoglobin is not a large enough population to be representative of the fluctuations that occur in a proteomics experiment. Only 11 peptides derived from the myoglobin spike-in were seen across all samples to derive coefficients for normalization. Normalization coefficients derived from spike-ins show poor correlation to those from the global scaling approach. Additionally, adding the spike-in resulted in a small decrease in the number of identified peptides (data not shown). It was decided to use the global scaling method as it provides considerable improvement in CV without significantly altering the data, which would complicate interpretation of downstream analysis.

Correlation between replicates for technical components

Because replicate samples were processed in separate batches and analyzed over the course of two weeks, it is important to ensure that there is good agreement between the datasets. To demonstrate this, we looked at the Pearson correlation within each individual technical component. For extraction, digestion, and instrumental stability replicates all ten LC-MS runs are used, five replicates from each patient. Biological difference between the two patients was removed by subtracting patient-specific mean abundances from the corresponding datasets. This was done to include any differences in variance between phenotypes. Additionally, this effectively doubled the number of replicates at each level and allowed to achieve more accurate estimates of variances.

When all technical variability is included, extraction replicates, the average correlation is 0.86. As expected, when components of variability are removed, the average correlation increases considerably. This phenomenon was similarly demonstrated by Qian et. al.⁷ Smaller increases in correlation are observed between digestion, instrumental stability, and instrumental variance replicates, indicating very good stability of these components throughout the study. As mentioned above, instrumental response drift is a common source of systematic error in quantitative proteomics. Obtaining an average correlation of 0.98 for instrumental stability replicates run across the entire 2 week analysis period suggests that

instrument “drift” is well controlled in our platform. This level of agreement is crucial to accurate assessment of variability contributions from upstream processes.

Another metric for comparing agreement between sample replicates, used commonly in the literature, is coefficient of variation (CV). The distribution of peptide CV's (calculated on non-log transformed intensities) after normalization for the four technical components are shown in Figure 4. In agreement with Figure 3, we see a decrease in variability as technical components are removed. This result gives us confidence in using this data to calculate individual contributions from technical components.

Contribution of technical components to variability

Using the model described above (eq. 2) we can calculate the contributions of the individual technical components. Figure 5A shows a boxplot of variance contributions for individual peptide measurements. The wide distribution that was obtained for each component clearly indicates that different peptides are affected differently by each processing step. This is to be expected considering the wide range of physical properties of proteins and their resultant peptides. It is also interesting to note that for some peptides, nearly 100% of their variability is contributed by LC-MS analysis. Figure 5B shows the median contribution to variability of each of the measured steps in the pipeline. From these results we can derive the following inequality: extraction >> instrumental variance > instrumental stability > digestion. Thus, for our proteomics platform controlling the variability of the protein extraction (dissection and homogenization) is most critical.

Figure 5: Contributions to variability. A.) Boxplot showing the distribution of contributions to variability calculated for individual peptides. B.) Piechart demonstrating the median contribution to variability for each of the four technical components: extraction 72%, digestion 3.1 %, instrumental stability 8.4%, instrumental variance 16%.

DISCUSSION

Comparison of variability with literature

By design, the measured CV for extraction represents the total variability observed in the study. Thus, we observe a median CV of 34% for peptide measurements which includes all technical variability associated with our high-throughput platform. This result compares well with the expected intrapersonal variability in other common sample types i.e. biofluids.^{49,50} In a study of urine, Nagaraj et. al.¹⁰ found an intrapersonal variability of 48%, depending on the day that it was collected. While this number may seem prohibitively large for meaningful discovery studies, it was found to be significantly smaller than the interpersonal variability measured (66%). Thus, it was deemed sufficient.

Technical variability depends greatly upon sample type and methodology. Gan et. al.⁵¹ report a technical replicate CV of 11% in a study of bacterial cell cultures, where cell type is uniform and homogenization is more readily achieved. This number also increases significantly when biological variability is considered. The CV obtained for sampling from two different cultures grown under the same conditions was found to be 25%. These results were obtained using an iTRAQ labeling approach and demonstrate that even when employing sophisticated labeling approaches, controlling upstream technical components is critical.

Instrumental variance replicates produced a median CV of 13%, and instrumental stability replicates a median 17% (Figure 4). This small increase CV obtained when spreading the analysis out over time, demonstrates control over instrument drift. These numbers compare quite favorably with numbers reported for other quantification studies that utilize AUC

measurements for label-free quantification: median CV of 18% in urine,¹⁰ and a CV range of 20%–50% in plasma.¹⁴ This is not a direct comparison however, as different approaches were utilized for peptide-to-spectrum match (PSM) filtering and normalization. In our study, only peptides identified in all datasets were used. This enriches for higher abundance peptide identifications and thus gives better results for reproducibility at the expense of coverage. Numerous reports using stable isotope labels such as SILAC¹⁵, super-SILAC⁵², and iTRAQ⁵¹, report instrumental variance CV's less than those obtained in this study. In the case of super-SILAC, median CV's of 5% have been reported. This suggests that although much of the variability in our platform occurs before instrumental analysis, reproducibility improvement could still be obtained by incorporating an appropriate labeling strategy.

Contributions of technical components

From the pie chart in Figure 5B, it is clear that the largest contributor to variance is the extraction step. This is not surprising, as this component includes variability that derives from different compositions of cell types within each extraction replicate, varying levels of blood contamination, and varying protein extraction efficiencies caused by well-to-well variations in homogenization efficiency. Reports of differences in protein expression caused by differing compositions of cell types are prevalent in the literature.^{53–55} For example, a 71% difference in MAO-B expression between grey and white matter has been reported.⁵⁵ However, to the best of our knowledge a comprehensive study of the impact of cell composition on global proteomics measurements is absent from the literature. Laser capture microdissection (LCM) represents a powerful potential tool for improvement in cell selection, and has been successfully used to isolate specific cell types in the brain and other tissue types for proteomics analysis.^{56–58} The downside of LCM is significantly decreased throughput.

In the case of tissue, homogenization has long been recognized to contribute variability to proteomics investigations.⁵⁹ As a result, numerous homogenization procedures have been investigated that utilize various detergent-chaotrope compositions.^{60,61} However, many of these reagents can pose challenges for downstream processing and LC-MS analysis. Our extraction buffer contained an 8M concentration of urea to aid in extraction, which was readily removed by C18 solid phase extraction. Detergents can be more difficult to remove but would likely aid in the protein extraction process. Recently, a number of protocols have been published to address this challenge, but were not investigated in this study.^{62,63} Butt et al. demonstrated between 36% and 260% improvement in extraction efficiency, depending on tissue type, using an automated frozen disruption approach compared with a manual approach. They also saw a small increase in number of proteins resolved and noted more uniformly and thoroughly homogenized samples.⁵⁹ Assessment of variability, analogous to extraction replicates in this study, was not performed.

To investigate the impact of blood contamination on our analysis, we looked at the extraction replicates and found 12 unique peptides that unambiguously match to hemoglobin. All 12 peptides were in the top quartile of abundance and have a median CV of 46% across extraction replicates. This is much larger than the 26% CV for the remaining peptides in the top quartile of abundance for this study. This suggests that there is indeed varying levels of blood contamination across our extraction replicates. Although these 12 peptides result in only a small increase in the reported median CV and can be easily identified and removed, the variable presence of these highly abundant peptides in the sample will still contribute to variability in the LC-MS analysis through inconsistent ionization suppression of co-eluting peptides. Controlling the cell type, sample composition and homogenization represent an area of great potential for future platform improvements.

The second largest contribution, 16%, comes from the LC-MS analysis variability. The CV values obtained for instrumental variance measurements agree well with others reported in the literature, *vide supra*, giving us confidence in the LC-MS performance in this study. This finding has significant implications for the field of quantitative proteomics. As mentioned, there has been significant investment in developing methods for improving measurement precision. These developments have been largely focused on instrument variability. This is especially true in the case of targeted proteomics measurements that have become somewhat of a “gold standard” for quantitation.^{64–68} However, even in the case of an ideal mass spectrometer (0% variability) significant variability would still exist for most sample types.

The remaining 11% of the variability comes from digestion and instrumental stability components. None of the studies mentioned above have decoupled the extraction and homogenization components from digestion and instrumental stability. Due to the detailed design of this study, we were able to demonstrate that the variability associated with our automated sample preparation makes only a small contribution (3%) to the overall variability. Moreover, instrument drift contributed less than 10%.

Study design and statistical power

Having an accurate estimate of the total variability of the analytical platform is essential to inform accurate study design. The impact of technical variability on necessary study size is shown in Figure 6. The estimated number of patients needed to obtain statistical significance increases drastically across the range of peptide CV's measured in this study, and quickly becomes prohibitively large for standard proteomic platforms. Large study sizes (>100) present a number of challenges for researchers, such as: long term platform stability, plate-to-plate differences, cost, and sample availability. It is also important to note that these calculations do not include biological variability within study groups, which will further increase the number of necessary analyses. This underscores the importance of building stable, high-throughput platforms with well-understood, closely-controlled technical components.

Figure 6: Plot of estimated study size necessary to make a statistically significant measurement vs. % effect size to be detected. A p-value cutoff of 0.05 was used after adjustment using the Bonferroni correction to conservatively account for multiple testing. Calculations are based on comparison between two different study groups using a two-sided t-test. Sample sizes were estimated using Cohen's d and a study power of 0.80.⁶⁹

CONCLUSIONS

In this study, a detailed experimental design was successfully utilized to dissect the contribution of 4 major technical components to the overall variability in our proteomics pipeline, using human brain tissue analysis as an example. Ideally, we would like to isolate and quantify all individual contributions to variation for precise guidance of the next-generation quantitative proteomics platform development. However, due to the practical constraints imposed by study size and limitations in sample availability, this is not feasible. From this analysis we found that the largest contribution came from the extraction component, which encompassed sample dissection and homogenization. Importantly, many quantitative proteomics strategies that use stable isotope-based approaches introduce the label downstream of the sample extraction step which contributes the most to overall variation.

Since this study unified both tissue dissection and homogenization steps it is unable to determine how much tissue dissection and homogenization individually contribute to the cumulative variability. However, a similar study design could be utilized to decouple the

two factors. These findings suggest that there are significant potential gains from in-depth analysis of the effect extraction area on tissue studies, as well as exploration and optimization of homogenization techniques.

The platform based on an Orbitrap (Thermo Fisher, San Jose, CA) mass spectrometer demonstrated good instrumental stability as evidenced by strong correlation throughout the study and a median CV of 33% for the analysis of brain tissue samples over a 2 week analysis time. The relatively small contribution to variability derived from the digestion (3.1%) and stability components (8.4%) demonstrates the effectiveness of the automated sample handling and the 4-column LC system used in these efforts. The remaining 16% of the variability was contributed by the fluctuations in instrumental response, and CV's obtained are in good agreement with other studies in the literature. Isolating of technical components allows a more detailed assessment of variability in our complex proteomic analysis pipeline, providing valuable insights to guide study design and future pipeline developments.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the patients and their families for their selfless donations. This project was supported by grants from the National Center for Research Resources (5 P41 RR018522-10) and the National Institute of General Medical Sciences (8 P41 GM103493-10) from the National Institutes of Health as well as NIH EUREKA grant R01-AG-034504 to A.J.M. Tissue resources from the University of Michigan were funded by NIH grant P50-AG08671. Portions of this research were supported by the National Institute of Allergy and Infectious Diseases (Y1-AI-8401) and by the Department of Energy Office of Biological and Environmental Research Genome Sciences Program under the Pan-omics project. Work was performed in the Environmental Molecular Science Laboratory, a U.S. Department of Energy (DOE) national scientific user facility at Pacific Northwest National Laboratory (PNNL) in Richland, WA. Battelle operates PNNL for the DOE under contract DE-AC05-76RLO01830.

REFERENCES

- (1). Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature*. 2003; 422:198–207. [PubMed: 12634793]
- (2). Ferguson PL, Smith RD. Proteome analysis by mass spectrometry. *Annu. Rev. Biophys. Biomolec. Struct.* 2003; 32:399–424.
- (3). Cravatt BF, Simon GM, Yates JR. The biological impact of mass-spectrometry-based proteomics. *Nature*. 2007; 450:991–1000. [PubMed: 18075578]
- (4). Xie F, Liu T, Qian WJ, Petyuk VA, Smith RD. Liquid Chromatography-Mass Spectrometry-based Quantitative Proteomics. *J. Biol. Chem.* 2011; 286:25443–25449. [PubMed: 21632532]
- (5). Zhu WH, Smith JW, Huang CM. Mass Spectrometry-Based Label-Free Quantitative Proteomics. *J. Biomed. Biotechnol.* 2010
- (6). Rivera-Burgos D, Regnier FE. Native Protein Proteolysis in an Immobilized Enzyme Reactor as a Function of Temperature. *Anal. Chem.* 2012; 84:7021–7028. [PubMed: 22845770]
- (7). Qian WJ, Jacobs JM, Liu T, Camp DG, Smith RD. Advances and challenges in liquid chromatography-mass spectrometry-based proteomics profiling for clinical applications. *Mol. Cell. Proteomics.* 2006; 5:1727–1744. [PubMed: 16887931]
- (8). Nilsson T, Mann M, Aebersold R, Yates JR, Bairoch A, Bergeron JJM. Mass spectrometry in high-throughput proteomics: ready for the big time. *Nat. Methods.* 2010; 7:681–685. [PubMed: 20805795]
- (9). Domon B, Aebersold R. Options and considerations when selecting a quantitative proteomics strategy. *Nat. Biotechnol.* 2010; 28:710–721. [PubMed: 20622845]

- (10). Nagaraj N, Mann M. Quantitative Analysis of the Intra- and Inter-Individual Variability of the Normal Urinary Proteome. *J. Proteome Res.* 2011; 10:637–645. [PubMed: 21126025]
- (11). Fenselau C, Yao XD. (18)O(2)-Labeling in Quantitative Proteomic Strategies: A Status Report. *J. Proteome Res.* 2009; 8:2140–2143. [PubMed: 19338309]
- (12). Yao XD, Freas A, Ramirez J, Demirev PA, Fenselau C. Proteolytic O-18 labeling for comparative proteomics: Model studies with two serotypes of adenovirus. *Anal. Chem.* 2001; 73:2836–2842. [PubMed: 11467524]
- (13). Petritis BO, Qian WJ, Camp DG, Smith RD. A Simple Procedure for Effective Quenching of Trypsin Activity and Prevention of (18)O-Labeling Back-Exchange. *J. Proteome Res.* 2009; 8:2157–2163. [PubMed: 19222237]
- (14). Qian WJ, Liu T, Petyuk VA, Gritsenko MA, Petritis BO, Polpitiya AD, Kaushal A, Xiao W, Finnerty CC, Jeschke MG, Jaitly N, Monroe ME, Moore RJ, Moldawer LL, Davis RW, Tompkins RG, Herndon DN, Camp DG, Smith RD. Inflammation; Host Response Injury, L. Large-Scale Multiplexed Quantitative Discovery Proteomics Enabled by the Use of an (18)O-Labeled “Universal” Reference Sample. *J. Proteome Res.* 2009; 8:290–299. [PubMed: 19053531]
- (15). Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics.* 2002; 1:376–386. [PubMed: 12118079]
- (16). Ong SE, Mann M. Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* 2005; 1:252–262. [PubMed: 16408053]
- (17). Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Hamon C. Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* 2003; 75:1895–1904. [PubMed: 12713048]
- (18). Ross PL, Huang YLN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlett-Jones M, He F, Jacobson A, Pappin DJ. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics.* 2004; 3:1154–1169. [PubMed: 15385600]
- (19). Dayon L, Hainard A, Licker V, Turck N, Kuhn K, Hochstrasser DF, Burkhard PR, Sanchez JC. Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. *Anal. Chem.* 2008; 80:2921–2931. [PubMed: 18312001]
- (20). Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 1999; 17:994–999. [PubMed: 10504701]
- (21). Hansen KC, Schmitt-Ulms G, Chalkley RJ, Hirsch J, Baldwin MA, Burlingame AL. Mass spectrometric analysis of protein mixtures at low levels using cleavable C-13-isotope-coded affinity tag and multidimensional chromatography. *Mol. Cell. Proteomics.* 2003; 2:299–314. [PubMed: 12766231]
- (22). Schmidt A, Kellermann J, Lottspeich F. A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics.* 2005; 5:4–15. [PubMed: 15602776]
- (23). Tebbe A, Schmidt A, Konstantinidis K, Falb M, Bisler B, Klein C, Aivaliotis M, Kellermann J, Siedler F, Pfeiffer F, Lottspeich F, Oesterhelt D. Life-style changes of a halophilic archaeon analyzed by quantitative proteomics. *Proteomics.* 2009; 9:3843–3855. [PubMed: 19670246]
- (24). Collier TS, Sarkar P, Franck WL, Rao BM, Dean RA, Muddiman DC. Direct Comparison of Stable Isotope Labeling by Amino Acids in Cell Culture and Spectral Counting for Quantitative Proteomics. *Anal. Chem.* 2010; 82:8696–8702. [PubMed: 20845935]
- (25). Collier TS, Randall SM, Sarkar P, Rao BM, Dean RA, Muddiman DC. Comparison of stable-isotope labeling with amino acids in cell culture and spectral counting for relative quantification of protein expression. *Rapid Commun. Mass Spectrom.* 2011; 25:2524–2532. [PubMed: 21818813]
- (26). Yao XD. Derivatization or Not: A Choice in Quantitative Proteomics. *Anal. Chem.* 2011; 83:4427–4439. [PubMed: 21495688]
- (27). Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000; 155:945–959. [PubMed: 10835412]

- (28). Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*. 2003; 164:1567–1587. [PubMed: 12930761]
- (29). Myers AJ, Gibbs JR, Awebster J, Rohrer K, Zhao A, Marlowe L, Kaleem M, Leung D, Bryden L, Nath P, Zismann VL, Joshupura K, Huentelman MJ, Hu-Lince D, Coon KD, Craig DW, Pearson JV, Holmans P, Heward CB, Reiman EM, Stephan D, Hardy J. A survey of genetic human cortical gene expression. *Nature Genet*. 2007; 39:1494–1499. [PubMed: 17982457]
- (30). Webster JA, Gibbs JR, Clarke J, Ray M, Zhang WX, Holmans P, Rohrer K, Zhao A, Marlowe L, Kaleem M, McCorquodale DS, Cuello C, Leung D, Bryden L, Nath P, Zismann VL, Joshupura K, Huentelman MJ, Hu-Lince D, Coon KD, Craig DW, Pearson JV, Heward CB, Reiman EM, Stephan D, Hardy J, Myers AJ, Grp NA-N. Genetic Control of Human Brain Transcript Expression in Alzheimer Disease. *Am. J. Hum. Genet*. 2009; 84:445–458. [PubMed: 19361613]
- (31). Livesay EA, Tang KQ, Taylor BK, Buschbach MA, Hopkins DF, LaMarche BL, Zhao R, Shen YF, Orton DJ, Moore RJ, Kelly RT, Udseth HR, Smith RD. Fully automated four-column capillary LC-MS system for maximizing throughput in proteomic analyses. *Anal. Chem*. 2008; 80:294–302. [PubMed: 18044960]
- (32). Maiolica A, Borsotti D, Rappsilber J. Self-made frits for nanoscale columns in proteomics. *Proteomics*. 2005; 5:3847–3850. [PubMed: 16130174]
- (33). Kelly RT, Page JS, Luo QZ, Moore RJ, Orton DJ, Tang KQ, Smith RD. Chemically etched open tubular and monolithic emitters for nanoelectrospray ionization mass spectrometry. *Anal. Chem*. 2006; 78:7796–7801. [PubMed: 17105173]
- (34). Zimmer JS, Monroe ME, Qian WJ, Smith RD. Advances in proteomics data analysis and display using an accurate mass and time tag approach. *Mass Spectrom Rev*. 2006; 25:450–482. [PubMed: 16429408]
- (35). Wang Y, Yang F, Gritsenko MA, Wang Y, Clauss T, Liu T, Shen Y, Monroe ME, Lopez-Ferrer D, Reno T, Moore RJ, Klemke RL, Camp DG II, Smith RD. Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics*. 2011; 11:2019–2026. [PubMed: 21500348]
- (36). Kim S, Mischerikow N, Bandeira N, Navarro JD, Wich L, Mohammed S, Heck AJR, Pevzner PA. The Generating Function of CID, ETD, and CID/ETD Pairs of Tandem Mass Spectra: Applications to Database Search. *Mol. Cell. Proteomics*. 2010; 9:2840–2852. [PubMed: 20829449]
- (37). Jaitly N, Mayampurath A, Littlefield K, Adkins J, Anderson G, Smith R. Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data. *BMC Bioinformatics*. 2009; 10:87. [PubMed: 19292916]
- (38). Monroe ME, Tolic N, Jaitly N, Shaw JL, Adkins JN, Smith RD. VIPER: an advanced software package to support high-throughput LC-MS peptide identification. *Bioinformatics*. 2007; 23:2021–2023. [PubMed: 17545182]
- (39). Stanley JR, Adkins JN, Slys GW, Monroe ME, Purvine SO, Karpievitch YV, Anderson GA, Smith RD, Dabney AR. A Statistical Method for Assessing Peptide Identification Confidence in Accurate Mass and Time Tag Proteomics. *Anal. Chem*. 2011; 83:6135–6140. [PubMed: 21692516]
- (40). Andreev VP, Petyuk VA, Brewer HM, Karpievitch YV, Xie F, Clarke J, Camp D, Smith RD, Lieberman AP, Albin RL, Nawaz Z, El Hokayem J, Myers AJ. Label-Free Quantitative LC-MS Proteomics of Alzheimer's Disease and Normally Aged Human Brains. *J. Proteome Res*. 2012; 11:3053–3067.
- (41). Zhang B, Chambers MC, Tabb DL. Proteomic parsimony through bipartite graph analysis improves accuracy and transparency. *J. Proteome Res*. 2007; 6:3549–3557. [PubMed: 17676885]
- (42). Kall L, Canterbury JD, Weston J, Noble WS, MacCoss MJ. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods*. 2007; 4:923–925. [PubMed: 17952086]
- (43). Polpitiya AD, Qian WJ, Jaitly N, Petyuk VA, Adkins JN, Camp DG, Anderson GA, Smith RD. DAnTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics*. 2008; 24:1556–1558. [PubMed: 18453552]

- (44). Milac TI, Randolph TW, Wang P. Analyzing LC-MS/MS data by spectral count and ion abundance: two case studies. *Stat. Interface.* 2012; 5:75–87.
- (45). Clough T, Braun S, Fokin V, Ott I, Ragg S, Schadow G, Vitek O. Statistical design and analysis of label-free LC-MS proteomic experiments: a case study of coronary artery disease. *Methods in molecular biology* (Clifton, N.J.). 2011; 728:293–319.
- (46). Chang CY, Picotti P, Huttenhain R, Heinzelmann-Schwarz V, Jovanovic M, Aebersold R, Vitek O. Protein Significance Analysis in Selected Reaction Monitoring (SRM) Measurements. *Mol. Cell. Proteomics.* 2012; 11
- (47). Callister SJ, Barry RC, Adkins JN, Johnson ET, Qian WJ, Webb-Robertson BJM, Smith RD, Lipton MS. Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *J. Proteome Res.* 2006; 5:277–286. [PubMed: 16457593]
- (48). Chambers RA, Cleveland WS, Kleiner B, Tukey PA. Graphical Methods for Data Analysis. 1983
- (49). Hu S, Loo JA, Wong DT. Human body fluid proteome analysis. *Proteomics.* 2006; 6:6326–6353. [PubMed: 17083142]
- (50). Ray S, Reddy PJ, Jain R, Gollapalli K, Moiyadi A, Srivastava S. Proteomic technologies for the identification of disease biomarkers in serum: Advances and challenges ahead. *Proteomics.* 2011; 11:2139–2161. [PubMed: 21548090]
- (51). Gan CS, Chong PK, Pham TK, Wright PC. Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ). *J. Proteome Res.* 2007; 6:821–827. [PubMed: 17269738]
- (52). Geiger T, Wisniewski JR, Cox J, Zanivan S, Kruger M, Ishihama Y, Mann M. Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. *Nat. Protoc.* 2011; 6:147–157. [PubMed: 21293456]
- (53). Zellner M, Veitinger M, Umlauf E. The role of proteomics in dementia and Alzheimer's disease. *Acta Neuropathol.* 2009; 118:181–195. [PubMed: 19259691]
- (54). Nakamura S, Kawamata T, Akiguchi I, Kameyama M, Nakamura N, Kimura H. Expression of monoamine oxidase B activity in astrocytes of senile plaques. *Acta Neuropathol.* 1990; 80:419–425. [PubMed: 2239154]
- (55). Gottfries CG. Neurochemical aspects on aging and diseases with cognitive impairment. *J Neurosci Res.* 1990; 27:541–547. [PubMed: 2079715]
- (56). Wilson KE, Marouga R, Prime JE, Pashby DP, Orange PR, Crosier S, Keith AB, Lathe R, Mullins J, Estibeiro P, Bergling H, Hawkins E, Morris CM. Comparative proteomic analysis using samples obtained with laser microdissection and saturation dye labelling. *Proteomics.* 2005; 5:3851–3858. [PubMed: 16145713]
- (57). Han MH, Hwang SI, Roy DB, Lundgren DH, Price JV, Ousman SS, Fernald GH, Gerlitz B, Robinson WH, Baranzini SE, Grinnell BW, Raine CS, Sobel RA, Han DK, Steinman L. Proteomic analysis of active multiple sclerosis lesions reveals therapeutic targets. *Nature.* 2008; 451:1076–U1072. [PubMed: 18278032]
- (58). Bagnato C, Thumar J, Mayya V, Hwang SI, Zebroski H, Claffey KP, Haudenschild C, Eng JK, Lundgren DH, Han DK. Proteomics analysis of human coronary atherosclerotic plaque - A feasibility study of direct tissue proteomics by liquid chromatography and tandem mass spectrometry. *Mol. Cell. Proteomics.* 2007; 6:1088–1102. [PubMed: 17339633]
- (59). Butt RH, Coorsen JR. Pre-extraction sample handling by automated frozen disruption significantly improves subsequent proteomic analyses. *J. Proteome Res.* 2006; 5:437–448. [PubMed: 16457611]
- (60). Tastet C, Charmont S, Chevallet M, Luche S, Rabilloud T. Structure-efficiency relationships of zwitterionic detergents as protein solubilizers in two-dimensional electrophoresis. *Proteomics.* 2003; 3:111–121. [PubMed: 12601803]
- (61). Luche S, Santoni V, Rabilloud T. Evaluation of nonionic and zwitterionic detergents as membrane protein solubilizers in two-dimensional electrophoresis. *Proteomics.* 2003; 3:249–253. [PubMed: 12627377]
- (62). Wisniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. *Nat. Methods.* 2009; 6:359–U360. [PubMed: 19377485]

- (63). Zhou JY, Dann GP, Shi TJ, Wang L, Gao XL, Su DA, Nicora CD, Shukla AK, Moore RJ, Liu T, Camp DG, Smith RD, Qian WJ. Simple Sodium Dodecyl Sulfate-Assisted Sample Preparation Method for LC-MS-Based Proteomics Applications. *Anal. Chem.* 2012; 84:2862–2867. [PubMed: 22339560]
- (64). Gallien S, Duriez E, Domon B. Selected reaction monitoring applied to proteomics. *J. Mass Spectrom.* 2011; 46:298–312. [PubMed: 21394846]
- (65). Picotti P, Rinner O, Stallmach R, Dautel F, Farrah T, Domon B, Wenschuh H, Aebersold R. High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nat. Methods.* 2010; 7:43–U45. [PubMed: 19966807]
- (66). Hossain M, Kaleta DT, Robinson EW, Liu T, Zhao R, Page JS, Kelly RT, Moore RJ, Tang KQ, Camp DG, Qian WJ, Smith RD. Enhanced Sensitivity for Selected Reaction Monitoring Mass Spectrometry-based Targeted Proteomics Using a Dual Stage Electrodynamical Ion Funnel Interface. *Mol. Cell. Proteomics.* 2011; 10
- (67). Gillet LC, Navarro P, Tate S, Rost H, Selevsek N, Reiter L, Bonner R, Aebersold R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics.* 2012; 11
- (68). Shi TJ, Su D, Liu T, Tang KQ, Camp DG, Qian WJ, Smith RD. Advancing the sensitivity of selected reaction monitoring-based targeted quantitative proteomics. *Proteomics.* 2012; 12:1074–1092. [PubMed: 22577010]
- (69). Cohen, J., editor. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed.. Lawrence Erlbaum Associates; Hillsdale, NJ: 1988.

Figure 1.

Flowchart describing the 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_1^2 , to determine the contribution to the total variability.

Figure 2.

Comparison of spike-in normalization with global scaling normalization. The first boxplot shows distribution of peptide CV's using raw data, the second boxplot shows peptide CV's normalized using spike-in peptide intensities, the third shows peptide CV's after global scaling normalization. A) digestion replicates B) instrumental stability replicates.

Figure 3. Boxplots showing the distribution of Pearson correlation coefficients among the different levels of replicates.

Figure 4. Boxplots showing the distributions of peptide-level CV's for the 4 technical components. Notches were added to the boxplot to demonstrate that the difference in median between components is statistically significant.⁴⁸

Figure 5. Contributions to variability. A.) Boxplot showing the distribution of contributions to variability calculated for individual peptides. B.) Piechart demonstrating the median contribution to variability for each of the four technical components: extraction 72%, digestion 3.1 %, instrumental stability 8.4%, instrumental variance 16%.

Figure 6.

Plot of estimated study size necessary to make a statistically significant measurement vs. % effect size to be detected. A p-value cutoff of 0.05 was used after adjustment using the Bonferroni correction to conservatively account for multiple testing. Calculations are based on comparison between two different study groups using a two-sided t-test. Sample sizes were estimated using Cohen's d and a study power of 0.80.⁶⁹