

The Proteomic Characterization of Plasma or Serum from HIV-Infected Patients

Nicole A. Haverland, Lance M. Villeneuve, Pawel Ciborowski, and Howard S. Fox

Abstract

Proteomics holds great promise for uncovering disease-related markers and mechanisms in human disorders. Recent advances have led to efficient, sensitive, and reproducible methods to quantitate the proteome in biological samples. Here we describe the techniques for processing, running, and analyzing samples from HIV-infected plasma or serum through quantitative mass spectroscopy.

Key words Mass spectroscopy, SWATH-MS, HIV, Proteomics, Serum, Plasma, Bayesian, Z-score

1 Introduction

Plasma/serum proteomics holds a vast potential for new biomarker discovery [1, 2]. Blood, which can repeatedly be harvested from patients with relatively low invasion and at a relatively low cost, is an attractive clinical material. Although blood contains a vast assortment of proteins and metabolites, numerous issues, such as reproducibility and sensitivity, have plagued plasma proteomics, thereby limited successes in the field. With the development of new mass spectrometry techniques and technologies, comprehensive, replicable analysis of the plasma proteome has become feasible.

Mass spectrometry-based quantitative techniques have been at the forefront of analytical approaches in biomarker discovery from the very beginning of proteomics [3, 4]. During one and a half decades of blood/plasma/serum/CSF proteomics, researchers have attempted to identify and quantify these proteomes using various methods including *in vitro* labeling techniques (such as iTRAQ, TMT, ICAT, super-SILAC) as well as label-free

Nicole A. Haverland and Lance M. Villeneuve have equally contributed to this chapter.

techniques, often using shotgun proteomics [5]. However, while each method has provided insight into various disorders, reproducibility and broad validation remains an unresolved issue. Each of these techniques suffers from various limiting factors with one common underlying limitation being data-dependent acquisition (DDA). The application of data-independent acquisition (DIA), in which all molecular species are recorded, opens new avenue in mass spectrometry-based, body fluid-based biomarker discovery. Sequential Window data-independent Acquisition of the Total High-resolution Mass Spectra (SWATH-MS) platform offers higher dynamic range of linearity of recorded ion intensities vastly improving precision and accuracy of quantification and sensitivity [6]. SWATH-MS uses DIA to generate the experimental data, which is searched against a library constructed of DDA acquired samples. In addition, the DIA data collected for the experimental samples creates a permanent spectral record that can be utilized to extract additional information in future analyses with other DDA libraries. Although at this time only a limited number of peer-reviewed studies employing SWATH-MS platform have been published, initial results are highly promising in the prospects of contribution to biomarker discovery.

SWATH-based proteomic analysis offers additional benefits compared to previous proteomic approaches. Using SWATH, additional fractionation is commonly unnecessary and detection of low abundance peptides is possible. However, to gain the most out of SWATH mass spectrometry (as with any experiment), optimization is necessary. Here we present our optimized methods for four critical steps of SWATH-MS: DDA for the building of the library, DIA for running of the experimental samples, data processing, and statistical analysis. In addition to their use for plasma/serum samples, these steps also apply to other applications of SWATH-MS.

2 Materials and Equipment

2.1 *Sample Processing*

1. Protease inhibitor cocktail.
2. Sodium dodecyl sulfate (SDS).
3. Seppro IgY14 Column (Sigma-Aldrich).
4. VIVASPIN 15R with 5000 MWCO.
5. Centrifuge with swinging bucket rotor.
6. HRM Calibration kit Standard Peptides (Biognosys).

2.2 *Instrumentation*

1. Centrifugal vacuum concentrator with rotor for 1.5–2.0 mL microcentrifuge tubes.
2. cHiPLC-nanoflex system (Eksigent).
3. TripleTOF 5600 Mass Spectrometer equipped with a NanoSpray III Ion Source (AB SCIEX).

2.3 Supplies

1. Nano-cHiPLC column 75 $\mu\text{m} \times 15$ cm ChromXp C18-CL 3 μm 300 Å (Eksigent).
2. Nano-cHiPLC Trap column 200 $\mu\text{m} \times 0.5$ mm ChromXp C18-CL 3 μm 300 Å (Eksigent).
3. Oasis MCX 1 cm^3 column (Waters).
4. Trap-elute jumper chip (Eksigent).
5. 0.1 % Formic Acid in Water.
6. Solution A: 0 % Aqueous Solution with 0.1 % Formic Acid.
7. Solution B: 100 % Acetonitrile with 0.1 % Formic Acid.
8. National Limited Volume Wide Opening Plastic Crimp Top Autosampler Vials; 450 μL capacity (Thermo Scientific).
9. 11 mm Snap-It Cap for Autosampler Vials, 6 mm hole (Thermo Scientific).

2.4 Data Processing: Generating the Spectral Library

1. 8-core computer with ProteinPilot installed and licensed (AB SCIEX).

2.5 Data Processing: Targeted Data Extraction

1. Computer with PeakView v. 2.1 Software installed and licensed (AB SCIEX) and the add-on, Protein Quantitation 1.0 MicroApp, installed and licensed.

3 Methods

3.1 Sample Processing for SWATH-MS Designed Experiments

1. Obtain plasma or serum sample from patients or biobanks (*see* **Notes 1** and **2**).
2. If samples are frozen, thaw, and immediately upon thawing add 50 μL of 20 \times Protease Inhibitors per milliliter of the sample to prevent protein degradation. Samples are mixed with the inhibitors by inversion or gentle vortexing and then immediately placed on ice. To delipidate, the samples should be centrifuged at 18,000 $\times g$ at 4 °C for 15 min and the middle layer of the cleared plasma/serum collected.
3. Deplete samples of highly abundant proteins using a commercial mix of immobilized antibodies (*see* **Note 3**). The standard manufacturer protocol for immunodepletion should be followed for the Seppro IgY14 columns. This kit is available in a spin column format or as HPLC columns of different sizes depending on the volume of sample to be depleted, and contains all necessary buffers. Concentrate the flow-through depleted samples using a VIVASPIN 15R spin concentrator, centrifuging at 4000 $\times g$ for approximately 1.5 h. Samples may be frozen at this stage if desired.

4. Bring samples to 4 % sodium dodecyl sulfate (SDS) (*see Note 4*). Centrifuge at $400 \times g$ to pellet any debris. Transfer supernatant to a clean sample container. Quantify the protein concentration within the sample using the Pierce 660 protein quantification kit (*see Note 5*).
5. Digest the samples using trypsin and using filter-assisted sample preparation (FASP; <http://www.nature.com/nmeth/journal/v6/n5/extref/nmeth.1322-S1.pdf>) protocol. FASP is compatible with SDS and has many benefits over in-gel or in-solution approaches [7] (*see Note 6*). We recommend digesting 50 μg of protein using a protein:enzyme ratio of <1:50.
6. Desalt the digested samples using an Oasis mixed cation exchange column following manufacturer's protocols. Desiccate the desalted sample using a vacuum concentrator.
7. Resuspend the peptides in a minimal volume of 0.1 % formic acid in HPLC-grade water (*see Note 7*). Perform peptide quantification based on spectral absorbance at 205 nm on a NanoDrop 2000 [8] (*see Note 8*). Remove an aliquot (2 μg or less; keep this consistent across biological and technical replicates) of cleaned peptides from each sample and transfer to a clean autosampler vial (*see Note 9*). If the total volume is >6 μL , desiccate the sample and resuspend in 6 μL 0.1 % formic acid in HPLC-grade water. If the total volume is <6 μL , bring the solution to a total volume of 6 μL .
8. (Optional for DIA) Spiking-in peptides: Peptide spiking-in is a process whereby artificial peptides are added to the database and to each experimental sample. These artificial peptides have various predicted elution times. After determining the change in experimental elution times from predicted elution times, the elution profile can be shifted to enable better matching of the SWATH library reference spectra to the experimentally obtained DIA spectra. Add an equal amount of artificial peptides from the HRM calibration kit to each DIA sample.

3.2 Mass Spectrometry

1. Replace nano-cHiPLC columns if necessary (*see Note 10*). The Eksigent cHiPLC system requires three chips: the cHiPLC column that is used for elution, the trap column that is used during sample loading, and a trap-and-elute jumper chip.
2. *See Table 1* for the LC method for LC-MS/MS analyses of tryptic digested peptides (*see Note 11*). Equilibrate cHiPLC columns. If the Eksigent LC system is used, insert pre-run flush for 0.1 min using 100 % initial flow rate into the LC in the second tab of Eksigent method.
3. *See Table 2* to prepare mass spectrometry data acquisition methods for DDA and DIA experiments. For both methods, the mass spectrometer will be operated in high sensitivity mode.

Table 1
LC method

Sample loading			
Flow mode	Independent		
Time (min)	Flow Soln. A ($\mu\text{L}/\text{min}$)	Flow Soln. B ($\mu\text{L}/\text{min}$)	Event
0	10	0	
8.5	10	0	
9	2	0	
10	2	0	
Elution			
Flow mode	Conserved. Total flow rate = $0.3 \mu\text{L}/\text{min}$		
Time (min)	% Soln. A	% Soln. B	Event (<i>see Note 26</i>)
0	95	5	AUX3 TTL Low
0	95	5	AUX4 TTL High
1	95	5	AUX3 TTL High
180	65	35	
182	10	90	
192	10	90	
193	95	5	AUX4 TTL Low
200	95	5	

Samples for the library must be run using the DDA method, while experimental samples for SWATH-MS analysis must be run using DIA method. For information on samples for library construction, *see Note 12*.

4. Retrieve the autosampler vial(s) prepared in **step 7**, which contain the resuspended peptides. Transfer the vials to the autosampler and assign the samples to queue accordingly (*see Note 13*). Ensure that the autosampler lids are on flush and the tubes are not crooked, as this may result in breaking the autosampler needle or unequal sample uptake.
5. Start queue. Each sample will take approximately 3.5 h for completion using the LC methods provided in **Table 1**. The total ion current (TIC) chromatogram can be used to monitor sample elution during the run within the Analyst program (*see Note 14*).
6. As samples finish mass spectrometry analysis, the TICs can be overlaid in the PeakView software using the open multiple

Table 2
Mass spectrometry methods

DDA	
Charge state	From +2 to +5
Intensity threshold	>100 counts
Switch after	50 spectra
Advanced settings	True
Always exclude	True
Exclude for	15 s
Mass tolerance units	mDa
Mass tolerance	50
Use inclusion list	False
Use exclusion list	False
Ignore peaks within	6 Da
Real time	None
Dynamic collision energy	True
Fragment intensity multiplier	2
Maximum accumulation	2 s
Allow standard filters for smart IDA	True
Number of cycles	7615
Polarity	Positive
Period cycle time	2798 ms
Pulser frequency	14.170 kHz
ISVF (IonSpray voltage floating)	2400 V
Pulser frequency	14.170 kHz
Precursor scan (MS1) experiment type	TOF MS
MS1 accumulation time	250.0 ms
MS1 start mass	400.0 Da
MS1 end mass	1800.0 Da
Precursor fragmentation (MS2) experiment type (50 selections)	TOF MS ²
MS2 accumulation time	50.0 ms
MS2 start mass	100.0 Da
MS2 end mass	1800.0 Da

(continued)

Table 2
(continued)

DDA	
DIA	
Number of cycles	3555
Polarity	Positive
Period cycle time	3363 ms
Pulser frequency	14.170 kHz
ISVF (IonSpray voltage floating)	2400 V
Pulser frequency	14.170 kHz
Precursor scan (MS1) experiment type	TOF MS
MS1 accumulation time	50.0 ms
MS1 start mass	400.0 Da
MS1 end mass	1250.0 Da
SWATH-MS (MS2) experiment type	TOF MS ²
MS2 accumulation time	96.0 ms
MS2 start mass	100.0 Da
MS2 end mass	1800.0 Da
SWATH-MS experiment mass window	25 + 1 Da overlap
Fragment conditions	Rolling collision energy, charge state +2, collision energy spread of 15 V

WIFF tool to compare the chromatograms and evaluate differences between samples (*see Note 15*).

7. Following completion of all mass spectrometry, transfer all files to the hard drive of the computer that will be used for database searching of library samples and targeted data extraction of SWATH-MS files (*see Note 16*).

3.3 Data Processing: Generating the Spectral Library

1. If not already performed, transfer all DDA-generated files that will be used for creating the spectral library to the computer hard drive. The computer must have ProteinPilot installed.
2. Compile the FASTA file that will be used for database searching (*see Note 17*). For this step, we use the UniProt-SwissProt (www.uniprot.org) database to export the reference proteomes for *Homo sapiens* (search “organism:9606 AND reviewed:yes AND keyword:1185”) and for HIV-1 (search: “taxonomy:11706 AND reviewed:yes AND keyword:1185”). A word processor,

such as notepad, can be used to merge the files and add any additional FASTA sequences, such as the file with common laboratory contaminants provided by the AB SCIEX. Transfer the newly generated FASTA file to the databases folder within the AB SCIEX ProteinPilot Application folder.

3. (Optional for DIA) Inclusion of artificial peptides in the FASTA database. Information on the spiked-in peptides must be added to the FASTA database manually (*see Note 18*). Open the FASTA database in Notepad. Provide an entry for each artificial peptide. The sequence and name (Artificial names are ok) must be provided for each peptide (*see Note 19*).
4. Launch ProteinPilot. In the workflow tasks panel, click “LC...” under the “Identify Proteins” tab. Use the “Add...” button to add DDA samples to the search file. Process the file using a new paragon method (Table 3) and save the method using the “Save As...” button. Back in the “Identify Proteins” dialog box, save the results file and assign its location using the “Save As...” button. Click the “Process” button to begin the search. The file that is generated is a .group file, which will be uploaded as the reference spectral library during targeted data extraction.

Table 3
Paragon method

Describe sample	
Sample type	Unlabeled
Cys alkylation	Iodoacetamide
Digestion	Trypsin
Instrument	TripleTOF 5600
Special factors	None selected
Species	None
Specify processing	
Quantitate, bias correction, background correction	Not able to be selected
ID Focus	Biological modifications, amino acid substitutions
Database	FASTA database compiled in Data Processing: Generating the Spectral Library Step 2 .
Search effort	Thorough ID
Results quality	Detected protein Threshold >0.05 (10 %)
Run false discovery rate analysis	Checked

3.4 Data Processing: Targeted Data Extraction

1. If not already performed, transfer all DIA-generated files that will be used for targeted data extraction to the computer hard drive. The computer must have PeakView installed and licensed with the Protein Quantitation MicroApp installed.
2. Launch PeakView. Under the “Quantitation” menu, click “Import Ion Library.” Select the group file that was generated in Data Processing: Generating the Spectral Library **step 4** (*see Note 20*). The upload time will vary from minutes to hours and is dependent on the size of the .group file and computer processor speed. After the library has successfully been loaded, a dialog box will automatically appear and request selection of the SWATH-MS files that will be used for targeted data extraction. Select all the files that will be used for export.
3. If applicable, select the peptides that will be used for retention time (RT) correction, using either the spiked-in peptides or selecting high-abundance endogenous peptides (*see Note 21*). To do this, search for the protein of interest and click the peptides that will be used for correction so that a check mark is apparent next to the peptide sequence. Next, click the “Add RT-Cal” button to add selected peptides to the RT calibration set. To edit the set of peptides used for RT calibration, click the “Edit RT-Cal” button and select peptides for deletion. In addition, use the “Edit RT-Cal” tool to calculate RT fit and apply RT modifications.
4. Following RT correction, click the processing settings button under the SWATH Processing dialog box. Set the processing settings accordingly. We use the following parameters: Up to 30 peptides, 6 transitions, 95 % peptide confidence threshold, 1 % false discovery rate threshold, exclude shared peptides, XIC window of 12 min, and XIC width of 75 ppm. These settings will likely require optimization dependent on the samples used for mass spectrometry (*see Note 22*).
5. Click process to perform targeted data extraction. Following processing, export all information using the Quantitation menu → SWATH processing → Export → All. The file that is generated is an .xlsx and can be opened in a database application or alternative statistical platforms capable of importing .xlsx files.

3.5 Data Processing: Normalization of Data and Statistical Analysis

We use one of two distinct approaches for normalization and statistical analysis of SWATH-MS data.

3.5.1 Normalization in MarkerView and Statistical analysis

When investigating any proteomic data, there is a necessity to normalize to correct for any error in preparation. To compensate for this error, we recommend normalization in MarkerView.

MarkerView offers a wide variety of normalization parameters that may be chosen dependent on the experimental design. Using MarkerView normalization in conjunction with a Bayesian analysis, a probabilistic statistical approach, offers additional benefits. Bayesian analysis can analyze data of high dimensionality by demonstrating the data must follow the rules of probability introduced by the Bayes theorem [9]. By following these rules, Bayesian analysis is able to correctly analyze data with fewer biological replicates than a simple *t*-test. This can be performed in eight steps:

- Step 1. Export the area under curve data from PeakView as a MarkerView file.
- Step 2. Open the extracted ion chromatogram (XIC) data in MarkerView.
- Step 3. Normalize the data choosing the best normalization method based upon the design of the experiment (*see Note 23*).
- Step 4. Statistical analysis of mass spectrometry data is necessary to draw strong conclusions from the data. Unfortunately, when comparing multiple proteins in multiple samples, common multiple testing corrections (e.g., Bonferroni) render everything insignificant. To combat these problems, Bayesian analysis followed our multiple testing correction is a viable method to analyze data.
- Step 5. CyberT (<http://cybert.ics.uci.edu> or <http://molgen51.biol.rug.nl/cybert/>), an online Bayesian analysis calculator, can be used for analyzing high dimension mass spectrometry data [9, 10].
- Step 6. Format the data for upload into CyberT in accordance with the recommendations provided by the online calculator.
- Step 7. Select the correct analysis parameters.

For normalization, CyberT can perform normalizations but is limited in number of normalization methods. We recommend loading MarkerView normalized data.

For the Bayesian analysis parameters, we recommend following CyberT instructions for Sliding window size. For the Bayesian confidence value, we recommend multiplying the number of replicates by 3 and using the corresponding value.

For multiple testing correction, multiple methods may be calculated through a single analysis by selecting to “Compute multiple test corrections” under “Standard Multiple Hypothesis Testing Corrections.”

We recommend computing the Posterior Probability of Differential Expression (PPDE). PPDE gives the prob-

ability of observing a real change. Cumulative PPDE is the best method for determining statistical significance because it corrects the PPDE to a false discovery rate of 0.05.

Proteins with a p -value <0.05 and a Cumulative PPDE >0.95 are considered to be significantly altered between samples.

Step 8. Export the data to Excel or other data formats for further analysis.

3.5.2 Normalization by Relative Abundance and Parametric Statistical Testing

Normalization by relative abundance uses the z-score to assign a value that represents the relative distribution of each protein within a given dataset/condition. This value is then used to measure alterations in the relative abundance of a given protein between conditions. An advantage of the z-score transformation is that normalization of the dataset to the standard normal distribution can be performed independently for each dataset, which allows for the rapid inclusion of multiple conditions, replicates, and comparisons. The change in relative abundance between conditions for any given protein is termed z-difference, and this measure can be used for parametric statistical testing, including the z-test [11, 12] and even other parametric tests, such as the ANOVA for multiple conditions and comparisons. The methods below provide a step-by-step procedure for normalization and statistical testing, including the paired and unpaired z-test.

1. Open the targeted data extraction file that was exported from PeakView and move (as a copy) the protein data to a new Excel spreadsheet. In this manner, the original export retains its integrity if the raw data is requested.
2. In the new spreadsheet, transform the raw intensity data using the natural log (ln). This transformation will normalize the data so that the entire dataset better approaches a normal distribution that is required for statistical testing (*see Note 24*).
3. Z-transformation of the data requires calculation of the mean and standard deviation of all proteins within a single dataset (one replicate, one condition). The z-score is a quantitative representation of the relative abundance of a protein and can be calculated using the following equation:

$$z = \frac{x - \mu}{\sigma}$$

where x is the natural log transformed raw intensity value for a given protein, μ is the overall average of natural log transformed raw intensity values for all proteins within a single dataset, and σ is the standard deviation of the natural log transformed raw intensity values for all proteins within a single dataset.

4. Based on the experimental design, choose the appropriate statistical test: the paired samples or independent samples z-test. For multiple comparisons/conditions, consider using a statistical test such as an ANOVA. The paired samples z-test is the more appropriate statistical test when comparing the expression of a protein before and after a condition within the same donor whereas the independent samples z-test is the more appropriate statistical test when comparing the overall expression of a given protein within a cohort of control subjects as compared to the overall expression of a given protein within a cohort of subjects with a defined disease or condition. The paired and independent samples z-tests are conceptually equivalent to the paired and independent samples Student's *t*-tests, respectively.
5. For paired samples, use the following formula to calculate the z-test statistic [12]:

$$\text{ztest}_{\text{paired}} = \frac{\bar{d} - D}{\frac{\sigma_d}{\sqrt{n}}}$$

where \bar{d} is the mean value of pairwise differences across all replicates, D is the hypothesized mean of the pairwise differences across all replicates (most often 0), σ_d is the standard deviation of the pairwise differences across all replicates, and n is the total number of pairwise comparisons (replicates).

6. For independent samples, use the following formula to calculate the z-test statistic [11]:
7. After the z-test statistic is calculated, determine the *p*-value using the standard normal distribution for a two-tailed test (e.g., a test statistic of 1.96 = 95 % confidence or $p=0.05$) (see **Note 25**)

$$\text{ztest}_{\text{ind}} = \frac{\overline{x_{\text{exp}}} - \overline{x_{\text{cont}}}}{\sqrt{\frac{\sigma_{\text{exp}}^2}{n_{\text{exp}}} + \frac{\sigma_{\text{cont}}^2}{n_{\text{cont}}}}}$$

where $\overline{x_{\text{exp}}}$ is the mean value of a given protein across all replicates in the “experimental condition,” $\overline{x_{\text{cont}}}$ is the mean value of a given protein across all replicates in the “control condition,” σ_{exp}^2 is the variance of the protein expression across all replicates in the “experimental condition,” σ_{cont}^2 is the variance of the protein expression across all replicates in the “control condition,” and n is the total number of samples for each condition.

3.6 Limitations of Proteomics

Despite unquestionable progress in acquisition of mass spectra, a major limitation of quantitative proteomics is the very high dynamic range of protein concentrations in highly complex mixtures of proteins and peptides generated by any method of controlled (i.e., enzymatic) fragmentation. This limitation applies to all methods: label free, chemical labeling, or metabolic labeling methods and researchers are advised to look for potential systemic bias. Approaches of extensive fractionation leading to reduction of complexity of samples have been used and will be refined in the future. These approaches help to reduce the impact of the high dynamic range of concentrations by providing high quality spectra for low abundant proteins as well as remove suppressive effect of many spectra from highly abundant proteins. For plasma, an immunodepletion of highly abundant proteins as described in Sample Processing for SWATH-MS Designed Experiments **Step 3** has been widely used to reduce interference from highly abundant proteins and is recommended.

No matter how refined and advanced the proteomics technology becomes, nothing can make up for problems in experimental design. In addition to the issues in **Note 1**, adequate group sizes, proper controls, consistency in specimen acquisition, processing and storage, and other aspects important in studies of biospecimens from diverse human populations apply to proteomic experiments.

4 Notes

1. People with HIV infection offer interesting problems for performing serum/plasma biomarker studies. People with HIV-1 infection have a high incidence of comorbidities including hepatitis C co-infection [13], cardiovascular, liver, and kidney disease [14], and all these factors may influence serum/plasma protein composition. When performing biomarker studies, these confounding issues have the potential to confound experimental results. Therefore, additional effort should be focused on obtaining a thorough background on each patient to identify and account for confounding variables.
2. All work on patients as well as specimens derived from patients must be done under approval from the proper regulatory bodies such as Institutional Review Boards. In addition, working with human samples (and here with known infectious agents) must be done under appropriate safety standards (in general, BSL-2). Both of these aspects should be done following your institutional requirements. Samples should be drawn, processed, and stored under standardized conditions.
3. Immunodepletion will reduce the concentration of high-abundance serum/plasma proteins and in doing so will help

improve the detection and quantification of other lower abundance proteins that may have otherwise been masked. However how many most abundant proteins should be removed to facilitate analysis, and the best means to do this, is still an open question. We currently recommend using the Seppro IgY14 column (with avian antibodies targeting the following proteins: albumin, α_1 -antitrypsin, IgM, haptoglobin, fibrinogen, α_1 -acid glycoprotein, apolipoprotein A-I and A-III, apolipoprotein B, IgG, IgA, transferrin, α_2 -macroglobulin, and complement C3) from Sigma-Aldrich.

4. SDS inactivates HIV-1 [15]; depending on one's institutional biosafety requirements the safety precautions may differ following viral inactivation.
5. We suggest using the Pierce 660 protein quantification methods because of the rapidity of analysis. Alternative methods (Bradford, BCA, etc.) can also be used if desired. However, we do not recommend using spectral absorbance for quantification as this measure is dependent on the presence of aromatic amino acids and can lead to a mistaken representation for protein concentration.
6. Numerous digest procedures are available online and provided in the literature and through protease manufacturer websites. In addition to trypsin, other enzymes may also be used, such as LysC; however these enzymes will require additional optimization. As an alternate to SDS treatment followed by FASP, samples may also be digested using a standard in-gel or in-solution tryptic digest protocol without the use of SDS; however remain aware of biosafety considerations.
7. For a sample with approximately 50 μg of starting protein, we recommend resuspending the sample in no more than 25 μL . Using this volume will allow for accurate analysis of peptide quantification by using a minimal amount of sample. In our experience, peptide quantity is usually 20–50 % of the amount of protein measured in Sample Processing for SWATH-MS Designed Experiments, **step 4**.
8. Using absorbance at a wavelength of 205 nm will quantify peptides by measuring at level of peptide bonds, rather than by inclusion of aromatic rings (measured at 280 nm). To customize the detection method, visit: <http://www.nanodrop.com/Library/A205%20Proteins%20&%20Peptides%20Custom%20Method.pdf>.
9. Peptide quantity to be loaded on the LC column is dependent on the type of cHiPLC columns being used. A variety of cHiPLC columns with different lengths, diameters, pore size, and resin are available, but will require additional optimization.
10. We recommend using the same cHiPLC columns for the duration of the project. Multiple columns are available and differ in

pore size and column length. Use of alternative columns will require additional optimization (also *see* **Note 9**).

11. Although Table 1 gives a suggested LC protocol (with a 180 min gradient from 5 to 35 % acetonitrile), optimization is necessary, and depending on the experimental setup, the gradient may be shortened or lengthen as necessary. However, it is important that all DIA and DDA for an experiment are performed with the same gradient.
12. When performing a SWATH-MS experiment, generation of the spectral library in DDA mode is important. Three methods are available to generate a library: (1) use a preconstructed library, (2) generate a library from experimental samples, or (3) generate a library from a variety of cell lines or other suitable samples containing proteins covering the range of those found in the experimental samples. Each method has benefits, but if limits on the sample availability exist the generation of a library through cell lines is an enticing option. We have previously performed such an analysis [16]. Until issues of alignment of elution times are resolved, we do not recommend using a preconstructed library.
13. At this step, we separate our samples into library or SWATH-MS runs and randomize the samples within each group. In this way, the mass spectrometry methods will only be changed once when transitioning from samples used for the library to SWATH-MS (or vice versa).
14. In our experience, peptides will begin eluting at approximately 30 min and maximum intensity readings occur between 70 and 110 min. However, elution times will change based on the elution gradient and will be influenced by sample composition. It can also be influenced by type of resin used for the reverse phase HPLC.
15. While some differences are expected between samples, TICs that are markedly dissimilar may indicate impurities in the sample, problems with the cHiPLC system, or issues with the mass spectrometry methods. We recommend testing all the procedures using a comparable sample to the experimental samples, but can be discarded in the event of procedural shortcomings.
16. Because of the processor intensive demands for searching and for targeted data analysis, we recommend using a computer distinct from the computer that is loaded with Analyst and operates the mass spectrometer.
17. The SwissProt section of UniProtKB is a high quality, manually curated database of protein sequences that eliminates redundancy. In contrast, the TrEMBL section contains computationally analyzed records that are obtained from the translation of annotated coding sequences of the EMBL-bank/

GenBank/DDBJ nucleotide databases. Although TrEMBL contains more information, this section of UniProtKB is limited in experimental validation of sequences and the use of TrEMBL sequences during database searching may increase the risk of inappropriate spectral assignment during targeted data extraction of SWATH-MS files.

18. Artificial peptides are added to the SWATH library in the same fashion as contaminants (*see* Data Processing: Generating the Spectral Library, **step 2**).
19. FASTA files for artificial peptides listed in the Materials and Equipment can be found on the manufacturer's website. http://www.biognosys.ch/fileadmin/Uploads/iRT/iRT_Peptides_Fusion.FASTA. An alternative to using spiked-in peptide standards to correct for retention time drift is to select peptide(s) from abundant proteins (actin, keratin, etc.) in the PeakView software during data analysis.
20. Although proteins with a lower confidence are being imported to PeakView for targeted data extraction, lower confidence proteins will be filtered out during targeted data extraction within the PeakView software (FDR, Score) and can also be filtered manually after export (e.g., filtered by number or peptides per protein ≥ 2).
21. When selecting spiked-in or endogenous peptides for RT correction, be sure to select only those peptides with high intensity readings, overlapping transition states, those that are free of background noise, and cumulatively have adequate coverage across the entire elution gradient.
22. Low confidence assignment of spectra for a number of given samples *may* impact the quality of the export for all samples if exported in unison. For this reason, perform targeted data extraction only for samples that will be used for direct comparison to one another. The processing settings will need to be optimized for each individual experiment with a particular emphasis on the extraction window. Using the RT correction tool will likely improve RT variability between samples and allow for narrowing of the extraction window. Additionally, setting a stringent FDR threshold (e.g., 1 %) will improve the quality of the exported data (as assessed by manual review of the overlay of transition states in the XIC pane and the alignment of the SWATH spectra to the library spectra in the spectra pane). It should be noted, however, that this stringency comes with the cost of decreasing the total number of proteins exported and will likely impact the number of peptides used for quantification. Also discussed in [6].
23. Four normalization methods are available in MarkerView: (1) Selected peak, (2) Total peak intensity, (3) median peak intensity,

and (4) manual scale factor. Additional information on the normalization can be obtained in the MarkerView program.

24. In some cases, such as working with supernatants or whole cell lysates from cell lines, the means and standard deviations between replicates and conditions will be comparable and as such, no further transformation is necessary. In these cases, it is suggested to perform statistical testing using the *t*-test (for two comparisons) or a variation of the ANOVA for multiple comparisons. However, when working with biological fluids obtained from primary donors, it should be expected that the mean intensity and standard deviations between donors is not comparable, and as such, the *z*-transformation can be applied in order to allow for parametric statistical testing.
25. If desired, multiple comparisons corrections can be employed, but it may limit the robustness and utility of continued analyses, including bioinformatic analysis.
26. “Event” describes the programming that is used to direct the sample path within the trap and column and is provided for reference. For additional information on this subject, please contact Eksigent.

References

1. Pernemalm M, Lehtio J (2014) Mass spectrometry-based plasma proteomics: state of the art and future outlook. *Expert Rev Proteomics* 11(4):431–448
2. Anderson L (2014) Six decades searching for meaning in the proteome. *J Proteomics* 107C:24–30
3. Burkhard PR, Rodrigo N, May D, Sztajzel R, Sanchez JC, Hochstrasser DF, Schiffer E, Reverdin A, Lacroix JS (2001) Assessing cerebrospinal fluid rhinorrhea: a two-dimensional electrophoresis approach. *Electrophoresis* 22(9):1826–1833
4. Vejda S, Posovszky C, Zelzer S, Peter B, Bayer E, Gelbmann D, Schulte-Hermann R, Gerner C (2002) Plasma from cancer patients featuring a characteristic protein composition mediates protection against apoptosis. *Mol Cell Proteomics* 1(5):387–393
5. DeSouza LV, Siu KW (2013) Mass spectrometry-based quantification. *Clin Biochem* 46(6):421–431
6. Gillet LC, Navarro P, Tate S, Rost H, Selevsek N, Reiter L, Bonner R, Aebersold R (2012) 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* 11(6):O111.016717
7. Wisniewski JR, Zougman A, Nagaraj N, Mann M (2009) Universal sample preparation method for proteome analysis. *Nat Methods* 6(5):359–362
8. Scopes RK (1974) Measurement of protein by spectrophotometry at 205 nm. *Anal Biochem* 59(1):277–282
9. Baldi P, Long AD (2001) A Bayesian framework for the analysis of microarray expression data: regularized *t*-test and statistical inferences of gene changes. *Bioinformatics* 17(6):509–519
10. Kayala MA, Baldi P (2012) Cyber-T web server: differential analysis of high-throughput data. *Nucleic Acids Res* 40(Web Server issue):W553–W559
11. Cheadle C, Vawter MP, Freed WJ, Becker KG (2003) Analysis of microarray data using *Z* score transformation. *J Mol Diagn* 5(2):73–81
12. Haverland NA, Fox HS, Ciborowski P (2014) Quantitative proteomics by SWATH-MS reveals altered expression of nucleic acid binding and regulatory proteins in HIV-1-infected macrophages. *J Proteome Res* 13(4):2109–2119

13. Taylor LE, Swan T, Mayer KH (2012) HIV coinfection with hepatitis C virus: evolving epidemiology and treatment paradigms. *Clin Infect Dis* 55(Suppl 1):S33–S42
14. Deeks SG, Lewin SR, Havlir DV (2013) The end of AIDS: HIV infection as a chronic disease. *Lancet* 382(9903):1525–1533
15. Krebs FC, Miller SR, Malamud D, Howett MK, Wigdahl B (1999) Inactivation of human immunodeficiency virus type 1 by nonoxynol-9, C31G, or an alkyl sulfate, sodium dodecyl sulfate. *Antiviral Res* 43(3):157–173
16. Villeneuve LM, Stauch KL, Fox HS (2014) Proteomic analysis of the mitochondria from embryonic and postnatal rat brains reveals response to developmental changes in energy demands. *J Proteomics* 109C: 228–239