Proteomics of Cerebrospinal Fluid: Throughput and Robustness Using a Scalable Automated Analysis Pipeline for Biomarker Discovery

ARTICLE *in* ANALYTICAL CHEMISTRY · OCTOBER 2015

Impact Factor: 5.64 · DOI: 10.1021/acs.analchem.5b02748

3 AUTHORS, INCLUDING:

Martin Kussmann

Nestlé Institute of Health Sciences S.A.

104 PUBLICATIONS **2,187** CITATIONS

SEE PROFILE

Loïc Dayon

Nestlé Institute of Health Sciences S.A.

41 PUBLICATIONS **1,113** CITATIONS

SEE PROFILE

analytical chemistry

Article

pubs.acs.org/ac

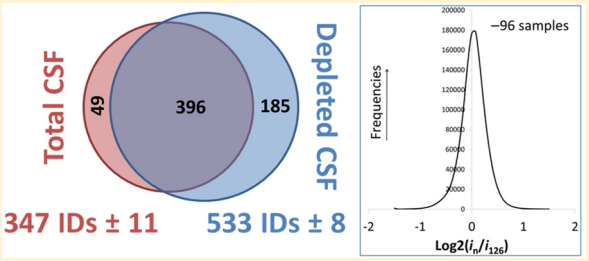
Proteomics of Cerebrospinal Fluid: Throughput and Robustness Using a Scalable Automated Analysis Pipeline for Biomarker Discovery

Antonio Núñez Galindo, Martin Kussmann, and Loïc Dayon*

Molecular Biomarkers Core, Nestlé Institute of Health Sciences SA, EPFL Innovation Park, Bâtiment H, Lausanne, Switzerland

Supporting Information

ABSTRACT: Cerebrospinal fluid (CSF) is a body fluid of high clinical relevance and an important source of potential biomarkers for brain-associated damages, such as traumatic brain injury and stroke, and for brain diseases, such as Alzheimer's and Parkinson's. Herein, we have implemented, evaluated, and validated a scalable automated proteomic pipeline (ASAP^2) for the sample preparation and proteomic analysis of CSF, enabling increased throughput and robustness for biomarker discovery. Human CSF samples were depleted from abundant proteins and subjected to automated reduction, alkylation, protein digestion, tandem mass tag (TMT) 6-plex labeling, pooling, and sample cleanup in a 96-well-plate format before reversed-phase liquid chromatography tandem mass spectrometry (RP-LC MS/MS). We showed the impact on the CSF proteome coverage of applying the depletion of abundant proteins, which is usually performed on blood plasma or serum samples. Using ASAP^2 to analyze 96 identical CSF samples, we determined the analytical figures of merit of our shotgun proteomic approach regarding proteome coverage consistency (i.e., 387 proteins), quantitative accuracy, and individual protein variability. We demonstrated that, as for human plasma samples, ASAP^2 is efficient in analyzing large numbers of human CSF samples and is a valuable tool for biomarker discovery. The data has been deposited to the ProteomeXchange with identifier PXD003024.



Accuracy, reproducibility, throughput, and robustness are compulsory for proteomic technologies to be applied to clinical research. To date, most biomarker discovery studies performed with mass spectrometry (MS)-based proteomics have been severely limited, in terms of sample size (i.e., number of subjects and samples encompassed by the cohort and the analysis), which has often compromised the study design. This limitation mostly originates from technological constraints and directly impacts the strength of the biomarker candidate findings and their translation into the clinics.[1,2] Therefore, we have recently developed a scalable automated proteomic pipeline (ASAP^2) for biomarker discovery to handle large numbers of human plasma samples with clinically relevant throughput (i.e., 192 samples within 3 weeks; 384 samples within 4 weeks; 576 samples within 5 weeks, ... from sample reception in the laboratory to MS raw data file generation).[1] This proteomic pipeline can be adapted and applied to other sample types (in particular, other bodily fluids).

Cerebrospinal fluid (CSF) is a bodily fluid present around the brain and in the spinal cord. It acts as a protective cushion against shocks and participates in the immune response in the brain. CSF normally contains proteins at concentrations between 150 and 600 mg L^{-1} in human subjects. Analysis of total CSF protein can be used for diagnostic purposes, as, for instance, a sign of a tumor, bleeding, inflammation, or injury.[3,4] In Alzheimer's disease (AD), specific peptides and proteins (Aβ1−42, total CSF tau protein, and CSF phosphorylated tau$_{181P}$)[5] have been proposed as criteria for diagnosis of the dementia.[6] Considering the high value of CSF as a source of potential biomarkers for brain-associated damages and pathologies, the development of robust automated platform for CSF proteomics is of great value.

The CSF proteome has been extensively studied using a variety of analytical techniques.[7,8] In particular, isobaric tagging was used to study relative protein changes in CSF related to neurodegenerative disorders[9,10] and brain injury.[11] Several thousands of proteins are identified in CSF when using prefractionation techniques.[12,13] Unfortunately, such separation precludes the analysis of large numbers of samples, because of time constraints and, in some cases, reproducibility issues. One-dimensional reversal-phase liquid chromatography tandem mass spectrometry (RP-LC MS/MS) can typically reveal a few hundred proteins in CSF[14] and is recommended to warrant sufficient throughput compatible with clinical research and proteomic workflows such as ASAP^2.

ASAP^2 was initially developed with the purpose of (i) discovering protein biomarkers in plasma, i.e., the sample type most routinely used in the clinics, and (ii) analyzing cohorts with hundreds to thousands of subjects and samples.[1,15] ASAP^2 is composed of analytical blocks that can be used as optional or

Received: April 9, 2015
Accepted: October 9, 2015

DOI: 10.1021/acs.analchem.5b02748
Anal. Chem. XXXX, XXX, XXX−XXX

mandatory steps. As a first step, abundant-protein immuno-affinity depletion is performed with antibody-based columns and LC systems equipped with a refrigerated autosampler and fraction collector. This procedure is the rate-limiting step of the workflow and requires 4 days to be sequentially completed for 96 samples using one LC apparatus. This block is linked to and followed by buffer exchange performed in a 96-well plate format by manual operations that require <1 h to be completed (excluding time for sample evaporation; see the Experimental Section). The rest of the process is fully automated and includes (i) reduction, alkylation, enzymatic digestion (performed overnight); (ii) tandem mass tag (TMT) labeling and pooling (processing time of 110 min for 96 samples); (iii) RP solid-phase extraction (SPE) purification (processing time of 100 min); and (iv) strong cation-exchange (SCX) SPE purification (processing time of 70 min). Basically, a liquid handling platform is used to dispense required volumes of reagents with a 4-channel pipetting arm, which acts also as grippers to move plates on the deck. Block units that agitate, and/or heat the 96-well plate are mounted. A dark chamber is installed for the alkylation step. The liquid handler also aspirates samples and dispenses them into new laboratory wares (e.g., sample pooling after TMT labeling and sample loading onto SPE cartridges). A vacuum manifold is mounted onto the platform to allow for fully automated SPE of up to 96 samples, with reservoirs for solvent storage. Finally, RP-LC MS/MS is performed with instrumentation equipped with autosamplers.

Herein, we report on the implementation, evaluation, and validation of ASAP^2 for sample preparation and proteomic analysis of human CSF samples in a 96-well plate. CSF samples were first depleted from abundant proteins by multiplexed immuno-affinity. Subsequently, reduction, alkylation, protein digestion, TMT 6-plex labeling,[11] pooling, and sample cleanup were performed in a 96-well-plate format using a liquid-handling robotic platform. Samples were analyzed with RP-LC MS/MS. We first evaluated the significance of the depletion of 14 highly abundant human plasma proteins on human CSF samples by assessment of proteome coverage and biological relevance of the identified proteins. To analytically characterize and finally validate ASAP^2 for CSF applications, 96 identical CSF samples were prepared using the highly automated ASAP^2 procedure. Figures of merit, with regard to proteome coverage consistency, quantitative precision and trueness, and individual protein variability, were determined. Finally, we demonstrated the ability of ASAP^2 to quantify differentially concentrated proteins by performing a calibration curve with a two-proteome model experiment.

EXPERIMENTAL SECTION

Materials. A commercial pool of normal CSF samples (70 mL) was purchased from Analytical Biological Services (Wilmington, DE). The samples constituting this pool were remnant samples collected from individuals for various testing. Each donor gave informed consent. The local institutional ethical committee board approved the clinical protocol. Iodoacetamide (IAA), tris(2-carboxyethyl) phosphine hydrochloride (TCEP), triethylammonium hydrogen carbonate buffer (TEAB) (1 M, pH 8.5), sodium dodecyl sulfate (SDS), and β-lactoglobulin (LACB) from bovine milk were purchased from Sigma (St. Louis, MO). Formic acid (FA, 99%) and acetonitrile were from BDH (VWR International, Ltd., Poole, U.K.). Hydroxylamine solution 50 wt % in H_2O (99.999%) was acquired from Aldrich (Milwaukee, WI). Water (18.2 MΩ cm at 25 °C) was obtained from a Milli-Q apparatus (Millipore, Billerica, MA). Trifluoroacetic acid (Uvasol) was sourced from Merck Millipore (Billerica, MA). The 6-plex TMTs were purchased from Thermo Scientific (Rockford, IL). Sequencing-grade modified trypsin/Lys-C was procured from Promega (Madison, WI). For immuno-affinity depletion of 14 highly abundant proteins from human biological fluids, multiple affinity removal system (MARS) columns, Buffer A, and Buffer B were obtained from Agilent Technologies (Wilmington, DE). The BCA protein assay was purchased from Thermo Scientific. Oasis HLB cartridges (1 cm^3, 30 mg) were acquired from Waters (Milford, MA) and SCX cartridges from Phenomenex (Torrance, CA).

E. coli protein-lyophilized sample (2.7 mg, Bio-Rad, Hercules, CA) was dissolved in H_2O and precipitated with prechilled acetone (−20 °C) overnight. After centrifugation at 8000*g* for 10 min (4 °C), the acetone was decanted and the pellets were dried for 10 min.

Sample Preparation. Aliquots of 400 μL of the commercial pooled CSF sample were evaporated with a vacuum centrifuge (Thermo Scientific). The dried samples were diluted in 125 μL of Buffer A containing 0.00965 mg mL^{-1} LACB. A volume of 120 μL was filtered with 0.22 μm filter plate from Millipore. Immuno-affinity depletion was performed as followed. Abundant plasma proteins were removed from the filtered CSF sample solutions (100 μL loaded on column), following the manufacturer instructions with slight modifications, using MARS columns and high-performance liquid chromatography (HPLC) systems (Thermo Scientific, San Jose, CA) equipped with an HTC-PAL (CTC Analytics AG, Zwingen, Switzerland) fraction collector. After immuno-depletion, samples were snap-freezed and stored. Buffer exchange was performed with Strata-X 33u Polymeric RP (30 mg/1 mL) cartridges mounted on a 96-hole holder and a vacuum manifold, all from Phenomenex as previously described.[1] Samples were subsequently evaporated and stored at −80 °C.

Nondepleted CSF samples were prepared as followed. Volumes of 100 μL CSF were evaporated with a vacuum centrifuge. The dried samples were diluted in 125 μL of H_2O containing 0.00965 mg mL^{-1} LACB. A volume of 120 μL was filtered with 0.22 μm filter plate from Millipore. The samples were evaporated again and stored at −80 °C before further use.

Dried samples (depleted or nondepleted) were subjected to reduction, alkylation, digestion, TMT 6-plex labeling, pooling and SPE sample purification (Oasis HLB and SCX) using a 4-channels Microlab Star liquid handler (Hamilton, Bonaduz, Switzerland) in a 96-well-plate format and according to a previously reported protocol.[1] Briefly, each sample was dissolved in 95 μL of TEAB 100 mM and 5 μL of 2% SDS. A volume of 5.3 μL of TCEP (20 mM) was added and incubation was performed for 1 h at 55 °C. A volume of 5.5 μL IAA 150 mM was added (incubation for 1 h in darkness). Enzymatic digestion was performed via the addition of 10 μL of trypsin/Lys-C at 0.25 μg μL^{-1} in 100 mM TEAB (incubation overnight at 37 °C). TMT labeling was performed via the addition of 0.8 mg of TMT 6-plex reagent in 41 μL of CH_3CN (incubation for 1 h at room temperature). After reaction, a volume of 8 μL of hydroxylamine 5% in H_2O was added to each tube to react for 15 min. Samples from a given TMT 6-plex experiment were pooled together in a new tube. Pooled samples were further purified with Oasis HLB, followed by

DOI: 10.1021/acs.analchem.5b02748
Anal. Chem. XXXX, XXX, XXX−XXX

SCX SPE. The purified pooled TMT-labeled samples were then evaporated to dryness before storage at −80 °C.

The same procedure was followed to evaluate the ability of ASAP^2 to quantify differentially concentrated proteins in CSF using a two-proteome model experiment. To perform a calibration curve experiment, a volume of 6.2 μL protein extracts of *E. coli* (see the Materials Section), dissolved in H_2O at 0, 0.33, 0.49, 0.82, 1.02, and 1.23 μg/μL, respectively, was spiked in 400 μL of human CSF samples (i.e., *E. coli* extract spiked amounts of 0, 2, 3, 5, 6.25, and 7.5 μg, respectively). Such sets of 6 spiked CSF samples were processed using ASAP^2 in triplicate.

RP-LC MS/MS. The samples were dissolved in 200 μL H_2O/CH_3CN/FA 96.9/3/0.1 for RP-LC MS/MS (750 μL H_2O/CH_3CN/FA 96.9/3/0.1 for nondepleted CSF). LC MS/MS was performed with a hybrid linear ion trap-Orbitrap (LTQ-OT) Elite and an Ultimate 3000 RSLC nano system (Thermo Scientific). Proteolytic peptides (injection of 5 μL of sample) were trapped on an Acclaim PepMap 75 μm × 2 cm (C18, 3 μm, 100 Å) precolumn and separated on an Acclaim PepMap RSLC 75 μm × 50 cm (C18, 2 μm, 100 Å) column (Thermo Scientific) coupled to a stainless steel nanobore emitter (40 mm, OD 1/32 in.) mounted on a Nanospray Flex Ion Source (Thermo Scientific).The analytical separation was run for 150 min using a gradient that reached 30% of CH_3CN after 140 min and 80% of CH_3CN after 150 min at a flow rate of 220 nL min^{-1}. For MS survey scans, the OT resolution was 120 000 (ion population of 1×10^6) with an m/z window from 300 to 1500. For MS/MS with higher-energy collisional dissociation (HCD) at 35% of the normalized collision energy, ion population was set to 1×10^5 (isolation width of 2), with a resolution of 15 000, first mass at m/z = 100, and a maximum injection time of 250 ms in the OT. A maximum of 10 precursor ions (most intense) were selected for MS/MS. Dynamic exclusion was set for 60 s within a ±5 ppm window. A lock mass of m/z = 445.1200 was used. Each sample was analyzed in triplicate.

Data Processing and Analysis. Proteome Discoverer (version 1.4, Thermo Scientific) was used as data processing interface. Identification was performed against the human UniProtKB/Swiss-Prot database (08/12/2014 release) including the LACB sequence (20194 sequences in total). Mascot[16] (version 2.4.2, Matrix Sciences, London, U.K.) was used as the search engine. Variable amino acid modifications were oxidized methionine, deamidated asparagine/glutamine, and 6-plex TMT-labeled peptide amino terminus (+229.163 Da); 6-plex TMT-labeled lysine (+229.163 Da) was set as fixed modifications as well as carbamidomethylation of cysteine. Trypsin was selected as the proteolytic enzyme, with a maximum of two potential missed cleavages. Peptide and fragment ion tolerance were set to, respectively, 10 ppm and 0.02 Da.

For the two-proteome model experiment, an additional search was performed with Mascot in the UniProtKB/SwissProt database restrained to the *E. coli* taxonomy. Other parameters remained unchanged.

All Mascot result files were loaded into Scaffold Q+S 4.3.2 (Proteome Software, Portland, OR) to be further searched with X! Tandem. Both peptide and protein FDRs were fixed at 1% maximum, with a two-unique-peptide criterion to report protein identification. Quantitative values were exported from Scaffold either as raw spectral data or as log2 of the protein ratio fold changes (i.e., log2 of the means of peptide ratios with isotopic purity correction but without normalization applied between samples and experiments). The intensities of the TMT reporter-ions at m/z = 126 were used as the denominators for ratio calculations. The MS proteomic data has been deposited to the ProteomeXchange Consortium[17] (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository[18] with the dataset identifier PXD003024.

Calculation and statistics were performed with Excel 2010 (Microsoft, Redmond, WA) and GraphPad Prism 5 (GraphPad Software, San Diego, CA). The DAVID[19,20] software was used for functional annotation and GO term enrichment.

RESULTS AND DISCUSSION

Our previously reported highly automated workflow used two immuno-depletion systems, one liquid handler, and one mass spectrometer and had the capacity to analyze 2 × 96 samples within 3 weeks.[1,15] This workflow is certainly scalable by multiplication and multiplexing of the same type of resources, both instrumental and human, and costs. It was employed herein without significant modification, apart from initial sample handling of the CSF samples (e.g., volume used, sample evaporation requirement before abundant protein depletion, loading amount for RP-LC MS/MS; see the Experimental Section). Abundant protein depletion constitutes the rate-limiting step as each sample is treated one after the other; it is fully supported by the LC system. The buffer exchange step is still performed by an operator but parallelized in a 96-well-plate format using multichannel pipettes. Then, the robotic platform processes the 96 samples that are subjected to reduction, alkylation, digestion, TMT labeling, pooling, and purification; manual operations consist only of providing the required reagents. RP-LC MS/MS is then performed using an analysis time of 3 h per TMT 6-plex experiment. The relative quantification based on TMT technology offers multiplexing that significantly reduces the overall analysis time.

Depletion of 14 Abundant Plasma Proteins in CSF Samples: Qualitative Assessment. Abundant protein depletion is commonly used when analyzing human plasma samples and specific immuno-affinity material has been developed for this purpose. We first evaluated the use of current commercial immuno-affinity LC column (see the Experimental Section) on human CSF samples and its impact on the ASAP^2 approach. Despite the different protein distributions in human plasma and CSF, several proteins are highly abundant in both body fluids (e.g., albumin, being by far the most abundant protein in both bodily fluids, IgG, transferrin, alpha-1-antitrypsin, ...),[21] masking the detection of less-abundant ones. The overall proportion of low-abundance proteins was reported to be higher (17%) in CSF than in plasma.[21] In addition, the pros and cons of depleting CSF with plasma depletion products had been previously addressed, yet without clear conclusion on universal adoption of either one of the workflows.[22−25]

Based on the commercial CSF sample used herein, we found that ∼93% of the total protein content was removed following immuno-affinity depletion of 14 proteins (see the Experimental Section), going from 310 ± 40 μg loaded on column (i.e., in 400 μL of CSF) to 22 ± 5 μg measured in the collected fractions. This result confirmed the overlap in terms of high abundance of some of the 14 targeted proteins between plasma and CSF (i.e., serum albumin, IgG, alpha-1-antitrypsin, IgA, serotransferrin, haptoglobin, fibrinogen, alpha-2-macroglobulin, alpha-1-acid glycoproteins, IgM, apolipoprotein A-I, apolipo-

DOI: 10.1021/acs.analchem.5b02748
Anal. Chem. XXXX, XXX, XXX−XXX

protein A-II, complement C3, and transthyretin).[21] Buffer exchange was applied to depleted samples for compatibility with downstream proteomic sample preparation.

Depleted and nondepleted samples initially originating from the same commercial CSF sample pool were prepared for RP-LC MS/MS analysis, following the ASAP^2 procedure. The first automated step, including sample solubilization, reduction of disulfide bridges, alkylation of thiol moieties, and digestion of proteins with trypsin/Lys-C, was performed on 18 previously and independently depleted samples and 18 nondepleted samples. The second robotic step consisted in 6-plex TMT labeling of the samples, followed by pooling. The last automated steps consisted of sample purification (see the Experimental Section).

With RP-LC MS/MS, we analyzed, in triplicate, the three TMT 6-plex experiments originating from depleted CSF samples, and the three TMT 6-plex experiments originating from nondepleted (i.e., total) CSF samples. In total, we identified 630 proteins in human CSF (Figure 1a). As shown in

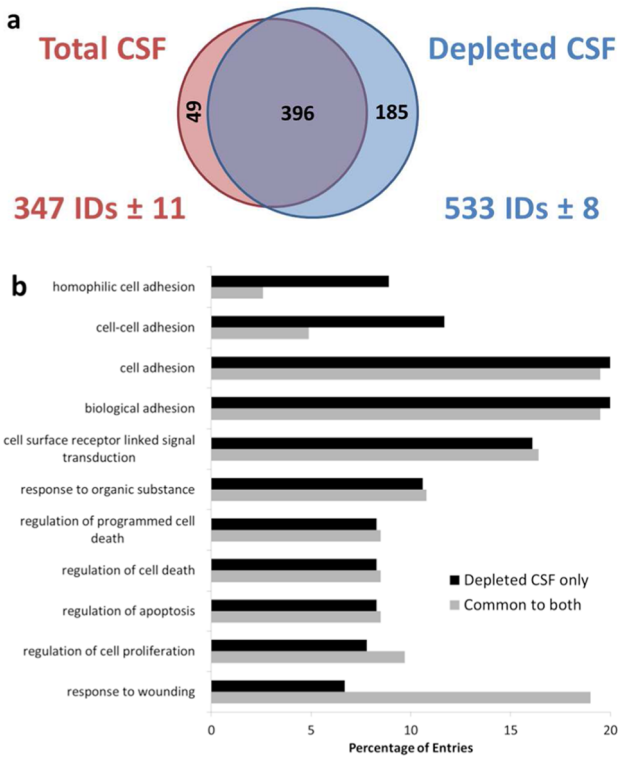


Figure 1. (a) Proteome coverages obtained with and without inclusion of the immuno-affinity depletion step in ASAP^2. Both total and averaged numbers of identified proteins from respectively all (see the Venn diagram) and per replicate(s) (see protein identification "IDs") are displayed. (b) Gene ontology/biological processes enriched in the list of proteins identified only in depleted CSF (i.e., 185 proteins), with respect to their percentage of representation in the list of proteins identified commonly in both total nondepleted CSF and depleted CSF (i.e., 396 proteins).

Figure 1a, 63% of the proteins were overall commonly identified in depleted CSF and total CSF. Depletion increased proteome coverage by 31% overall, which is an improvement that, in our opinion, justifies its incorporation in the workflow. Among the 49 proteins only identified in the total nondepleted CSF samples (see the Supporting Information), 7 proteins were direct targets of the immuno-affinity depletion (i.e., alpha-1-antitrypsin, serotransferrin, haptoglobin, alpha-2-macroglobulin, alpha-1-acid glycoproteins (1 and 2), apolipoprotein A-I), and 31 proteins were immunoglobulins. The presence of 11 other unexpected proteins may have resulted from unspecific binding or co-removal with the targeted proteins, as previously shown in plasma.[26] Therefore, protein removal with depletion is to be cautiously applied, particularly with regard to targeted workflows.[27] In addition and based on spectral counting, serum albumin and complement C3 were very efficiently depleted (0.5% and 2.7% of the spectra remaining, respectively) and transthyretin to some extent (39% of spectra remaining; see the Supporting Information). Based on the same semiquantitative spectral counting criteria, we found that fibrinogen and apolipoprotein A-II were also still present in the depleted CSF samples (see the Supporting Information). Using the same immuno-affinity depletion system as we did, Liebler and co-workers demonstrated the same partial removal for those proteins in human plasma.[28] Overall, the abundant-protein immuno-affinity depletion technique, initially developed and commonly used for human plasma and serum samples, turned out to be also efficient when applied to human CSF samples. Performing depletion required a volume that was four times greater (i.e., 400 μL). This sample requirement must be taken into account when designing clinical studies.

Finally, we investigated the relevance of the depletion step with regard to the proteomic coverage of biological processes, accessible in addition to those delivered by total CSF analysis (Figure 1b). The cell−cell adhesion and homophilic cell adhesion subsets were clearly over-represented in the depleted CSF samples, as revealed by detection of several cadherins. This is an important observation because cell adhesion molecules—and, in particular, cadherins and protocadherins—have been proposed to be involved in the pathophysiology of several neurological diseases.[29−31] Those qualitative observations argued in favor of the depletion of abundant proteins. The impact on the quantitative readouts still needs to be carefully evaluated.

Depletion of 14 Abundant Plasma Proteins in CSF Samples: Quantitative Assessment. We then used the quantitative data obtained from the TMT isobaric labeling technology to assess the quantitative performance of ASAP^2 with and without immuno-affinity depletion. According to the design of the experiments, all quantitative measurements were expected to provide 1:1 protein ratios. The three TMT 6-plex experiments with depleted CSF samples and the three TMT 6-plex experiments with total CSF samples were compared, in terms of quantitative performance. The distribution of the log2 of the ratio fold changes of the matched raw spectrum data obtained with MS/MS showed that the accuracy of quantitative measurements decreased when depletion (including buffer exchange) was added to the sample preparation workflow (see Figure 2 and Table 1, and Figures S1 and S2 in the Supporting Information), as previously shown with plasma samples, and as expected, because of the addition of technical steps.[1] It was recently shown that the depletion step induces underestimation of the endogenous CSF protein concentrations with multiple-reaction monitoring MS.[24] The relative quantification strategy used in ASAP^2 should not suffer from such an underestimation as long as the workflow ensures reproducibility. Standard deviation (SD) values were 0.1856 and 0.1279, for the log2 of ratio fold changes corresponding to theoretical ratios of 1, for depleted CSF and total nondepleted CSF samples, respectively (see Table 1). We concluded that this difference in quantitative

DOI: 10.1021/acs.analchem.5b02748
Anal. Chem. XXXX, XXX, XXX−XXX

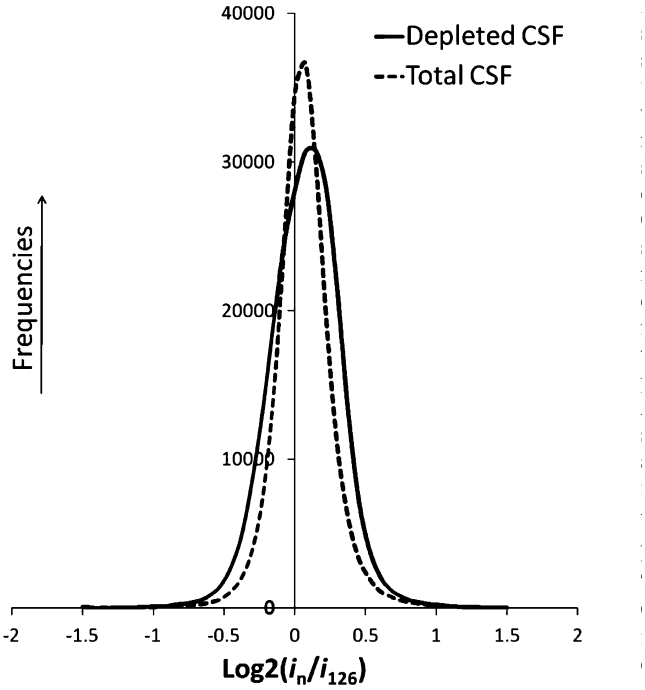


Figure 2. Distribution of log2(i_n/i_{126}) obtained from raw spectral data matched to a peptide sequence with and without the immuno-affinity depletion step (i.e., depleted and total nondepleted CSF samples, respectively). Each 6-plex TMT experiment was performed in triplicate and each was analyzed with RP-LC MS/MS in triplicate.

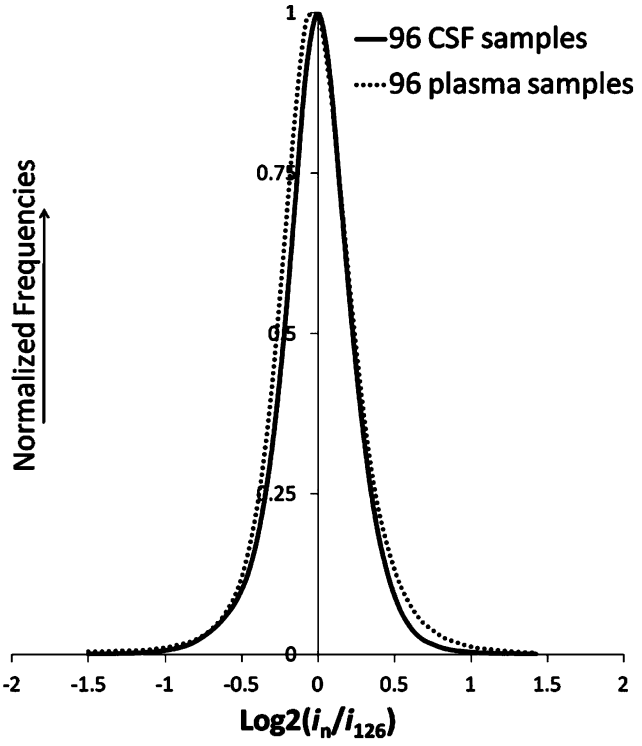


Figure 3. Normalized distribution of log2(i_n/i_{126}) obtained from matched raw spectral data. In total, 96 identical human CSF samples were individually prepared using the highly automated MS-based proteomic workflow, ASAP^2, including individual immuno-affinity depletion of each sample as a first step (see the Experimental Section). RP-LC MS/MS analysis was performed in triplicate. Previously published distribution for 96 identical human plasma samples is displayed.[1]

performance does not preclude the implementation of immuno-affinity depletion to the workflow. Indeed, these results confirmed that the depletion procedure for human CSF samples, as applied in ASAP^2, is quantitatively robust and consistent,[28] in addition of showing substantial value in terms of both proteome coverage and biological readouts. Therefore, immuno-affinity depletion was implemented in the ASAP^2 procedure for CSF analysis.

Validation of ASAP^2 with 96 Identical Human CSF Samples. The integrated ASAP^2 workflow was applied to 96 identical CSF pools. Each such CSF sample was prepared individually in a 96-well-plate format, including initial immuno-affinity depletion, buffer exchange, reduction, alkylation, protein digestion, TMT 6-plex labeling, pooling, and cleanup. Relative quantification using TMT 6-plex delivered the normalized distribution of the log2 of the ratio fold changes of the matched raw spectrum data, as displayed in Figure 3. The analytical figures of merit for CSF samples appeared very similar to those previously obtained with plasma samples.[1] As shown in Figure 3, the CSF and plasma-derived distributions of log2(i_n/i_{126}) obtained from matched raw spectral data are almost superimposed. We determined that 95% of the raw quantitative spectral ratio data were comprised between 0.69 and 1.45. A coefficient of variation of 15% was found for these raw data.

In total, 387 proteins were consistently quantified across all 96 CSF samples (see Figure S3 in the Supporting Information) (including LACB, from 726 identified proteins in total; see Supporting Information). On this subdata set without any missing value, an averaged SD of 0.2080 per protein was estimated for the log2 of the protein ratio fold changes, ranging from 0.1405 to 0.8852 for individual proteins (see the Supporting Information). It is very consistent with the global SD of 0.2174 reported in Table 1. The 95% confidence CI of the mean was 0.996−0.998 for ratio fold changes correspond-

Table 1. Descriptive Statistics of the Datasets (i.e., Reported for Each Workflow Evaluation and on the Proteins Consistently Quantified Across All Replicates)[a]

expt	short description	expected ratio fold change	expected log2 (ratio fold change)	median	mean	standard deviation, SD	lower 95% CI[b] of mean	upper 95% CI[b] of mean	kurtosis
1	ASAP^2 Evaluation Excluding Depletion	1	0	−0.0902	−0.07996	0.1279	−0.08243	−0.07749	133.1
2	ASAP^2 Evaluation Including Depletion	1	0	−0.1827	−0.1852	0.1856	−0.1888	−0.1816	0.1023
3	ASAP^2 Validation with 96 Samples	1	0	0.0021	−0.004654	0.2174	−0.006052	−0.003257	6.221

[a]The details of the experiments are given in the text. In all cases, quantitative ratios (e.g., fold changes) were calculated as i_n/i_{126}. [b]Confidence interval.

DOI: 10.1021/acs.analchem.5b02748
Anal. Chem. XXXX, XXX, XXX−XXX

ing to theoretical ratios of 1. For the 387 proteins, 95% of the measured protein ratio fold changes (considered in all experiments and replicates) were comprised between 0.74 and 1.35 (0.75 and 1.34 when considering the averages of the triplicate measurements per sample). When artificially taking the protein averages of all experiments and replicates, those boundaries were delimitated between 0.96 and 1.04. Taken together, these results showed that a low number of measurement replicates can already suffice for biomarker discovery and that our ASAP^2 workflow can be applied to the measurements of human CSF samples in clinical research.

From the Supporting Information, we extracted the 25 proteins that showed the lowest variability when analyzing the 96 identical CSF pools with ASAP^2 (see Figure S4 in the Supporting Information). This protein set overlapped very little with the previously found 25 plasma proteins that showed the lowest variability when analyzed with our highly automated pipeline.[1] Only leucine-rich alpha-2-glycoprotein and hemopexin were present on both lists of proteins that show the lowest variability when analyzed with ASAP^2.

Ability to Quantify Differentially Concentrated Proteins in CSF with ASAP^2. Lastly, we performed a calibration experiment using a two-proteome model design, by spiking increasing amounts of an *E. coli* protein extract (0:0.4:0.6:1:1.25:1.5 ratios) into identical pooled human CSF samples (1:1:1:1:1:1 ratios) (see Figure 4). In those samples,

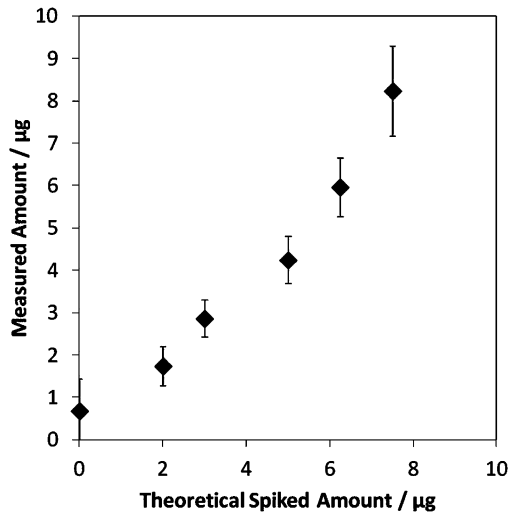


Figure 4. Averaged quantitative performances obtained with a two-proteome model using ASAP^2. *E. coli* total protein extract was spiked at 0, 2, 3, 5, 6.25, and 7.5 μg in 400 μL of human CSF, i.e., final concentration of 0, 0.0050, 0.0075, 0.0125, 0.0156, and 0.0188 μg/μL, respectively. A calibration curve was obtained for the *E. coli* proteins ($y = 0.9774x + 0.0896$; $R^2 = 0.9563$). TMT 6-plex experiments were performed three times and each was analyzed in triplicate by RP-LC MS/MS. The measured amounts were recalculated from the measured relative abundances obtained with MS and the total spiked amount.

we consistently quantified 350 human proteins and 95 *E. coli* proteins across experiments and replicates. Human CSF proteins were quantitatively stable (average ratio of 1.09; coefficient of variation of 20%). A calibration curve was obtained for the spiked *E. coli* proteins ($y = 0.9774x + 0.0896$; $R^2 = 0.9563$). We concluded that ASAP^2 applied to CSF provides sufficiently accuracy and can be further used for real-life applications.

CONCLUSION

We have characterized, validated, and implemented a highly automated proteomic pipeline for the analysis of human CSF samples. Importantly, the stability and robustness of our workflow was demonstrated on a second body fluid (i.e., CSF), after its initial application to human blood plasma,[1] showing a potential wide range of applications. ASAP^2 delivers accuracy, reproducibility, and throughput and is therefore expected to be recognized as reference workflow for human protein biomarker discovery based on MS. Inherently to shotgun proteomic discovery workflow, our methodology could miss information present at the protein level that might be relevant as biomarker information.

The automation of data processing with an advanced data analysis pipeline is a further improvement that we are now implementing into the ASAP^2 laboratory workflow. Laboratory information management systems (LIMS) that support both Mascot and Scaffold proteomic tools, and integrated data storage solutions, are used. Such an implementation is crucial to minimize the risk of technical variability and errors in both the wet and dry laboratory and facilitates real-life, large-scale studies in clinical proteomic research.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.5b02748.

Comparison of SD values per protein with and without the immuno-affinity depletion step (Figures S1 and S2); numbers of identified and quantified proteins across successive RP-LC MS/MS analyses (Figure S3); proteins presenting the lowest variability across the ASAP^2 workflow (Figure S4) (PDF)

List of proteins identified and quantified in the depleted and total nondepleted CSF samples (XLSX)

List of proteins identified and quantified in the 96 identical human CSF samples (XLSX)

List of peptides identified in the 96 identical human CSF samples (XLSX)

AUTHOR INFORMATION

Corresponding Author

*Tel.: +41 21 632 6114. Fax: +41 21 632 6499. E-mail: loic.dayon@rd.nestle.com.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank John Corthésy and Seu-Ping Guiraud for help and discussions.

REFERENCES

(1) Dayon, L.; Núñez Galindo, A.; Corthésy, J.; Cominetti, O.; Kussmann, M. *J. Proteome Res.* **2014**, *13*, 3837−3845.

(2) Liu, Y.; Buil, A.; Collins, B. C.; Gillet, L. C. J.; Blum, L. C.; Cheng, L. Y.; Vitek, O.; Mouritsen, J.; Lachance, G.; Spector, T. D.; Dermitzakis, E. T.; Aebersold, R. *Mol. Syst. Biol.* **2015**, *11*, 786.

(3) Griggs, R. C.; Józefowicz, R. F.; Aminoff, M. J. In *Goldman's Cecil Medicine*; Elsevier: Amsterdam, 2012; pp 2228−2235.

DOI: 10.1021/acs.analchem.5b02748
Anal. Chem. XXXX, XXX, XXX−XXX

(4) Rosenberg, G. A. In *Bradley's Neurology in Clinical Practice*; Daroff, R. B., Fenichel, G. M., Jankovic, J., Mazziotta, J. C., Eds.; Saunders Elsevier: Philadelphia, PA, 2012.
(5) De Meyer, G.; Shapiro, F.; Vanderstichele, H.; Vanmechelen, E.; Engelborghs, S.; De Deyn, P. P.; Coart, E.; Hansson, O.; Minthon, L.; Zetterberg, H.; Blennow, K.; Shaw, L.; Trojanowski, J. Q. *Arch. Neurol.* **2010**, *67*, 949−956.
(6) McKhann, G. M.; Knopman, D. S.; Chertkow, H.; Hyman, B. T.; Jack, C. R., Jr; Kawas, C. H.; Klunk, W. E.; Koroshetz, W. J.; Manly, J. J.; Mayeux, R.; Mohs, R. C.; Morris, J. C.; Rossor, M. N.; Scheltens, P.; Carrillo, M. C.; Thies, B.; Weintraub, S.; Phelps, C. H. *Alzheimer's Dementia* **2011**, *7*, 263−269.
(7) Maurer, M. H. *Mass Spectrom. Rev.* **2010**, *29*, 17−28.
(8) Yuan, X.; Desiderio, D. M. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2005**, *815*, 179−189.
(9) Abdi, F.; Quinn, J. F.; Jankovic, J.; McIntosh, M.; Leverenz, J. B.; Peskind, E.; Nixon, R.; Nutt, J.; Chung, K.; Zabetian, C.; Samii, A.; Lin, M.; Hattan, S.; Pan, C.; Wang, Y.; Jin, J.; Zhu, D.; Li, G. J.; Liu, Y.; Waichunas, D.; Montine, T. J.; Zhang, J. *J. Alzheimers Dis.* **2006**, *9*, 293−348.
(10) Choe, L.; D'Ascenzo, M.; Relkin, N. R.; Pappin, D.; Ross, P.; Williamson, B.; Guertin, S.; Pribil, P.; Lee, K. H. *Proteomics* **2007**, *7*, 3651−3660.
(11) Dayon, L.; Hainard, A.; Licker, V.; Turck, N.; Kuhn, K.; Hochstrasser, D. F.; Burkhard, P. R.; Sanchez, J. C. *Anal. Chem.* **2008**, *80*, 2921−2931.
(12) Guldbrandsen, A.; Vethe, H.; Farag, Y.; Oveland, E.; Garberg, H.; Berle, M.; Myhr, K. M.; Opsahl, J. A.; Barsnes, H.; Berven, F. S. *Mol. Cell. Proteomics* **2014**, *13*, 3152−3163.
(13) Schutzer, S. E.; Liu, T.; Natelson, B. H.; Angel, T. E.; Schepmoes, A. A.; Purvine, S. O.; Hixson, K. K.; Lipton, M. S.; Camp, D. G., II; Coyle, P. K.; Smith, R. D.; Bergquist, J. *PLoS ONE* **2010**, *5*, e10980 10.1371/journal.pone.0010980.
(14) Hölttä, M.; Minthon, L.; Hansson, O.; Holmén-Larsson, J.; Pike, I.; Ward, M.; Kuhn, K.; Rüetschi, U.; Zetterberg, H.; Blennow, K.; Gobom, J. *J. Proteome Res.* **2015**, *14*, 654−663.
(15) Cominetti, O.; Núñez Galindo, A.; Corthésy, J.; Oller Moreno, S.; Irincheeva, I.; Valsesia, A.; Astrup, A.; Saris, W. H. M.; Hager, J.; Kussmann, M.; Dayon, L. Submitted for Publication 2015.
(16) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551−3567.
(17) Vizcaíno, J. A.; Deutsch, E. W.; Wang, R.; Csordas, A.; Reisinger, F.; Ríos, D.; Dianes, J. A.; Sun, Z.; Farrah, T.; Bandeira, N.; Binz, P. A.; Xenarios, I.; Eisenacher, M.; Mayer, G.; Gatto, L.; Campos, A.; Chalkley, R. J.; Kraus, H. J.; Albar, J. P.; Martinez-Bartolomé, S.; Apweiler, R.; Omenn, G. S.; Martens, L.; Jones, A. R.; Hermjakob, H. *Nat. Biotechnol.* **2014**, *32*, 223−226.
(18) Vizcaíno, J. A.; Côté, R. G.; Csordas, A.; Dianes, J. A.; Fabregat, A.; Foster, J. M.; Griss, J.; Alpi, E.; Birim, M.; Contell, J.; O'Kelly, G.; Schoenegger, A.; Ovelleiro, D.; Pérez-Riverol, Y.; Reisinger, F.; Ríos, D.; Wang, R.; Hermjakob, H. *Nucleic Acids Res.* **2013**, *41*, D1063−D1069.
(19) Huang, D. W.; Sherman, B. T.; Lempicki, R. A. *Nat. Protoc.* **2008**, *4*, 44−57.
(20) Huang, D. W.; Sherman, B. T.; Lempicki, R. A. *Nucleic Acids Res.* **2009**, *37*, 1−13.
(21) Roche, S.; Gabelle, A.; Lehmann, S. *Proteomics: Clin. Appl.* **2008**, *2*, 428−436.
(22) Borg, J.; Campos, A.; Diema, C.; Omeñaca, N.; De Oliveira, E.; Guinovart, J.; Vilaseca, M. *Clin. Proteomics* **2011**, *8* (1), 6 10.1186/1559-0275-8-6.
(23) Ogata, Y.; Charlesworth, M. C.; Muddiman, D. C. *J. Proteome Res.* **2005**, *4*, 837−845.
(24) Percy, A. J.; Yang, J.; Chambers, A. G.; Simon, R.; Hardie, D. B.; Borchers, C. H. *J. Proteome Res.* **2014**, *13*, 3733−3747.
(25) Shores, K. S.; Knapp, D. R. *J. Proteome Res.* **2007**, *6*, 3739−3751.
(26) Yadav, A. K.; Bhardwaj, G.; Basak, T.; Kumar, D.; Ahmad, S.; Priyadarshini, R.; Singh, A. K.; Dash, D.; Sengupta, S. *PLoS ONE* **2011**, *6* (9), e24442 10.1371/journal.pone.0024442.
(27) Jedrychowski, M. P.; Wrann, C. D.; Paulo, J. A.; Gerber, K. K.; Szpyt, J.; Robinson, M. M.; Nair, K. S.; Gygi, S. P.; Spiegelman, B. M. *Cell Metab.* **2015**, *22*, 734−740.
(28) Tu, C.; Rudnick, P. A.; Martinez, M. Y.; Cheek, K. L.; Stein, S. E.; Slebos, R. J. C.; Liebler, D. C. *J. Proteome Res.* **2010**, *9*, 4982−4991.
(29) Carrasquillo, M. M.; Zou, F.; Pankratz, V. S.; Wilcox, S. L.; Ma, L.; Walker, L. P.; Younkin, S. G.; Younkin, C. S.; Younkin, L. H.; Bisceglio, G. D.; Ertekin-Taner, N.; Crook, J. E.; Dickson, D. W.; Petersen, R. C.; Graff-Radford, N. R.; Younkin, S. G. *Nat. Genet.* **2009**, *41*, 192−198.
(30) Hirabayashi, T.; Yagi, T. *Adv. Neurobiol.* **2014**, *8*, 293−314.
(31) Liu, G.; Jiang, Y.; Wang, P.; Feng, R.; Jiang, N.; Chen, X.; Song, H.; Chen, Z. *J. Neurochem.* **2012**, *120*, 190−198.

DOI: 10.1021/acs.analchem.5b02748
Anal. Chem. XXXX, XXX, XXX−XXX