

RESEARCH ARTICLE

Protein variability in cerebrospinal fluid and its possible implications for neurological protein biomarker research

Lukas M. Schilde¹, Steffen Kösters¹, Simone Steinbach¹, Karin Schork¹, Martin Eisenacher¹, Sara Galozzi¹, Michael Turewicz¹, Katalin Barkovits¹, Brit Mollenhauer², Katrin Marcus^{1†*}, Caroline May^{1‡}

1 Medizinisches Proteom-Center, Ruhr-University Bochum, Universitaetsstrasse, Bochum, Germany, **2** Paracelsus-Elena-Klinik, Klinikstraße, Kassel, and University Medical Center Göttingen, Department of Neurology, Göttingen, Germany

☞ These authors contributed equally to this work.

‡ These authors also contributed equally to this work.

* Katrin.Marcus@rub.de



OPEN ACCESS

Citation: Schilde LM, Kösters S, Steinbach S, Schork K, Eisenacher M, Galozzi S, et al. (2018) Protein variability in cerebrospinal fluid and its possible implications for neurological protein biomarker research. *PLoS ONE* 13(11): e0206478. <https://doi.org/10.1371/journal.pone.0206478>

Editor: Stephen D. Ginsberg, Nathan S Kline Institute, UNITED STATES

Received: May 15, 2018

Accepted: October 12, 2018

Published: November 29, 2018

Copyright: © 2018 Schilde et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (<http://www.proteomexchange.org/>) with the dataset identifier PXD007555.

Funding: This work was supported by the Deutsche Parkinson Gesellschaft, Medical Faculty at RUB (FoRUM), the European Union (NISCI, GA no. 681094), the German Federal Ministry of Education and Research (WTZ with Brasil, FKZ

Abstract

Cerebrospinal fluid is investigated in biomarker studies for various neurological disorders of the central nervous system due to its proximity to the brain. Currently, only a limited number of biomarkers have been validated in independent studies. The high variability in the protein composition and protein abundance of cerebrospinal fluid between as well as within individuals might be an important reason for this phenomenon. To evaluate this possibility, we investigated the inter- and intraindividual variability in the cerebrospinal fluid proteome globally, with a specific focus on disease biomarkers described in the literature. Cerebrospinal fluid from a longitudinal study group including 12 healthy control subjects was analyzed by label-free quantification (LFQ) via LC-MS/MS. Data were quantified via MaxQuant. Then, the intra- and interindividual variability and the reference change value were calculated for every protein. We identified and quantified 791 proteins, and 216 of these proteins were abundant in all samples and were selected for further analysis. For these proteins, we found an interindividual coefficient of variation of up to 101.5% and an intraindividual coefficient of variation of up to 29.3%. Remarkably, these values were comparably high for both proteins that were published as disease biomarkers and other proteins. Our results support the hypothesis that natural variability greatly impacts cerebrospinal fluid protein biomarkers because high variability can lead to unreliable results. Thus, we suggest controlling the variability of each protein to distinguish between good and bad biomarker candidates, e.g., by utilizing reference change values to improve the process of evaluating potential biomarkers in future studies.

01DN14023), the HUPO Brain Proteome Project (HBPP), PURE, a project of Northrhine Westfalia, a federal German state and by de.NBI, a project of the German Federal Ministry of Education and Research [FKZ 031 A 534A]. No funding bodies had any role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Human cerebrospinal fluid (CSF) is a clear body fluid that is produced by filtration of blood in the choroid *plexus* of the first three brain ventricles. Normally, CSF contains less than five cells per μL . The total protein concentration of CSF varies between 0.2% and 0.5% of the total protein concentration of blood [1]. It is considered that 80% of CSF proteins originate in blood and that CSF proteins are diluted in a molecule-size-dependent concentration gradient [2]. The concentration of blood-derived proteins increases from the ventricles to the cistern to the lumbar CSF. It is assumed that the remaining 20% of CSF proteins are released from the central nervous system (CNS) [2]. The primary role of CSF is to protect the CNS from mechanical shocks [3, 4]. Another important function of CSF is to maintain metabolite clearance from the adult brain by circulation. Moreover, CSF supports the homeostatic balance in the brain and, therewith, normal brain activity [4].

Protein biomarkers are the focus of many studies investigating CNS disorders because the main issue for the clinical diagnosis of such disorders is nonspecific clinical symptomatology [5]. For example, Parkinson's disease (PD), Alzheimer's disease (AD), Huntington's disease (HD), multiple sclerosis (MS) and other neurological disorders usually show many similar symptoms. This similarity often results in a delayed diagnosis or even misdiagnosis [6]. The problem of an ambiguous diagnosis occurs particularly in the early stages of these diseases, when the clinical symptoms (e.g., depression) are not specific to any one of the mentioned disorders [7]. A reliable and definitive diagnosis requires pathological confirmation. However, *in vivo* brain tissue sampling by biopsy to support the clinical diagnosis is often not realizable and results in too many health risks for routine workup in many disorders. Thus, to improve *in vivo* diagnoses and to evaluate disease progression or potential therapeutic effects, new reliable biomarkers are needed [8]. For clinical application, biomarkers need to be easily accessible. In contrast to brain biopsy, CSF is easily and safely accessible by established methods [9]. Several studies already detected CSF alterations decades before the first clinical symptoms appeared [10]. Therefore, many disease and progression biomarkers for neurological disorders were investigated in CSF in the past [11–14]. Regardless of the disease, there is a pronounced lack of reproducibility of these unbiased proteomic studies [12, 15]. For example, for PD, many proteins found in CSF by proteome analysis have been investigated as biomarkers but showed inconclusive or conflicting results [16–18]. These findings also apply with other neurological disorders [14, 19]. One of the factors contributing to this issue could be the high variability of CSF proteins for various reasons [20]. In contrast, AD studies revealed that a combination of the main components of the disease-specific pathological hallmarks are promising biomarkers [21, 22]. However, reliable biomarkers are needed for early diagnosis and upcoming putative neuroprotective trials for all neurological disorders [8, 22, 23].

Multiple studies have previously investigated the CSF proteome in the context of variability and biomarkers in different diseases. For example, in 2005, Hu *et al.* used samples from six subjects obtained within a sampling interval of two weeks. Four subjects were cognitively normal, and two subjects showed indications of very mild dementia. Two-dimensional difference gel electrophoreses (DIGE) was used for differential analysis of protein spots between the sampled time points, followed by tandem mass spectrometry for identification of these proteins [24]. The researchers investigated the intraindividual variability within two weeks between sampling as well as the interindividual variability, but only for the differential protein spots. Furthermore, Hühmer and colleagues summarized the efforts and progress in CSF protein profiling. Evaluating CSF composition studies, the authors published a list of detected CSF proteins that were identified in at least two independent studies of human CSF [15]. A further CSF proteome study was performed by a Canadian research group in 2016. The main purpose

of this study was to identify brain-related proteins in CSF that are suitable for the development of diagnostic assays. The authors found 78 brain-related proteins in CSF of at least 4 of 6 healthy individuals [25]. Another study analyzed the interindividual variability in CSF protein abundances in samples of 9 patients undergoing routine, nonneurological surgical procedures [20]. The researchers presumed that an understanding of the biological variation in CSF proteins healthy individuals is essential for reliable interpretation of studies for neurological disorder biomarkers [20]. In 2018, Trombetta *et al.* implemented a fit-for-purpose modeled approach to qualify a broad selection of commercially available immunoassays [26]. Therefore, paired baseline and eight-week CSF samples from twenty participants with mild cognitive impairment or mild dementia due to AD were used [26]. Trombetta *et al.* demonstrated consistent sensitivity, reliability and biotemporal stability of different immunoassays for 32 different CSF analytes i.a. by calculating the coefficients of variation (CVs) [26].

The aim of this study was to investigate the implications of the high variability in CSF protein composition and protein abundance in healthy control subjects for biomarker discovery studies. It is known that the variability is influenced by the analytical variability as well as the biological variability [27, 28]. In addition to standard deviation, variance and CV, a measure of variability is the reference change value. This value represents the change in the abundance of a protein (e.g., a potential biomarker) that is necessary for the change to be considered larger than expected after taking into account biological and technical variance [29]. Via unbiased LFQ via LC-MS/MS, we investigated CSF samples obtained from a longitudinal study group of 12 healthy control subjects (12 samples for every time point: 0 months, T_0 ; 24 months, T_{24} ; 48 months, T_{48}) and determined the intra- and interindividual variability, as well as the reference change value, for each protein.

Materials and methods

Ethical statement

The ethics committee of the Physician's Board Hessen, Germany (approval no. FF89/2008) approved this study, which is registered at the German Register for Clinical Trials (DRKS00000540) according to the WHO Trial Registration Data Set. All participants provided written, fully informed consent to participate in this study. The relevant documents relating to this process are archived at Paracelsus-Elena-Klinik in Kassel, Germany. Only the anonymous data and materials from the participants were provided to the scientists carrying out the research. The data concerning this study were stored separately from the hospital charts of the patients.

Subjects and samples

This CSF core proteome study has a longitudinal design with 12 human subjects evaluated at three different time points. The subjects were selected as previously described by Mollenhauer *et al.* [18]. These 12 healthy subjects are a representative group of 90 neurologically unimpaired healthy subjects who are participating in the longitudinal DeNoPa study (de Novo Parkinson study), which is a prospective, single center study performed as described by Mollenhauer *et al.* [18]. All subjects (volunteer CSF donors) were recruited at Paracelsus-Elena-Klinik in Kassel, Germany. The subjects were not eligible for the study if they demonstrated cognitive impairment (Mini-Mental State Examination score < 27) or any other neurological disease. A neurologist evaluated qualified subjects according to a standardized assessment protocol, including a neurological examination and MRI, among tests (S1 Table). This examination was repeated at every time point (T_0 , T_{24} , T_{48}) for every subject. No subject showed symptoms of any neurological disease. Further, routine laboratory blood analysis was

carried out. Blood was collected with BD Vacutainer system tubes (BD, Franklin Lakes, NJ, USA) by venous puncture and processed according to published standard operating procedures (SOPs) between 7 and 9 a.m. after 12 hours fasting [18]. Aliquots were stored at -80 °C within 30 min following the venous puncture.

Characteristics of the study group

The age and gender characteristics of our study group are shown in Table 1. The median age of human subjects was 65 at the beginning of the study, with 39.1% of the subjects being female. The mean time between each of the follow-up assessments (T₀ (first), T₂₄ and T₄₈) was two years. Routine blood and CSF laboratory analyses were performed to exclude severe diseases that can affect CSF. Furthermore, blood contamination of CSF was clinically assessed by the erythrocyte count. In some samples, slight contamination from blood was found (0 to 64 erythrocytes, median: 0) but were negligible, as described by Reiber [1]. Otherwise, no significant deviations related to default values were revealed in the routine laboratory analysis of the CSF. With all the tests and information, the selected study group covered all characteristics necessary for our study.

CSF sampling for routine laboratory analysis and mass spectrometric analysis

CSF was obtained from the subarachnoid space of the lumbar spinal cord by lumbar puncture. Sampling was performed as described by Mollenhauer *et al.* [18]. Routine clinical variables were determined using established routine protocols (Table 2 and S2 Table). The CSF was centrifuged at 2,500 x g for 10 min at room temperature, and the supernatant was collected for further analysis. The time between obtaining the sample and the centrifugation step was under 30 min. Within 30 min, the samples were stored in aliquots at -80 °C until further analysis.

CSF sample preparation for mass spectrometric analysis

Before performing mass spectrometric analysis, protein digestion of the CSF samples was performed according to the protocol published by Stoop *et al.* with slight modifications [20]. In short, CSF samples (50 µL) were incubated 1:1 with 0.2% (s/v) RapiGest SF Surfactant (Waters Corporation, Milford, MA, USA) (in 50 mM ammonium bicarbonate). Proteins were reduced

Table 1. Gender, age and neuropsychological characteristics of the representative study group.

	Standard value	Study group						
		N	f*	m*	Mean	SD*	CV*	Median
Age [years]	40–85	12	5	7	69.1	4.7	6.7%	69.0
NMS Quest total [points]	rating scale	12	5	7	4.3	2.9	67.7%	4.0
NMS Quest sum [points]	rating scale	12	5	7	0.1	0.1	68.4%	0.1
UPDRS total [points]	rating scale	12	5	7	2.2	2.2	99.6%	2.0
MDS UPDRS total [points]	rating scale	12	5	7	6.3	5.0	79.8%	5.5
BDI total [points]	0–8	12	5	7	5.0	4.5	90.2%	3.1
MMSE total [points]	24–30	12	5	7	28.3	1.5	5.3%	29.0
Clock test: [points]	<2	12	5	7	1.3	0.6	43.3%	1.0

Of the 90 human subjects participating in the DeNoPa study twelve were selected for this CSF study. The representative study group analyzed in the present study reflects the total study group in terms of gender frequencies, age distribution as well as neuropsychological characteristics.

*SD = standard deviation, CV = coefficient of variation, f = female, m = male.

<https://doi.org/10.1371/journal.pone.0206478.t001>

Table 2. CSF and routine laboratory analysis for the representative study group analyzed in the present study.

	Standard value	Study group				
		N	Mean	SD*	CV*	Median
Proteins [mg/L]	200–400	12	424.6	106.9	25.2%	413.5
White blood cell count [cells/mm ³]	0–4	12	0.4	0.6	151.9%	0.0
Erythrocyte count [cells/mm ³]	0–10	12	4.9	13.3	271.8%	0.0
CSF albumin [mg/L]	0.0–350.0	12	276.2	77.3	28.0%	255.5
Serum albumin [g/L]	35.00–55.00	12	42.7	2.6	6.1%	42.8
Albumin Quotient [ratio]	< 8	12	6.5	1.8	27.4%	6.2
CSF IgG [mg/L]	10.0–40.0	12	29.4	8.0	27.4%	30.5
Serum IgG [g/L]	8.0–18.0	12	10.0	1.8	18.2%	9.7
IgG Quotient [ratio]	~ 2,3	12	3.0	0.9	29.4%	2.8
CSF IgA [mg/L]	1.5–6.0	12	5.3	2.9	55.3%	4.6
Serum IgA [g/L]	0.9–4.5	12	3.4	1.6	46.2%	3.1
IgA Quotient [ratio]	~ 1,3	12	1.5	0.6	39.9%	1.4
CSF IgM [mg/L]	0.0–1.0	12	0.2	0.1	37.5%	0.2
Serum IgM [g/L]	0.6–2.5	12	1.0	0.5	48.0%	0.8
IgM Quotient [ratio]	~ 0.3	12	0.3	0.3	82.7%	0.2

Clinical CSF parameters (e.g., red and white blood cell count and protein concentrations) were determined to ensure a healthy study group. In addition CSF/serum quotients were determined.

*SD = standard deviation, CV = coefficient of variation.

<https://doi.org/10.1371/journal.pone.0206478.t002>

with 5 mM dithiothreitol (final concentration) at 60 °C for 30 min. Next, iodoacetamide was added for alkylation (final concentration of 15 mM), and the samples were incubated at room temperature for 30 min in the dark. Tryptic digestion was performed at 37 °C overnight (16 hours) at a 1:50 (w/w) trypsin-to-protein ratio. Digestion was stopped by adding 25% (v/v) trifluoroacetic acid (TFA) for a final concentration of 0.5% (v/v). After incubation at 37 °C for 40 min, the samples were centrifuged (Centrifuge 5417R, Rotor: FC45-30-11, Eppendorf AG, Hamburg, Germany) at 17,000 x g and 4 °C for 10 min. Clear supernatants of the samples were transferred into new reaction tubes and dried with a rotational vacuum concentrator (SpeedDry RVC 2–25 CDplus, Martin Christ Gefriertrocknungsanlagen GmbH, Osterode, Germany). Finally, samples were resuspended in 0.1% (v/v) TFA. The peptide concentration was measured by amino acid analysis, as described by Plum *et al.* [30].

Mass spectrometric analysis

Protein digests were analyzed on a nanoHPLC system (UltiMate 3000, Dionex, Idstein, Germany) coupled on-line to a quadrupole orbitrap mass spectrometer (Q Exactive, Thermo Fisher Scientific, Bremen, Germany). Samples were injected into the nanoHPLC system by an autosampler and were loaded on a C18 trap column (PepMap100 C18, 100 µm ID x 2 cm, particle size 5 µm and pore size 100 Å; Thermo Scientific, Rockford, IL, USA) using 0.1% (v/v) TFA and a flow rate of 30 µL/min. After sample loading, the trap column was switched to an analytical C18 column (PepMap C18, 75 µm x 50 cm, particle size 5 µm and pore size 100 Å; Thermo Scientific, Rockford, IL, USA). For peptide separation, the following solvent system was used: buffer A: 0.1% (v/v) formic acid; buffer B: 84% (v/v) acetonitrile, 0.1% (v/v) formic acid. A linear gradient of 4–50% buffer B was carried out for 180 min at a flow rate of 400 nL/min, followed by a washing step with 95% B for 5 min and an equilibration step with 5% B for 5 min. The column oven temperature was set to 60 °C. Ionization took place in a nano

electrospray ionization source (ESI), and mass spectrometric analysis was performed in data-dependent scan mode. For MS/MS analysis, full MS spectra were scanned in the range from 350–1,400 m/z with a resolution of 70,000 at 200 m/z (automatic gain control (AGC) target 2e6, 80 ms maximum injection time). The spray voltage was set to 1,600 V, and the capillary temperature was set to 250 °C. For internal recalibrations, the lock mass polydimethylcyclosiloxane (m/z: 445.120) was used. The 10 ions with the highest intensities were selected for higher energy collision-induced dissociation (HCD) fragmentation. MS/MS fragments were generated with a 27% normalized collision energy, an isolation window of 2.2 m/z and a fixed first mass of 130 m/z. An orbitrap analyzer with a resolution of 35,000 at 200 m/z (AGC 5e5, maximum injection time 120 ms) was used for fragment analysis. For the assessment of the analytical variability, seven replicates of pooled samples were measured.

Protein quantification

Mass spectrometric data were analyzed with the intensity-based quantification software MaxQuant (version 1.5.3.12) [31]. MS/MS spectra were searched against the UniProt/Swiss-Prot [32] human proteome database (UniProtKB/Swiss-Prot UniProt release 2017_01; downloaded 2017-01-26; number of entries 553,474) using the search engine Andromeda [33]; the search included 262 common contaminants and concatenated with the reversed versions of all sequences [34]. The precursor and fragment ion mass tolerance were set to 5 ppm and 20 ppm, respectively. The enzyme specificity was set to trypsin, and two missed cleavages were allowed. The minimum peptide length was set to 7 amino acids. Cysteine carbamidomethylation was set as fixed, and methionine oxidation and N-terminal acetylation were set as variable modifications. A maximum of 5 modifications per peptide was set. For both peptide spectrum matches (PSMs) and protein level, the false discovery rate (FDR) was set to 1%. For the calculation of the protein abundances, label-free quantification (LFQ, [35]) was performed with an LFQ minimum ratio count of two. Normalized LFQ intensities were used for further data analysis.

Mathematical calculation of the variation in protein abundance

For each protein, its CV is defined as the ratio of the standard deviation (s) to the mean (\bar{x}) in %:

$$CV = s/\bar{x} * 100 \tag{1}$$

In this paper, several types of CVs are used. First, to assess the analytical variability or inaccuracy, CV_a can be calculated using the standard deviation and mean of repeated measurements of the same sample. To this end, in this paper, 4 technical replicates were used. Second, for each subject, protein abundances were measured at three different time points (i.e., T₀, T₂₄ and T₄₈), and for each protein, a CV for the three abundances was calculated. The median of these CVs across all samples is called CV_t and represents the intraindividual variability over time. Finally, to obtain the measured interindividual CV between different subjects (CV_g), the mean and the standard deviation of the protein abundances across all samples at T₀ are used in formula (1) to obtain CV_g.

The reference change value (RCV) describes the change in the protein abundance (e.g., a potential protein biomarker) in % that is added to the original abundance (i.e., 100%) and is necessary to reflect an “unexpectedly high” change [29], i.e., a change that is unlikely to occur based only on the analytical, inter- or intraindividual variability and might be biologically relevant (e.g., a potential protein biomarker). The original formula is given in (2):

$$RCV = 1.96 * \sqrt{2} * \sqrt{CV_a^2 + CV_i^2} \tag{2}$$

with CV_i standing for the intraindividual biological CV given by

$$CV_i = \sqrt{CV_t^2 - CV_a^2} \quad (3)$$

The factor 1.96 corresponds to a 95% confidence level assuming a normal distribution. By inserting formula (3) into formula (2), CV_a^2 is canceled out, and the square root vanishes. Therefore, the formula for the RCV can be simplified to

$$RCV_t = 1.96 * \sqrt{2} * CV_t \quad (4)$$

This RCV is called RCV_t because it uses CV_t and applies to the intraindividual variability. We modify the definition for RCV_g by exchanging CV_t for CV_g in order to reflect “unexpectedly high” changes between different individuals:

$$RCV_g = 1.96 * \sqrt{2} * CV_g \quad (5)$$

Generally, the RCV is always larger than zero, and in order to compare the RCV to fold changes, the original abundance (i.e., 100%) needs to be added so that $(RCV + 100\%)/100\%$ has to be considered. Then, whether a fold change is unexpectedly high can be assessed. For a CV_g value of 50%, the following RCV_g can be calculated:

$$RCV_g = 1.96 * \sqrt{2} * 50\% \quad (6)$$

$$RCV_g = 139\% \quad (7)$$

Consequently, for this example, the fold change should be larger than 2.39 to reflect an “unexpectedly high” change in protein abundance.

Results

Basic technical information of LC-MS/MS runs

In quantitative mass spectrometry, accuracy is highly important. Therefore, in Fig 1, box plots of the LFQ intensities and peptide retention times plotted in the order in which the samples were run are shown. Furthermore, we investigated the analytical variation in our method (CV_a). To calculate the analytical variability CV_a , pooled samples were measured four times during the whole measuring sequence for the 36 individual samples. The resulting average CV_a was 8.49%.

Global description of proteins identified in CSF

We identified a total of 5601 peptides across all 12 analyzed samples from neurologically healthy individuals (Fig 2), and the three time points, T_0 , T_{24} and T_{48} , together with 36 samples, were analyzed. These peptides were matched to 791 protein groups hereinafter briefly referred to as “proteins” by MaxQuant’s protein inference (e.g., see [36] regarding the relevance of protein inference). The selection criterion used to obtain this number of proteins was at least one identified unique peptide per protein across all samples. To investigate the variability, only 223 proteins that were found in all 36 samples were used (S1 File). Then, to increase the significance of the identified proteins, the selection criterion was set to two unique peptides per protein. With this choice, the number of proteins decreased from 791 to 610 proteins (S2 File) and from 223 to 216 (S3 File). In the following, this resulting protein list (S3 File) will be referred to as the “CSF core proteome”.

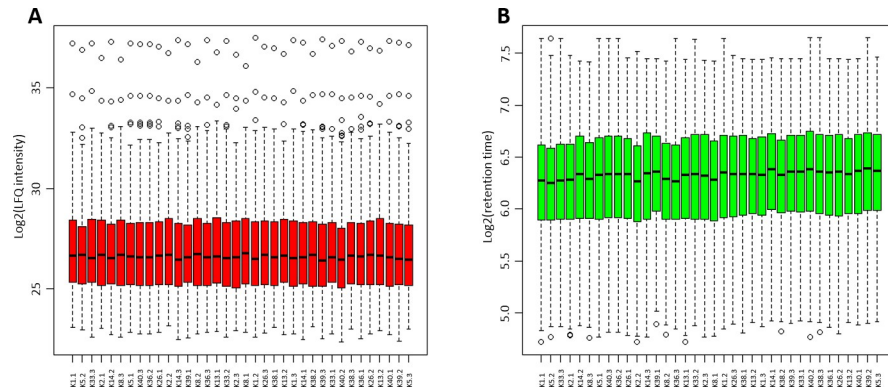


Fig 1. Box plots of LFQ intensities and peptide retention times. Log₂ of the LFQ intensities of all proteins, which were quantified in 100% of the samples, are shown in A (red box plots). The data were not further normalized or transformed once the values from MaxQuant were obtained. B shows the log₂ of peptide retention times for every LC-MS/MS run (green box plots).

<https://doi.org/10.1371/journal.pone.0206478.g001>

High variability of protein abundance in the CSF core proteome

To assess intraindividual variability, first, the protein overlap of samples from different time points (T₀, T₂₄ and T₄₈) for each individual human subject was compared. For visualization of the intraindividual overlaps, Venn diagrams were generated (Fig 3A). Within one human subject, between 284 and 370 proteins were found at all three time points. Furthermore, between 5 and 47 proteins were identified in only one subject at a particular time point. There was no subject with a 100% overlap of identified proteins at all three time points.

Additionally, to assess the intra- and interindividual variability, CV_t and CV_g were calculated. The calculated CV_t values were between 2.6% and 29.3%, with a mean of 8.2% (S2 File). The determined CV_g was between 9.3% and 101.5%, with a mean of 25.7% (S1 File). To visualize the intra- and interindividual variability, the CV_t and CV_g values of the CSF core proteins were plotted against their abundance rank (Fig 3B and 3C), and the top 50 proteins of the CSF

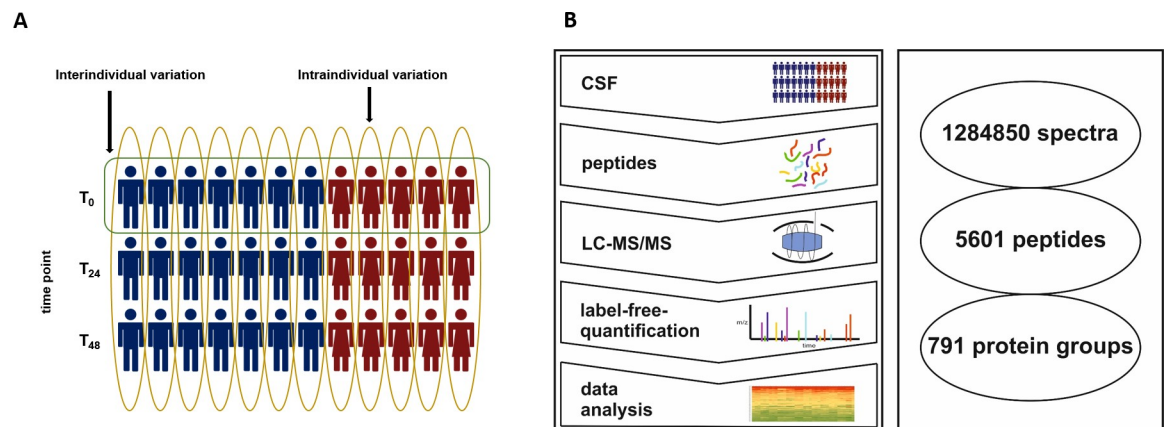


Fig 2. Study overview. A: Two variables, time point (T₀, T₂₄ and T₄₈) and human subject (red figures represent female subjects, and blue figures represent male subjects), can characterize every sample in this study, and for every human subject, three samples were obtained. In this study, the inter- and intraindividual variability of the CSF were analyzed. Hereby, the intraindividual variability is defined as the variation within one human subject between the three time points (yellow circle), and the interindividual variability is defined as the variation between all subjects at T₀ (green circle). B: The workflow included CSF sampling, sample preparation, tryptic digestion, LC-MS/MS with LFQ and data analysis. We identified 1,284,850 spectra, which resulted in 5601 peptides and 791 protein groups.

<https://doi.org/10.1371/journal.pone.0206478.g002>

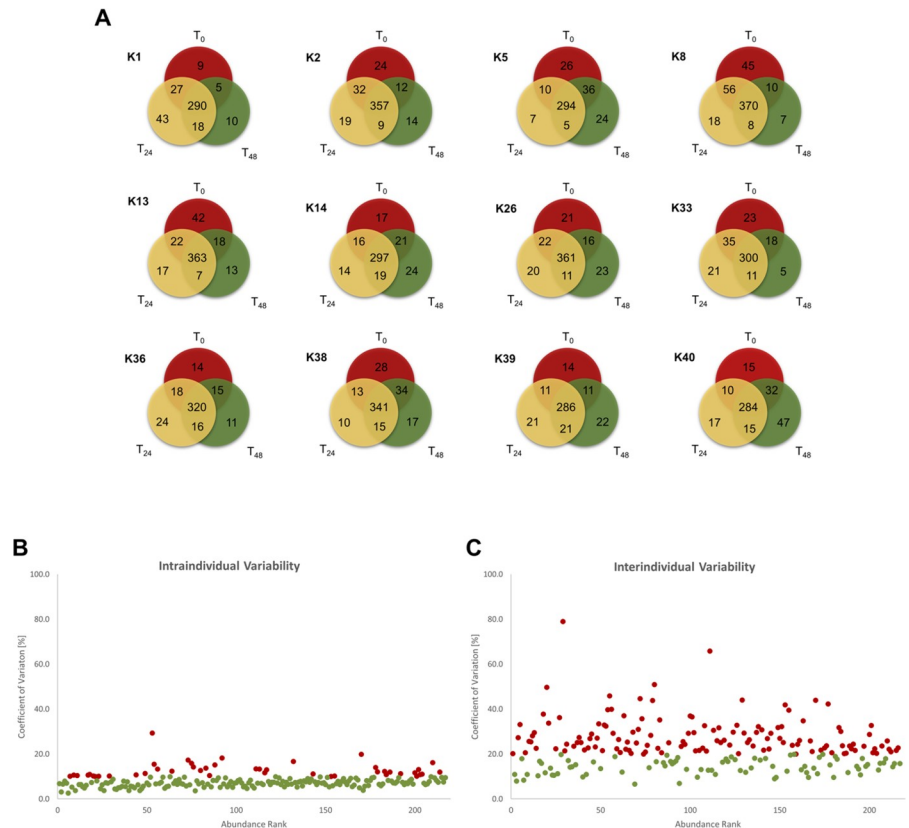


Fig 3. Intra- and interindividual variability. **A:** This figure provides twelve Venn diagrams, one for each human subject. Only proteins that were found with at least 2 unique peptides in one of three samples were counted. The three time points are represented by different colors. Proteins that were identified in all three samples are shown in the middle overlap. Each overlap between two circles gives the number of proteins that were identified in these two samples. **B:** CV_t values were calculated for all 216 proteins of the CSF core proteome and were plotted against their abundance rank. Proteins with $CV_t > 10\%$ are colored in red, and those with $CV_t < 10\%$ are colored green. **C:** CV_g values were calculated for all 216 proteins of the CSF core proteome and were plotted against their abundance rank. Proteins with $CV_g > 20\%$ are colored red, and those with $CV_g < 20\%$ are colored green.

<https://doi.org/10.1371/journal.pone.0206478.g003>

core proteome showing the highest intensities were plotted in a heatmap to give a broad sense of the variability within and between individuals (Fig 4). Here, the protein with the highest mean LFQ intensity ($1.43e11$) was serum albumin (P02678; ALB), followed by serotransferrin (P02787; TF; mean LFQ intensity: $2.47e10$), prostaglandin-H2 D-isomerase (P4122; PTGDS; mean LFQ intensity: $8.86e9$) and Ig gamma-1 chain C region (P01857; IGHG1; mean LFQ intensity: $7.68e9$). The CV_g values from the top 50 abundant proteins ranged from 9.8% (prostaglandin-H2 D-isomerase) to 101.5% (haptoglobin). The CV_t values varied between 2.6% (complement C3) and 11.3% (apolipoprotein A-IV) (S3 Table).

We determined RCV_t and RCV_g for all proteins of the CSF core proteome. The obtained RCV_t values ranged from 7.2% to 81.3% (S3 File), with a mean of 25.7%. The RCV_g values varied between 19.2% and 281.4%, with a mean of 71.2%. The highest RCV_g was calculated for haptoglobin, and the highest RCV_t was calculated for beta-2-microglobulin.

In conclusion, there is a wide range of variability in protein abundance in the CSF core proteome between subjects as well as within one subject. This variability is also indicated by the broad range of RCVs. This high variability might have been attributed to a false identification of potential protein biomarkers. To evaluate this assumption, we looked more closely at

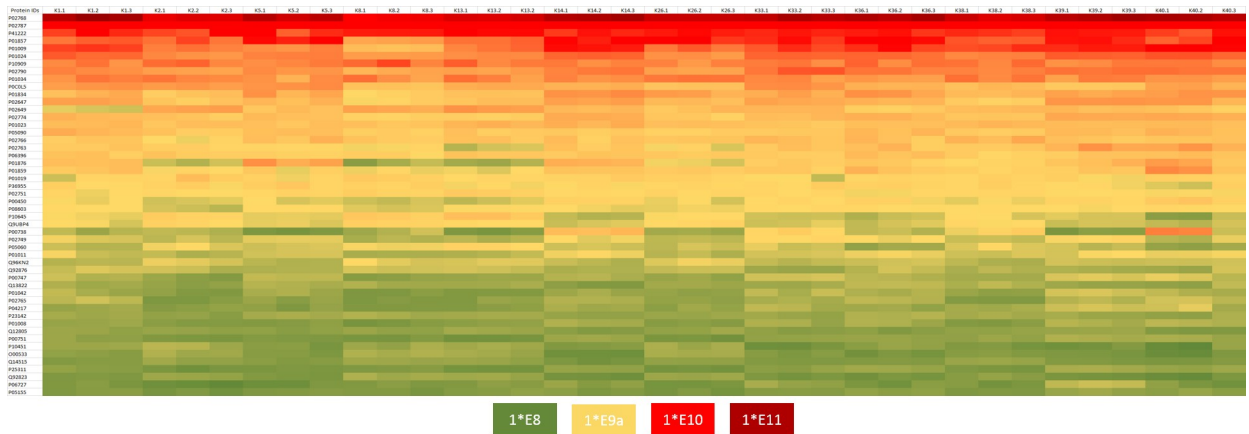


Fig 4. LFQ intensity-based heatmap of the top 50 proteins. LFQ intensity-based heatmap of the top 50 proteins. Top 50 proteins were defined by their mean LFQ intensity of all samples. In each column, the LFQ intensity for one sample is visualized. The color range of LFQ intensity extends from 1e8 (green) over 1e9 (yellow) and 1e10 (red) to 1e11 (dark red).

<https://doi.org/10.1371/journal.pone.0206478.g004>

specific proteins within our CSF core proteome that were proposed in previous studies as potential protein biomarkers for different neurological diseases.

Protein biomarkers often identified in various diseases

Several proteins have been proposed as biomarker candidates for neurological disorders of the CNS. However, these proteins exhibited partially contradictory results. In this context, we generated a list of CSF protein biomarker candidates by searching the literature regarding potential CSF biomarkers with a particular focus on the neurological diseases AD, PD, MS and HD. Table 3 presents 20 proteins of the CSF core proteome that were mentioned in different publications as potential biomarkers for several diseases and were detected by mass spectrometric approaches. The CV and RCV values of these proteins were assessed. Generally, low CVs and RCVs might hint at the possibility that a protein could be a suitable biomarker. However, these proposed biomarkers show a wide range of variability.

Discussion

Generally, the suitability of a molecule as a biomarker depends on its biological variability [27, 29]; thus, we presume that this condition also applies to protein biomarkers in neurological diseases. Unfortunately, the current knowledge about protein variability in CSF is insufficient, although there are some CSF protein variability studies. In 2010, Schutzer *et al.* investigated a pair of individual CSF samples obtained four weeks apart from 10 healthy persons by LC-MS/MS. The authors concluded that there might be general variability, which is relatively limited in a single individual over a short time [51]. Perrin *et al.* went further in 2013. Based on their LC-MS/MS study of two aliquots of CSF samples from 6 cognitively normal individuals, the researchers suggest that interindividual variability has strong implications for the potential of a protein to serve as a biomarker [52]. Unfortunately, both studies have limitations because neither of these studies performed an investigation over a long period. Furthermore, Perrin *et al.* focused more on the technical process for CSF biomarker discovery [52]. In this context, the purpose and scientific contribution of this study was to investigate the variability of the CSF proteome in healthy control subjects in order to determine its impact on biomarker discovery studies and provide a statistical strategy to assess protein variability in future studies.

Table 3. Protein biomarker candidates with their CVs and RCVs.

Protein IDs	Gene name	Protein name	CV _t / CV _g [%]	RCV _t / RCV _g [%]	PD/PDD	AD	HD	MS	Other diseases
P00738	HP	Haptoglobin	10.2 / 101.5	28.2 / 281.4	[37]	[37, 38]		[39]	
P02675	FGB	Fibrinogen beta chain	13.4 / 45.5	37.0 / 126.0	[37]		[40]		
P02652	APOA2	Apolipoprotein A-II	17.3 / 46.6	48.0 / 129.3	[37]			[41]	[42]
P06727	APOA4	Apolipoprotein A-IV	11.3 / 33.1	31.4 / 91.7				[39, 41]	
P61769	B2M	Beta-2-microglobulin	29.3 / 52.5	81.3 / 145.5	[37]	[43]			
P02649	APOE	Apolipoprotein E	6.5 / 31.8	17.9 / 88.2		[44, 45]		[46]	[42, 44]
P02753	RBP4	Retinol-binding protein 4	15.1 / 27.7	42.0 / 76.9	[37]		[40]	[41]	
P25311	AZGP1	Zinc-alpha-2-glycoprotein	4.5 / 23.5	12.6 / 65.1	[37]	[47]			
P02774	GC	Vitamin D-binding protein	5.5 / 22.6	15.2 / 62.7		[44]		[39, 41]	[44]
P02749	APOH	Beta-2-glycoprotein 1	7.7 / 23.6	21.2 / 65.5	[37]		[40]	[41]	
P00441	SOD1	Superoxide dismutase [Cu-Zn]	15.9 / 32.3	44.1 / 89.6		[37]			[37]
P02768	ALB	Serum albumin	6.9 / 21.0	19.0 / 58.3		[48]		[41]	
P01024	C3	Complement C3	2.6 / 19.0	7.2 / 52.6			[40]	[41, 49, 50]	
P05090	APOD	Apolipoprotein D	6.1 / 19.9	17.0 / 55.1		[37]		[41, 49]	
P36222	CHI3L1	Chitinase-3-like protein 1	5.2 / 30.2	14.5 / 83.8		[44, 45]		[39]	[44]
P02766	TTR	Transthyretin	10.7 / 19.0	29.5 / 52.8		[44]			[37]
Q13822	ENPP2	Extonucleotide pyrophosphatase	5.6 / 14.7	15.5 / 40.9		[44]	[40]		[37]
Q92876	KLK6	Kallikrein-6	5.0 / 16.4	13.7 / 45.5	[37]	[37]		[41, 49]	
P01034	CST3	Cystatin-C	10.6 / 15.4	29.4 / 42.7	[37]	[38, 44]		[41, 49]	[44]
P10909	CLU	Clusterin	10.2 / 12.8	28.2 / 35.4	[38]	[44]		[41, 49]	[44]

20 protein biomarker candidates with their respective intra- and interindividual coefficient of variation (CV_t and CV_g) and reference change values (RCV_t and RCV_g) are listed as well as the studies in which they have been proposed as biomarker candidate.

<https://doi.org/10.1371/journal.pone.0206478.t003>

Updating the basic protein composition of CSF

In the last two decades, several studies investigated and expanded our knowledge of the CSF proteome [15, 20, 25]. By comparing our results to those of the three most relevant publications, we could confirm the previous findings and further expand them.

First, the proteins detected in the CSF core proteome were compared to the proteins in Table 3 of Hühmer *et al.*; the latter proteins included 90 proteins that were measured in two or more independent compositional CSF studies with gel-based analysis or gel-free LC/MS approaches [15]. In the CSF core proteome, 70 of these 90 proteins were also detected. Furthermore, 2 proteins listed by Hühmer *et al.* were obsolete in the UniProt database (P01028 and P62988; gene symbols n.a.). Of the 18 missing proteins, 8 were not identified in the CSF core proteome: O14791 (APOL1), P09211 (GSTP1), P35542 (SAA4), P60709 (ACTB), P68871 (HBB), P69905 (HBA1), Q06830 (PRDX1) and Q9NQX5 (NPDC1). Ten could not be identified in our complete data set of 791 proteins.

Second, the proteins identified in the CSF core proteome were compared to a total of 126 CSF proteins identified by Stoop *et al.* [20]. The results showed that 114 proteins of these proteins were also detected in the CSF core proteome. However, 11 of the missing 12 proteins were identified in our data set but not in 100% of our samples. One protein was identified as a contamination, and one protein was obsolete in the UniProt database (P62988; gene symbol n.a.).

Third, the proteins detected in the CSF core proteome were compared to 78 brain tissue-enriched and group-specific candidates, as provided in Supplemental Table 5 (SM5) within Begcevic *et al.* [25]. The results showed that 20 of these 78 proteins were identified in our CSF

core proteome. However, 38 of the missing proteins were identified in our data set but again not in 100% of our samples.

In summary, our results are in accordance with the literature. Thus, we propose a summarizing CSF core proteome of 244 proteins in healthy human individuals ([S4 File](#)). This list is based on proteins identified by our group, Stopp *et al.*, and Hühmer *et al.*

General view— CV_a , CV_t and CV_g

First, we determined the analytical variability by measuring pooled samples during the whole measuring sequence of the 36 samples. It is recommended to assess analytical variability because it may influence statistical values, such as fold changes or p -values, may distort other statistical methods, such as subgroup detection [53], and should be as small as possible. Moreover, CV_a should be definitely lower than CV_t or CV_g to enable the possibility of detecting differences in protein abundance between study groups. As reported above, we determined an average CV_a of 8.49%. Generally, high CV_a values should be critically reviewed because imprecision in the analytical workflow may be responsible for high variability in protein abundance. Therefore, we assume that CV_t and CV_g values higher than 8.49% were caused by additional biological variation.

As reported above, for the CSF core proteome, CV_g values up to 101.5% were observed. These results are in line with the findings of Stoop *et al.*, who assessed the interindividual variability of 126 CSF proteins [30] and concluded that brain-specific proteins are more likely to have a higher CV_g than blood proteins [30]. However, the criteria for whether a protein is blood or brain specific are not described in detail. In our study, blood-specific proteins also exhibited high CV_g values (e.g., apolipoprotein a-II, fibrinogen beta chain and Ig alpha-2 chain C region) ([S1 File](#)). From our point of view, the origin of a protein is not the determining factor for its interindividual variability (CV_g). The range of CV_g for the 216 proteins measured in our study is relatively wide (9.3% to 101.5%). We checked our data to determine whether processed peptides are responsible for some high CV_g s. After comparison of single peptides between the samples, we found no indication that there are possible processed peptides or at least small peptides that behave differently across time or between subjects. Furthermore, we examined whether age influences the variability. We did not find any relationship between these two variables (see [S1 Fig](#)). Thus, we anticipate that interindividual variability negatively influences the outcome of biomarker discovery studies, indicating that CV_g should be taken into account when rating changes in protein abundance between study groups. For some proteins with high CV_g values, only substantial changes might be relevant.

As reported above, in our data set, we observed that intraindividual variability (CV_t) ranged between 2.6% and 29.3%. Our results show a certain degree of variability within an individual over a four-year period. This finding indicates that CSF proteins vary in their abundance within one individual to different degrees.

In summary, CV_t is usually lower than CV_g , which is in line with Fraser *et al.* [27], and the variability of a given protein is mainly influenced by its CV_g value. Consequently, a biomarker with low CV_t and CV_g values may be useful for reflecting changes in clinical status. Ideally, the variability in both the disease group and the control group would be low, with a large difference between the means of the groups. A higher variability leads to a decreased power to find changes in protein abundance between healthy and diseased subjects. As a consequence, the results for potential biomarkers with higher CV_g values are likely to be less reproducible in further studies. Furthermore, large sample sizes are needed due to the high variability, which might not be feasible for rare diseases. To evaluate the relevance of changes, we calculated the reference change value for each protein within subjects (RCV_t) and between subjects (RCV_g)

to assess whether changes in protein abundances are expected or unexpected. As reported above, in our CSF core proteome, we found RCV_t values up to 81.3% and RCV_g values up to 281.4%. Due to these high abundance variabilities, for CSF proteome studies, we recommend calculating the RCV_t and RCV_g values for a meaningful interpretation of detected protein changes [24]. Moreover, it is possible to estimate a fold change threshold in order to decide whether a given fold change is expectedly high after taking into account its specific natural variability or whether it is unexpectedly high and a potential biomarker candidate. To this end, we suggest calculating RCV_g for each quantified protein and computing the “maximum fold change assessment value”, FC_{max}^* , as follows:

$$FC_{max} = \max\{A/B, B/A\} \tag{8}$$

where FC_{max} = maximum fold change, A and B = mean protein abundances

$$FC_{max}^* = FC_{max} - \left(\frac{RCV_g}{100}\right) - 1 \tag{9}$$

where FC_{max}^* = maximum fold change assessment value; RCV_g = interindividual reference change value

We interpret FC_{max}^* as an unexpectedly high change if $FC_{max}^* > 0$. Otherwise, the change in protein abundance is probably mainly caused by protein-specific natural variability.

In progression marker or clinical improvement studies, RCV_t should be used instead of RCV_g to compute the above assessment value because in these cases, the intraindividual variability is more important.

Closer look–Focus on proposed protein biomarkers

To evaluate whether high CV_t , CV_g , and RCV values might be a cause for the lack of reproducibility of CSF protein biomarkers in neurological diseases (PD, AD, MS, HD and others), we focused on published data from the last few years and generated a list of 20 proposed CSF protein biomarkers. These proteins were detected in studies with mass spectrometric approaches and were found in our CSF core proteome. As reported above, the maximum CV_t value for these 20 proteins was 29.3%. The protein with the highest CV_g was haptoglobin (CV_g : 101.5%), and clusterin had the lowest value (CV_g : 12.8%). Compared to the CV_t values, the CV_g values were usually higher. These results indicate that the published biomarkers have a certain degree of variability, which is different for each protein. To examine whether the changes detected in these studies are expectedly high, we calculated RCV_g for each of the 20 proteins. The resulting RCV_g values ranged from 35.4% to 281.4%. Most of the proposed biomarker candidates had an RCV_g value above 50% in our study. This value would imply $FC_{max}^* = 0.5$ in the worst case of $FC_{max} = 1$. Consequently, a fold change of at least 1.5 is needed for an FC_{max}^* value above 0, representing an unexpectedly high change in abundance. As another example, haptoglobin showed an RCV_g value of 281.4% in our study, which would indicate an FC_{max}^* value of 2.814 if $FC_{max} = 1$. Thus, the fold change in haptoglobin must be larger than 3.814; otherwise, there is no unexpectedly high change in protein abundance. On the other hand, proteins that exhibit a low RCV_g (i.e., under 100%) might be better suited as biomarker candidates. As a consequence, some of the 20 proposed protein biomarkers might not have a relevant change in protein abundance between the study groups. Meijers *et al.* showed that similar assumptions are true for established protein markers measured in routine laboratory tests in the context of chronic heart failure prognosis [29].

Strengths and weaknesses

In our study, we assessed the variability in the CSF proteome within neurologically healthy control subjects over a period of four years. Our identified CSF proteins correspond to the literature [15, 20]. The concordance with Hühmer *et al.* and Stoop *et al.* ensures the good reproducibility, high confidence, and transferability of our defined CSF core proteome. This proteome may serve as a reference list for future CSF proteome studies. To our knowledge, this was the first study to investigate the variability in the CSF proteome by employing bottom-up mass spectrometry in order to identify and quantify as many CSF proteins as possible and by using the RCV for fold change assessment. A targeted MS approach would have allowed us to focus on single biomarker candidates. However, we preferred to gain a broader picture and provide data for future biomarker studies. To this end, we used a label-free MS approach. Owing to our study design, we focused on the variability in protein abundance in neurologically unimpaired healthy control subjects. Further studies must evaluate whether the variability within diseased groups might be comparable to our data. It can be assumed that the variability is strongly dependent on the disease. Furthermore, we suggest investigating whether the variability assessed over shorter periods is similar to our results. However, CSF sampling within short periods can increase the risk of adverse side effects for the subjects. We chose longer periods on the order of years because the long-term behavior of biomarkers is clinically important. These biomarkers should reflect the course of the disease. For an alternative to the RCV for evaluating “unexpectedly high” changes in protein abundance, it might also be reasonable to conduct a receiver operating characteristic (ROC) analysis for biomarker studies. By considering sensitivity and specificity, the ability of a potential biomarker to discriminate between two patient groups can be better investigated. Furthermore, the interrelation between abundance variability and fold changes is also considered in statistical test theory. For example, the t-test is a generally accepted standard method to assess reliable biomarker candidates on the basis of group variances and the difference in group means. Nevertheless, the RCV-based assessment suggested here may be more intuitive and better interpretable for most researchers.

Conclusions

We investigated whether high variability in CSF proteins may be a reason that only a limited number of protein biomarker candidates could be validated. In conclusion, our results indicate that high intra- and interindividual variability of a protein might lead to the identification of unsuitable biomarker candidates for multiple diseases. Thus, natural variability has a considerable impact on CSF protein biomarker research and must be taken into account in order to generate relevant results. The degree of variability in protein abundance is strongly protein dependent. Therefore, the variability should not be estimated over all proteins but instead calculated for each protein specifically. To this end, we recommend calculating the analytical variability, CV_a , as well as the intraindividual variability, CV_v , and interindividual variability, CV_g , for each protein. Then, to assess whether the change in protein abundance between experimental groups might be an unexpectedly high change indicating biological relevance, we further recommend calculating RCV_g and RCV_v , depending on the study design.

Supporting information

S1 Table. Neuropsychological characteristics of all participants and the representative study group analyzed in the present study. The analyzed representative study group reflects the total study group in terms of age and neuropsychological characteristics (e.g., Beck Depression Inventory (BDI) and Mini-Mental State Examination (MMSE)).
(DOCX)

S2 Table. Blood and routine laboratory analysis for all participants and the representative study group analyzed in the present study. For all participants, clinically important blood parameters (e.g., red and white blood cell count, electrolyte concentrations and function parameters) were determined to ensure a neurologically healthy control group. The representative study group analyzed in the present study reflects the total study groups in terms of their blood analysis.

(DOCX)

S3 Table. Coefficient of variation (CV) and reference change value (RCV) of the top 50 cerebrospinal fluid (CSF) core proteins in terms of label-free quantification (LFQ) intensity. The top 50 proteins of all samples were defined by their mean LFQ intensity. This table shows their protein names, UniProt ID, and intra- (CV_t) and interindividual (CV_g) coefficient.

(DOCX)

S1 File. CSF protein list (223). All listed proteins were identified with at least one unique peptide in 100% of the samples. The file provides information regarding protein IDs, protein names, gene names, numbers of unique peptides, and LFQ intensities, as well as CV_t , CV_g , RCV_t and RCV_g , for every protein.

(XLSX)

S2 File. CSF protein list (610). All listed proteins were identified with at least two unique peptides in at least one sample. The file provides information regarding protein IDs, protein names, gene names, numbers of unique peptides and LFQ intensities for every sample.

(XLSX)

S3 File. CSF core proteome (216). All listed proteins were identified with at least two unique peptides in 100% of the samples. The file provides information regarding protein IDs, protein names, gene names, numbers of unique peptides, and LFQ intensities, as well as CV_t , CV_g , RCV_t and RCV_g , for every protein.

(XLSX)

S4 File. CSF compositional proteins in healthy human individuals. Merging of the strict protein list ([S1 File](#)) and the extended table of proteins provided by *Hühmer et al.* and *Stoop et al.* The file provides information regarding protein IDs and protein names, as well as which proteins were listed by *Hühmer et al.* or *Stoop et al.* and were simultaneously identified in our samples.

(XLSX)

S1 Fig. CV_t box plots. This graph shows the $\log_2 CV_t$ values for every subject depending on the subject's age.

(TIF)

Acknowledgments

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [1] partner repository with the dataset identifier PXD007555.

Author Contributions

Conceptualization: Brit Mollenhauer, Katrin Marcus, Caroline May.

Data curation: Lukas M. Schilde, Steffen Kösters, Simone Steinbach, Sara Galozzi, Michael Turewicz, Katalin Barkovits.

Funding acquisition: Martin Eisenacher, Brit Mollenhauer, Katrin Marcus, Caroline May.

Investigation: Lukas M. Schilde, Steffen Kösters, Simone Steinbach, Sara Galozzi, Michael Turewicz, Katalin Barkovits, Caroline May.

Methodology: Lukas M. Schilde, Steffen Kösters, Simone Steinbach, Sara Galozzi, Michael Turewicz, Caroline May.

Project administration: Brit Mollenhauer, Katrin Marcus, Caroline May.

Resources: Brit Mollenhauer, Katrin Marcus.

Software: Martin Eisenacher, Katrin Marcus.

Validation: Lukas M. Schilde, Steffen Kösters, Simone Steinbach, Karin Schork, Sara Galozzi, Michael Turewicz, Katalin Barkovits, Caroline May.

Writing – original draft: Lukas M. Schilde, Steffen Kösters, Simone Steinbach, Michael Turewicz, Katrin Marcus, Caroline May.

Writing – review & editing: Lukas M. Schilde, Steffen Kösters, Simone Steinbach, Karin Schork, Martin Eisenacher, Sara Galozzi, Michael Turewicz, Katalin Barkovits, Brit Mollenhauer, Katrin Marcus, Caroline May.

References

1. Reiber H. Dynamics of brain-derived proteins in cerebrospinal fluid. *Clin Chim Acta*. 2001; 310(2):173–86. Epub 2001/08/11. PMID: [11498083](#).
2. Reiber H. Proteins in cerebrospinal fluid and blood: barriers, CSF flow rate and source-related dynamics. *Restor Neurol Neurosci*. 2003; 21(3–4):79–96. Epub 2003/10/08. PMID: [14530572](#).
3. Simon MJ, Iliff JJ. Regulation of cerebrospinal fluid (CSF) flow in neurodegenerative, neurovascular and neuroinflammatory disease. *Biochim Biophys Acta*. 2016; 1862(3):442–51. Epub 2015/10/27. <https://doi.org/10.1016/j.bbadis.2015.10.014> PMID: [26499397](#); PubMed Central PMCID: PMC4755861.
4. Spector R, Robert Snodgrass S, Johanson CE. A balanced view of the cerebrospinal fluid composition and functions: Focus on adult humans. *Exp Neurol*. 2015; 273:57–68. Epub 2015/08/08. <https://doi.org/10.1016/j.expneurol.2015.07.027> PMID: [26247808](#).
5. Gomez-Rio M, Caballero MM, Gorris Saez JM, Minguez-Castellanos A. Diagnosis of Neurodegenerative Diseases: The Clinical Approach. *Curr Alzheimer Res*. 2016; 13(5):469–74. Epub 2015/11/17. PMID: [26567736](#).
6. Kalia LV, Lang AE. Parkinson's disease. *Lancet*. 2015; 386(9996):896–912. Epub 2015/04/24. [https://doi.org/10.1016/S0140-6736\(14\)61393-3](https://doi.org/10.1016/S0140-6736(14)61393-3) PMID: [25904081](#).
7. Schrag A, Horsfall L, Walters K, Noyce A, Petersen I. Prediagnostic presentations of Parkinson's disease in primary care: a case-control study. *Lancet Neurol*. 2015; 14(1):57–64. Epub 2014/12/02. [https://doi.org/10.1016/S1474-4422\(14\)70287-X](https://doi.org/10.1016/S1474-4422(14)70287-X) PMID: [25435387](#).
8. Thal LJ, Kantarci K, Reiman EM, Klunk WE, Weiner MW, Zetterberg H, et al. The Role of Biomarkers in Clinical Trials for Alzheimer Disease. *Alzheimer Dis Assoc Disord*. 2006; 20(1):6–15. <https://doi.org/10.1097/01.wad.0000191420.61260.a8> PMID: [16493230](#); PubMed Central PMCID: PMC1820855.
9. Doherty CM, Forbes RB. Diagnostic Lumbar Puncture. *Ulster Med J*. 2014; 83(2):93–102. PMID: [25075138](#); PubMed Central PMCID: PMC4113153.
10. Olsson B, Lautner R, Andreasson U, Ohrfelt A, Portelius E, Bjerke M, et al. CSF and blood biomarkers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. *Lancet Neurol*. 2016; 15(7):673–84. Epub 2016/04/14. [https://doi.org/10.1016/S1474-4422\(16\)00070-3](https://doi.org/10.1016/S1474-4422(16)00070-3) PMID: [27068280](#).
11. Babic M, Svob Strac D, Muck-Seler D, Pivac N, Stanic G, Hof PR, et al. Update on the core and developing cerebrospinal fluid biomarkers for Alzheimer disease. *Croat Med J*. 2014; 55(4):347–65. Epub 2014/08/29. <https://doi.org/10.3325/cmj.2014.55.347> PMID: [25165049](#); PubMed Central PMCID: PMC4157375.

12. Halbgebauer S, Ockl P, Wirth K, Steinacker P, Otto M. Protein biomarkers in Parkinson's disease: Focus on cerebrospinal fluid markers and synaptic proteins. *Mov Disord.* 2016; 31(6):848–60. Epub 2016/05/03. <https://doi.org/10.1002/mds.26635> PMID: 27134134.
13. Matejckikova Z, Mares J, Prikrylova Vranova H, Klosova J, Sladkova V, Dolakova J, et al. Cerebrospinal fluid inflammatory markers in patients with multiple sclerosis: a pilot study. *J Neural Transm (Vienna).* 2015; 122(2):273–7. Epub 2014/06/05. <https://doi.org/10.1007/s00702-014-1244-9> PMID: 24894698.
14. Ross CA, Aylward EH, Wild EJ, Langbehn DR, Long JD, Warner JH, et al. Huntington disease: natural history, biomarkers and prospects for therapeutics. *Nat Rev Neurol.* 2014; 10(4):204–16. Epub 2014/03/13. <https://doi.org/10.1038/nrneurol.2014.24> PMID: 24614516.
15. Huhmer AF, Biringer RG, Amato H, Fonteh AN, Harrington MG. Protein analysis in human cerebrospinal fluid: Physiological aspects, current progress and future challenges. *Dis Markers.* 2006; 22(1–2):3–26. Epub 2006/01/18. <https://doi.org/10.1155/2006/158797> PMID: 16410649; PubMed Central PMCID: PMCPMC3850820.
16. Kang JH, Mollenhauer B, Coffey CS, Toledo JB, Weintraub D, Galasko DR, et al. CSF biomarkers associated with disease heterogeneity in early Parkinson's disease: the Parkinson's Progression Markers Initiative study. *Acta Neuropathol.* 2016; 131(6):935–49. Epub 2016/03/30. <https://doi.org/10.1007/s00401-016-1552-2> PMID: 27021906; PubMed Central PMCID: PMCPMC5031365.
17. Hall S, Ohrfelt A, Constantinescu R, Andreasson U, Surova Y, Bostrom F, et al. Accuracy of a panel of 5 cerebrospinal fluid biomarkers in the differential diagnosis of patients with dementia and/or parkinsonian disorders. *Arch Neurol.* 2012; 69(11):1445–52. Epub 2012/08/29. <https://doi.org/10.1001/archneurol.2012.1654> PMID: 22925882.
18. Mollenhauer B, Trautmann E, Sixel-Doring F, Wicke T, Ebentheuer J, Schaumburg M, et al. Nonmotor and diagnostic findings in subjects with de novo Parkinson disease of the DeNoPa cohort. *Neurology.* 2013; 81(14):1226–34. Epub 2013/09/03. <https://doi.org/10.1212/WNL.0b013e3182a6cbd5> PMID: 23997153.
19. Byrne LM, Wild EJ. Cerebrospinal Fluid Biomarkers for Huntington's Disease. *J Huntingtons Dis.* 2016; 5(1):1–13. Epub 2016/04/01. <https://doi.org/10.3233/JHD-160196> PMID: 27031730.
20. Stoop MP, Coulier L, Rosenling T, Shi S, Smolinska AM, Buydens L, et al. Quantitative proteomics and metabolomics analysis of normal human cerebrospinal fluid samples. *Mol Cell Proteomics.* 2010; 9(9):2063–75. Epub 2010/09/03. <https://doi.org/10.1074/mcp.M900877-MCP200> PMID: 20811074; PubMed Central PMCID: PMCPMC2938111.
21. Rosen C, Hansson O, Blennow K, Zetterberg H. Fluid biomarkers in Alzheimer's disease—current concepts. *Mol Neurodegener.* 2013; 8:20. Epub 2013/06/27. <https://doi.org/10.1186/1750-1326-8-20> PMID: 23800368; PubMed Central PMCID: PMCPMC3691925.
22. Simonsen AH, Herukka SK, Andreassen N, Baldeiras I, Bjerke M, Blennow K, et al. Recommendations for CSF AD biomarkers in the diagnostic evaluation of dementia. *Alzheimers Dement.* 2017; 13(3):274–84. Epub 2017/03/28. <https://doi.org/10.1016/j.jalz.2016.09.008> PMID: 28341065.
23. Kalia LV, Kalia SK, Lang AE. Disease-modifying strategies for Parkinson's disease. *Mov Disord.* 2015; 30(11):1442–50. Epub 2015/07/25. <https://doi.org/10.1002/mds.26354> PMID: 26208210.
24. Hu Y, Malone JP, Fagan AM, Townsend RR, Holtzman DM. Comparative proteomic analysis of intra- and interindividual variation in human cerebrospinal fluid. *Mol Cell Proteomics.* 2005; 4(12):2000–9. Epub 2005/10/04. <https://doi.org/10.1074/mcp.M500207-MCP200> PMID: 16199891.
25. Begcevic I, Brinc D, Drabovich AP, Batruch I, Diamandis EP. Identification of brain-enriched proteins in the cerebrospinal fluid proteome by LC-MS/MS profiling and mining of the Human Protein Atlas. *Clin Proteomics.* 2016; 13:11. Epub 2016/05/18. <https://doi.org/10.1186/s12014-016-9111-3> PMID: 27186164; PubMed Central PMCID: PMCPMC4868024.
26. Trombetta BA, Carlyle BC, Koenig AM, Shaw LM, Trojanowski JQ, Wolk DA, et al. The technical reliability and biotemporal stability of cerebrospinal fluid biomarkers for profiling multiple pathophysiologies in Alzheimer's disease. *PLoS One.* 2018; 13(3):e0193707. Epub 2018/03/06. <https://doi.org/10.1371/journal.pone.0193707> PMID: 29505610; PubMed Central PMCID: PMCPMC5837100.
27. Fraser CG. Inherent biological variation and reference values. *Clin Chem Lab Med.* 2004; 42(7):758–64. Epub 2004/08/26. <https://doi.org/10.1515/CCLM.2004.128> PMID: 15327011.
28. Fraser CG, Harris EK. Generation and application of data on biological variation in clinical chemistry. *Crit Rev Clin Lab Sci.* 1989; 27(5):409–37. Epub 1989/01/01. <https://doi.org/10.3109/10408368909106595> PMID: 2679660.
29. Meijers WC, van der Velde AR, Muller Kobold AC, Dijck-Brouwer J, Wu AH, Jaffe A, et al. Variability of biomarkers in patients with chronic heart failure and healthy controls. *Eur J Heart Fail.* 2017; 19(3):357–65. Epub 2016/10/22. <https://doi.org/10.1002/ehfj.669> PMID: 27766733; PubMed Central PMCID: PMCPMC5347881.

30. Plum S, Helling S, Theiss C, Leite REP, May C, Jacob-Filho W, et al. Combined enrichment of neuromelanin granules and synaptosomes from human substantia nigra pars compacta tissue for proteomic analysis. *Journal of Proteomics*. 2013; 94:202–6. <https://doi.org/10.1016/j.jprot.2013.07.015> WOS:000330493400015. PMID: 23917253
31. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*. 2008; 26(12):1367–72. <https://doi.org/10.1038/nbt.1511> WOS:000261591300022. PMID: 19029910
32. UniProt: a hub for protein information. *Nucleic Acids Res*. 2015; 43(Database issue):D204–12. Epub 2014/10/29. <https://doi.org/10.1093/nar/gku989> PMID: 25348405; PubMed Central PMCID: PMC4384041.
33. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *Journal of Proteome Research*. 2011; 10(4):1794–805. <https://doi.org/10.1021/pr101065j> WOS:000288924000033. PMID: 21254760
34. Michalski A, Cox J, Mann M. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J Proteome Res*. 2011; 10(4):1785–93. Epub 2011/02/12. <https://doi.org/10.1021/pr101060v> PMID: 21309581.
35. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Molecular & Cellular Proteomics*. 2014; 13(9):2513–26. <https://doi.org/10.1074/mcp.M113.031591> WOS:000341785800025. PMID: 24942700
36. Turewicz M, Kohl M, Ahrens M, Mayer G, Uszkoreit J, Naboulsi W, et al. BioInfra.Prot: A comprehensive proteomics workflow including data standardization, protein inference, expression analysis and data publication. *J Biotechnol*. 2017; 261:116–25. Epub 2017/06/14. <https://doi.org/10.1016/j.jbiotec.2017.06.005> PMID: 28606611.
37. Abdi F, Quinn JF, Jankovic J, McIntosh M, Leverenz JB, Peskind E, et al. Detection of biomarkers with a multiplex quantitative proteomic platform in cerebrospinal fluid of patients with neurodegenerative disorders. *J Alzheimers Dis*. 2006; 9(3):293–348. Epub 2006/08/18. PMID: 16914840.
38. Yin GN, Lee HW, Cho JY, Suk K. Neuronal pentraxin receptor in cerebrospinal fluid as a potential biomarker for neurodegenerative diseases. *Brain Res*. 2009; 1265:158–70. Epub 2009/04/17. <https://doi.org/10.1016/j.brainres.2009.01.058> PMID: 19368810.
39. Comabella M, Fernandez M, Martin R, Rivera-Vallve S, Borrás E, Chiva C, et al. Cerebrospinal fluid chitinase 3-like 1 levels are associated with conversion to multiple sclerosis. *Brain*. 2010; 133(Pt 4):1082–93. Epub 2010/03/20. <https://doi.org/10.1093/brain/awq035> PMID: 20237129.
40. Fang Q, Strand A, Law W, Faca VM, Fitzgibbon MP, Hamel N, et al. Brain-specific Proteins Decline in the Cerebrospinal Fluid of Humans with Huntington Disease*S. *Mol Cell Proteomics*. 82009. p. 451–66.
41. Kroksveen AC, Guldbrandsen A, Vedeler C, Myhr KM, Opsahl JA, Berven FS. Cerebrospinal fluid proteome comparison between multiple sclerosis patients and controls. *Acta Neurol Scand Suppl*. 2012; (195):90–6. Epub 2013/01/04. <https://doi.org/10.1111/ane.12029> PMID: 23278663.
42. de Bont JM, den Boer ML, Reddingius RE, Jansen J, Passier M, van Schaik RH, et al. Identification of apolipoprotein A-II in cerebrospinal fluid of pediatric brain tumor patients by protein expression profiling. *Clin Chem*. 2006; 52(8):1501–9. Epub 2006/06/10. <https://doi.org/10.1373/clinchem.2006.069294> PMID: 16762995.
43. Ward MA, Proteome Sciences plc L, Soininen H, University of Kuopio K, O'Brien D, Proteome Sciences plc L, et al. Evaluation of CSF cystatin C, beta-2-microglobulin, and VGF as diagnostic biomarkers of Alzheimer's disease using SRM. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*. 2011; 7(4). <https://doi.org/10.1016/j.jalz.2011.05.405>
44. Heywood WE, Galimberti D, Bliss E, Sirka E, Paterson RW, Magdalino NK, et al. Identification of novel CSF biomarkers for neurodegeneration and their validation by a high-throughput multiplexed targeted proteomic assay. *Mol Neurodegener*. 2015; 10:64. Epub 2015/12/03. <https://doi.org/10.1186/s13024-015-0059-y> PMID: 26627638; PubMed Central PMCID: PMC4666172.
45. Paterson RW, Heywood WE, Heslegrave AJ, Magdalino NK, Andreasson U, Sirka E, et al. A targeted proteomic multiplex CSF assay identifies increased malate dehydrogenase and other neurodegenerative biomarkers in individuals with Alzheimer's disease pathology. *Transl Psychiatry*. 2016; 6(11):e952. Epub 2016/11/16. <https://doi.org/10.1038/tp.2016.194> PMID: 27845782; PubMed Central PMCID: PMC4666172.
46. Jia Y, Wu T, Jelinek CA, Bielekova B, Chang L, Newsome S, et al. Development of protein biomarkers in cerebrospinal fluid for secondary progressive multiple sclerosis using selected reaction monitoring mass spectrometry (SRM-MS). *Clin Proteomics*. 92012. p. 9.
47. Magdalino NK, Noyce AJ, Pinto R, Lindstrom E, Holmen-Larsson J, Holtta M, et al. Identification of candidate cerebrospinal fluid biomarkers in parkinsonism using quantitative proteomics. *Parkinsonism*

- Relat Disord. 2017; 37:65–71. Epub 2017/02/19. <https://doi.org/10.1016/j.parkreldis.2017.01.016> PMID: 28214264.
48. Vafadar-Isfahani B, Ball G, Coveney C, Lemetre C, Boocock D, Minthon L, et al. Identification of SPARC-like 1 protein as part of a biomarker panel for Alzheimer's disease in cerebrospinal fluid. *J Alzheimers Dis.* 2012; 28(3):625–36. Epub 2011/11/03. <https://doi.org/10.3233/JAD-2011-111505> PMID: 22045497.
 49. Kroksveen AC, Aasebo E, Vethe H, Van Pesch V, Franciotta D, Teunissen CE, et al. Discovery and initial verification of differentially abundant proteins between multiple sclerosis patients and controls using iTRAQ and SID-SRM. *J Proteomics.* 2013; 78:312–25. Epub 2012/10/13. <https://doi.org/10.1016/j.jprot.2012.09.037> PMID: 23059536.
 50. Stoop MP, Dekker LJ, Titulaer MK, Burgers PC, Sillevius Smitt PA, Luider TM, et al. Multiple sclerosis-related proteins identified in cerebrospinal fluid by advanced mass spectrometry. *Proteomics.* 2008; 8(8):1576–85. Epub 2008/03/21. <https://doi.org/10.1002/pmic.200700446> PMID: 18351689.
 51. Schutzer SE, Liu T, Natelson BH, Angel TE, Schepmoes AA, Purvine SO, et al. Establishing the Proteome of Normal Human Cerebrospinal Fluid. *PLoS One.* 52010.
 52. Perrin RJ, Payton JE, Malone JP, Gilmore P, Davis AE, Xiong C, et al. Quantitative Label-Free Proteomics for Discovery of Biomarkers in Cerebrospinal Fluid: Assessment of Technical and Inter-Individual Variation. *PLoS One.* 82013.
 53. Ahrens M, Turewicz M, Casjens S, May C, Pesch B, Stephan C, et al. Detection of patient subgroups with differential expression in omics data: a comprehensive comparison of univariate measures. *PLoS One.* 2013; 8(11):e79380. Epub 2013/11/28. <https://doi.org/10.1371/journal.pone.0079380> PMID: 24278130; PubMed Central PMCID: PMC3838370.