Resource

TRANSPARENT PROCESS OPEN ACCESS

THE EMBO JOURNAL

A region-resolved proteomic map of the human brain enabled by high-throughput proteomics

Johanna Tüshaus[1], Amirhossein Sakhteman[1], Severin Lechner[1], Matthew The[1], Eike Mucha[2], Christoph Krisp[2], Jürgen Schlegel[3], Claire Delbridge[3] & Bernhard Kuster[1,4,*]

Abstract

Substantial efforts are underway to deepen our understanding of human brain morphology, structure, and function using high-resolution imaging as well as high-content molecular profiling technologies. The current work adds to these approaches by providing a comprehensive and quantitative protein expression map of 13 anatomically distinct brain regions covering more than 11,000 proteins. This was enabled by the optimization, characterization, and implementation of a high-sensitivity and high-throughput microflow liquid chromatography timsTOF tandem mass spectrometry system (LC–MS/MS) capable of analyzing more than 2,000 consecutive samples prepared from formalin-fixed paraffin embedded (FFPE) material. Analysis of this proteomic resource highlighted brain region-enriched protein expression patterns and functional protein classes, protein localization differences between brain regions and individual markers for specific areas. To facilitate access to and ease further mining of the data by the scientific community, all data can be explored online in a purpose-built R Shiny app (https://brain-region-atlas.proteomics.ls.tum.de).

Keywords brain proteome atlas; brain regions; high-throughput proteomics; micro-flow LC; timsTOF

Subject Categories Neuroscience; Proteomics

DOI 10.15252/embj.2023114665 | Received 3 June 2023 | Revised 16 October 2023 | Accepted 17 October 2023 | Published online 2 November 2023

The EMBO Journal (2023) 42: e114665

Introduction

Around 86 billion neurons along with a similar number of glial cells make up the most complex organ in the human body—the brain (Azevedo *et al*, 2009). The functions of the brain are highly diverse and include the control of motion, the processing of sensory information, learning, and memory formation to name a few. To fulfill these complex tasks, the brain exhibits a highly organized substructure of anatomically distinct, but well-connected brain regions. Tremendous efforts have been expended to disentangle the molecular differences of the brain regions as well as mapping their connectivity. Consortia such as the Human Protein Atlas (HPA) (Sjöstedt *et al*, 2020), the PsychENCODE project (Akbarian *et al*, 2015), the Allen brain project (Lein *et al*, 2007), and the Human brain project (HBP) (Amunts *et al*, 2019) published detailed atlases of the human brain visualizing molecular differences in a spatial dimension based on RNA-seq, MRI, in situ hybridization, and (immuno)histochemistry data. However, a comprehensive proteomic profile of the brain regions has been lacking; arguably, a substantial gap as proteins are the major functional executors of cellular processes and are the targets of almost all neurological drugs. The few proteomic studies performed so far were either focused on the mouse brain (Sharma *et al*, 2015; Distler *et al*, 2020), used pooled human tissue samples (Carlyle *et al*, 2017), were confined to one specific brain region (Guo *et al*, 2022), or suffered from limited proteome coverage (Carlyle *et al*, 2017; Biswas *et al*, 2021; Melliou *et al*, 2022). In part, this may be due to the technical hurdles involved in performing deep proteome profiling at scale. While tremendous technical improvements have been made over the years, scaling the technology to the analysis of large numbers of samples has only recently come into focus. One way of achieving greater throughput while maintaining high data quality is the use of higher chromatographic flow rates (Bian *et al*, 2020, 2021a, 2021b; Messner *et al*, 2021). We and others have demonstrated that this enables the analysis of thousands of proteomes at a moderate loss of sensitivity. To take full advantage of such improvements in peptide separation technology, mass spectrometers capable of generating data at a very rapid rate and maintaining sensitivity at the same time are required. This has, for instance, been achieved by combining trapped ion mobility and time-of-flight mass spectrometry (timsTOF), to enable the parallel accumulation and serial fragmentation (PASEF) of peptides, leading to data acquisition rates of more than 150 Hz and highly efficient utilization of the available peptide ions (Meier *et al*, 2015, 2018, 2021).

Here, we report on the coupling of microflow liquid chromatography (LC) to a timsTOF mass spectrometer to combine the assets of rapid and high-resolution peptide chromatography with rapid and high sensitivity mass spectrometry (LC–MS/MS). We characterize the performance of this system by analyzing diverse biological samples

1 Proteomics and Bioanalytics, Department of Molecular Life Sciences, School of Life Sciences, Technical University of Munich, Munich, Germany
2 Bruker Daltonics GmbH & Co. KG, Bremen, Germany
3 Department of Neuropathology, Klinikum Rechts der ISAR, School of Medicine, Technical University Munich, Munich, Germany
4 German Cancer Consortium (DKTK), Munich Site, Heidelberg, Germany
*Corresponding author. Tel: +49 8161 71 5696; E-mail: kuster@tum.de

© 2023 The Authors. Published under the terms of the CC BY NC ND 4.0 license

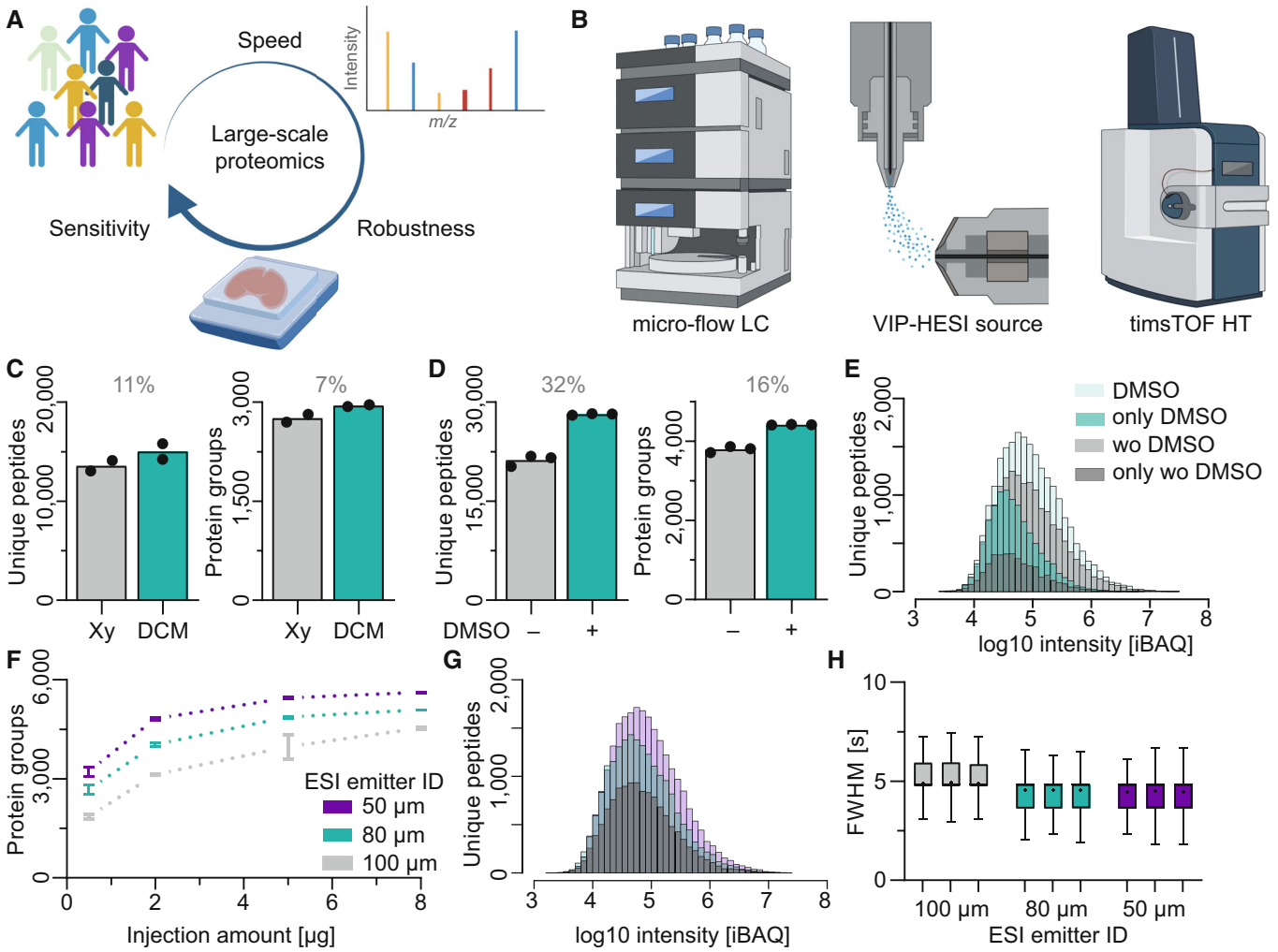

Figure 1. Establishing a sensitive, robust, and rapid LC–MS/MS setup.

A Schematic representation of the study aims: to establish a sensitive, robust, and rapid LC–MS/MS setup able to support large-scale proteomic studies.

B Coupling a microflow liquid chromatography system via a VIP-HESI ion source to a timsTOF mass spectrometer.

C Bar graph indicating the number of unique peptides and protein groups identified from FFPE human brain samples using xylene (Xy) deparaffinization with or without an additional overnight delipidation step using dichloromethane (DCM) ($N = 2$).

D Bar graph showing the number of unique peptides identified from a HeLa cell line digest in the presence or absence of 3% DMSO in LC solvents using a 30-min LC gradient and 2 µg injected digest ($N = 3$).

E iBAQ intensity distribution of peptides identified in panel (D).

F Number of protein groups identified using a 30-min LC gradient as a function of the amount of injected HeLa digest and using electrospray emitters of 50, 80, or 100 µm inner diameter (ID) ($N = 3$).

G iBAQ intensity distribution of peptides identified in panel (F) using 2 µg HeLa cell line digest loading.

H Box plots showing the chromatographic peak width distribution (full width at half maximum, FWHM) of peptides in panel (G) ($N = 3$, central plus: mean, boxes: middle 50% of the data, whiskers: Tukey).

including human cell lines, plasma, CSF, and formalin-fixed paraffin embedded (FFPE) tissue. In addition, we present an optimized end-to-end workflow ready for large-scale human FFPE brain proteome analysis and demonstrate its robustness by profiling 13 regions of the human brain to a depth of ~10,000 proteins each and representing a total of > 2,000 individual LC–MS/MS measurements. The resultant molecular resource, which is publically available online, constitutes the most comprehensive map of a human brain proteome to date and systematic data evaluation revealed distinct proteomic signatures of each brain region including new marker proteins.

Results

Optimization of a microflow LC timsTOF MS/MS setup

In order to establish a LC–MS/MS setup fulfilling the aforementioned requirements of sensitivity, robustness, and speed, we coupled a microflow LC system to a Bruker timsTOF mass spectrometer equipped with a VIP-HESI ion source (Fig 1A and B). Aiming for deep proteome coverage while restricting overall analysis time, we optimized the setup in a systematic manner including sample

 © 2023 The Authors

preparation, chromatographic, ion source, and MS instrument parameters (see Materials and Methods for details).

First, we adapted the SP3 sample preparation approach (Müller *et al*, 2020) for FFPE tissue and cerebrospinal fluid (CSF) to a 96-well format using a liquid handling platform. Protein extraction from FFPE tissue was aided by sonication using the Adaptive Focused Acoustics (AFA) technology (Green *et al*, 2014; Marchione *et al*, 2020). We also added a delipidation step using dichloromethane (DCM) to the brain tissue workflow akin to methods used for tissue clearing (Molbay *et al*, 2021). This improved chromatographic stability and enhanced protein identification by 7% at the same time (Fig 1C).

Second, 3% DMSO was added to LC solvents to boost electrospray ionization (ESI) efficiency (Hahne *et al*, 2013). This had a substantial effect on performance as it increased peptide and protein group identifications by 32 and 16%, respectively, using a standard 30-min LC gradient (Fig 1D). DMSO led to an overall improvement in peptide intensity and the peptides gained in the presence of DMSO were generally of low abundance (Figs 1E and EV1A). We note that the addition of DMSO neither negatively affects the performance of the chromatographic system nor the mass spectrometer even after months of operation and thousands of sample injections. DMSO did also not change the ion mobility characteristics of analyzed peptides (Fig EV1B).

Third, the ion source parameters were optimized including the fabrication of novel ESI emitters with smaller inner diameters (ID) of 50 and 80 μm compared with the standard emitter (100 μm ID) to improve ESI efficiency. Smaller ID emitters also improve ESI stability and ion desolvation characteristics at lower liquid flow rates (Covey *et al*, 2009). As anticipated, the 50 μm emitter outperformed the wider bore emitters irrespective of the amount of material analyzed in terms of protein and peptide identifications. This was the result of a threefold increase in peptide intensity as well as slightly narrower chromatographic peaks widths at half maximum (FWHM) (Figs 1F–H, and EV1C and D).

Fourth, MS data acquisition parameters were optimized to take advantage of the narrow LC peak widths provided by the microflow LC separations. Adjustment of the threshold parameters for MS2 scheduling resulted in a gain of 3% on protein level (Fig EV1E). Reducing the time spent per MS2 scan from the standard 4.4 ms to 3.2 ms enabled an average scheduling of 18 instead of 15 precursor ions per 100 ms ramp (Fig EV1F and G). Increasing the collision energy improved protein identification by another 3% compared with standard settings and also led to a higher fraction of MS2 spectra that lead to a peptide identification (from 62 to 68%; Fig EV1H and I). Last, we compared the timsTOF Pro2 to the timsTOF HT, the latter containing a tims-analyzer and detector designed for higher ion capacity (Appendix Fig S1). At sample loadings of up to 2 μg, the timsTOF Pro2 outperformed the timsTOF HT in terms of sensitivity. However, higher sample loads resulted in a reduced number of tryptic but an increased number of semitryptic peptides (19%). This was likely due to overloading of the timsTOF Pro2, in turn, leading to peptide fragmentation inside the tims-analyzer. The timsTOF HT showed no such effects independent of the peptide load and outperformed the timsTOF Pro2 at loadings of more than 2 μg peptide using a 30-min method. The following results were all obtained using the timsTOF HT.

Evaluation of achievable proteome coverage at different levels of sample throughput

To evaluate what proteome coverage can be achieved for different levels of sample throughput, we analyzed different amounts of HeLa cell line digests using data-dependent (DDA, MaxQuant) and data-independent (DIA, Spectronaut 17) MS methods that would allow the analysis of between 24 and 192 samples per day (60 min, respectively, 7.5-min total time from injection to injection including all overhead times for sample loading, column re-equilibration, etc.). For the highest throughput method, > 3,100 and > 5,000 proteins could be identified by using DDA or DIA, respectively, and the corresponding figures for the lowest throughput method (60 min) tested were > 6,600 and > 8,000 proteins (Figs 2A, and EV2A and C). Exemplified by data collected with the 30-min method, DDA protein identification results were almost completely contained in the DIA results (Fig 2B). In addition, most proteins identified by DDA showed a quantitative precision of below 10%. DIA quantified a similar number of proteins with the same precision and quantified > 2,500 more at still acceptable levels (up to 20% CV) (Figs 2C and EV2B).

We next extended the evaluation to human body fluids notably plasma and CSF. In plasma, proteome coverage ranged from 243 (7.5 min) to 451 (60 min) proteins groups in single-shot DDA mode (Figs 2D and EV2D). Analysis of five individual plasma samples (5 μg each; 30 min method) resulted, on average, in 333 and 362 proteins using DDA and DIA, respectively (Fig 2E). Notably, 45 (DDA) and 47 (DIA) of the 49 FDA-approved biomarkers previously identified by Geyer *et al* (2016)) were covered in the dataset with more than two unique peptides. At the same level of sample throughput, 1,283 and 1,375 proteins were, on average, identified in CSF samples in DDA and DIA mode, respectively (Figs 2F and G, and EV2E). Comparable CSF proteome coverage was recently reported using nano-flow MS/MS and LC gradients of 45–80 min (Bader *et al*, 2020; Karayel *et al*, 2022).

In a third application, we used the 30-min DDA method to profile the targets of HDAC-inhibitors using a competition binding assay previously established by the authors requiring 110 min of total nanoflow LC–MS/MS time (Lechner *et al*, 2022). Exemplified by the three HDAC inhibitors Quisinostat, TSA and Panobinostat, the microflow timsTOF setup provided very similar data quality in terms of detection of target proteins as well as drug:target interaction strength but in a fraction of total analysis time (Figs 2H–J, and EV2F and G).

Deep proteomic profiling of 13 formalin-preserved human brain regions

With an optimized, high-throughput-capable microflow LC timsTOF HT system at hand, we set out to establish a region-resolved proteomic atlas of the human brain. Thirteen brain regions were collected from one formalin-preserved postmortem human brain (Nucleus accumbens, Substantia nigra, Olives, Thalamus, Red nucleus, Hippocampus, Putamen, Claustrum, Caudate nucleus, Cerebellum and Cerebral cortex (gray & white matter of the letter two)). From each brain region, three tissue cubes (~5 × 5 × 5 mm) were collected and independently processed (two in case of olives). Following protein extraction and digestion, each of the 38 resulting peptide

© 2023 The Authors

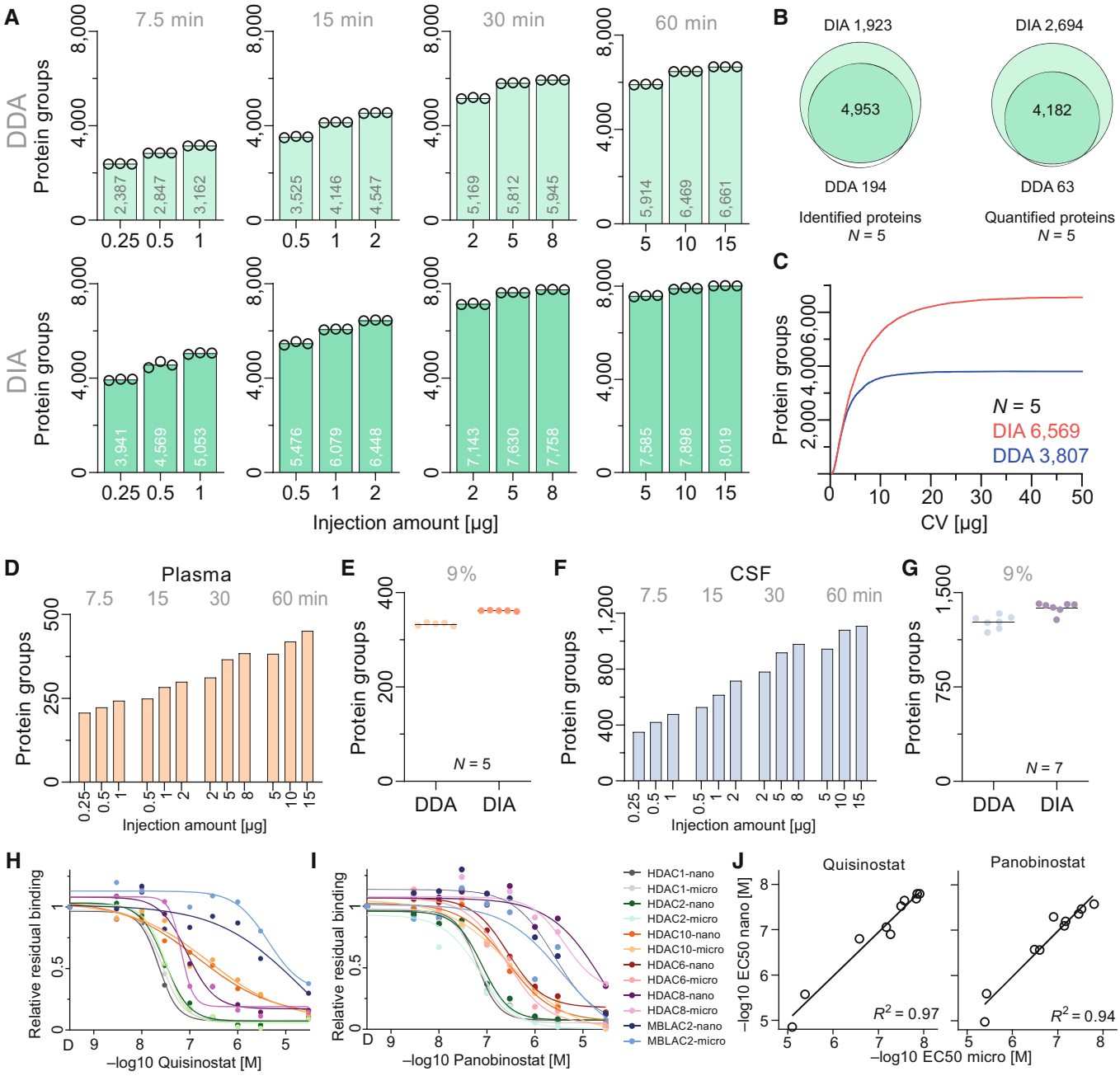

Figure 2. Benchmarking the microflow LC timsTOF HT setup.

A Number of identified proteins analyzed by data dependent acquisition (DDA) (upper panels) or data independent acquisition (DIA) (lower panels) analyzed by different LC gradients and as a function of the amount of HeLa digest injected ($N = 3$). Only proteins identified by at least two unique peptides were considered.

B Venn diagram indicating the overlap between the number of peptides identified or quantified by DDA and DIA runs (2 µg and 30-min gradient data) ($N = 5$).

C Cumulative density plot of the number of proteins as a function of quantitative precision (coefficient of variation (CV)) using protein LFQ intensities from data of panel (B) ($N = 5$).

D Same as panel (A) but for neat digested human plasma (DDA).

E Number of proteins identified from neat human plasma digest (five human individuals) injecting 5 µg digest and analyzed using a 30-min LC gradient in DDA and DIA mode.

F Same as panel (D) but for human cerebrospinal fluid (CSF).

G Same as panel (E) but for human CSF.

H Dose–response binding curves of targets of the HDAC inhibitor Quisinostat obtained by competition binding assays to HDAC beads and analyzed by micro-flow timsTOF LC–MS/MS (light colors) or nanoflow Orbitrap MS/MS (dark colors). The latter data were reproduced from Lechner *et al* (2022).

I Same as panel (H) but for Panobinostat.

J Correlation analysis of −log 10 EC 50 values of dose–response curves shown in panels (H) and (I). R, Pearson correlation coefficient.

 © 2023 The Authors

samples was separated into 48 fractions using high-pH reversed phase chromatography and analyzed by microflow LC–MS/MS using the 30-min method described above (Fig 3A). Including control samples (see below), this led to collecting data from 2,271 LC–MS/MS runs, all using the same online C18 column and the same 50 μm ESI emitter. To control for stable technical performance of the microflow LC–MS/MS setup, full proteome HeLa digests were analyzed between tissue samples ($N = 21$). Almost all proteins (98% of 2,845) and 76% of all peptides (10,488) showed CV values of below 20% (Fig 3B). The median CV was 7% on protein and 12% on peptide level. For the brain regions, the respective median protein CVs were somewhat higher (10–20%). This is due to the detection of a far greater number of proteins in each sample (> 9,000 proteins), the additional step of peptide fractionation, and the biological diversity within a brain region (Fig 3C). To measure LC performance stability, synthetic (PROCAL) peptides, designed to span the entire peptide elution spectrum across an LC gradient (Zolg *et al*, 2017), were run either alone or spiked into each tissue sample ($N = 1{,}968$). The CVs of retention times were very small (1%) and nearly identical when run alone or as spike-in (Figs 3D and EV3A).

Deep proteome coverage was achieved of all brain regions with, on average, 9,498 proteins and 98,434 peptides identified in each tissue (Fig 3E) and a total of 11,325 proteins across all regions. As a further indicator of data quality and achievable proteome coverage, three replicates of fresh frozen tissue of the cerebral cortex of a different postmortem human brain was analyzed in the same manner to estimate the losses caused by the fixation process. Proteome coverage of the fresh frozen CC material (average of 10,629 protein groups and 146,970 peptides per replicate) was ~7% deeper than that of FFPE CC material and the difference was much more pronounced at peptide level (28%). Comparing the intensities of the two samples showed that both proteomes span about six orders of magnitude in protein expression (Figs 3F and EV3B). Almost all proteins (98.6%) detected in the fixed tissue were contained in the data of the fresh frozen sample but had systematically lower intensities. Most proteins exclusively detected in the fresh frozen tissue populate the lower abundance range (87% of quartile 4), which was more pronounced than for proteins exclusively detected in the fixed material (Figs 3F and EV3C). iBAQ ratios of proteins identified in common between fresh and fixed tissue revealed an abundance-dependent pattern indicating some low abundance proteins in fresh tissue to be overrepresented in fixed tissue (Fig EV3D). According to functional clustering analysis, nuclear proteins related to transcription were overrepresented in the top 10% of proteins high in the fresh relative to fixed tissue. In contrast, the bottom 10% of proteins of low abundance in fresh relative to fixed tissue were associated with secretion and the extracellular matrix (Fig EV3E). An open modification search using FragPipe (Kong *et al*, 2017; Yu *et al*, 2020a, 2020b) returned 46% of all peptide spectrum matches (PSMs) in fresh tissue as unmodified, compared with 42% in the fixed tissue. In contrast, many more modifications that can be rationalized by the use of formaldehyde in the fixation process (Metz *et al*, 2004) were indeed found in the fixed tissue (Figs 3G and EV3F). The above data generally indicate that formaldehyde cross-linking cannot be fully reversed leading to substantial loss of peptides that can be recovered or identified from FFPE material. At the same time, these data show that certain proteins can be more efficiently extracted from FFPE than fresh tissue.

Analysis of the brain region-resolved human proteome atlas

The proteome atlas generated in this study offers a multitude of analyses only a few of which can be highlighted in the following. According to our atlas, 64% of all proteins were detected in all 13 brain regions and 6% were exclusive to a single one. Most of the exclusive proteins were found in the cerebellum (CBM), followed by cerebral cortex (CC), olives (OB), and hippocampus (HPC) (Figs 4A, and EV4A and B). Interestingly, as many as 5,750 of the 9,719 proteins in the atlas were significantly differentially expressed between brain regions (ANOVA, followed by pairwise *T*-tests for multiple comparisons of independent brain regions) (Dataset EV1). Following a concept adopted from the Human Protein Atlas (HPA) (Sjöstedt *et al*, 2020), we defined four regional protein enrichment classifications. Class 1 proteins (593) were only detected in a single brain region, Class 2 (218) includes proteins that are at least fourfold enriched in one brain region compared to all others, Class 3 (77) are the so-called group-enriched proteins (i.e., at least fourfold enriched in 2–7 brain regions) and Class 4 (441) comprises regionally enhanced proteins (at least fourfold enriched in one region over the average of all other regions) (Figs 4B and EV4C). The rational for the fourfold enrichment was derived from modeling the quantitative variance from the three replicates of PUT (the region with the highest variance in the dataset) (Fig 3C). Using this criterion, protein expression differences between brain regions can be detected with 99.998% confidence (Appendix Fig S2). By applying this extremely stringent cutoff, we aimed to focus attention on likely

Figure 3. Region-resolved proteomic map of the human brain.

A Illustration of the brain regions included in the proteomic atlas and numerical summary of the scope of the study.
B Cumulative density plot of the number of identified proteins and peptides as a function of quantitative precision (coefficient of variation (CV)) using protein LFQ intensities (2 μg HeLa peptides, 30-min LC gradients, $N = 21$).
C Same as panel (B) but for brain proteins from the different regions analyzed in this study.
D Same as panel (B) but for chromatographic retention time precision of synthetic peptides (PROCAL) run as LC quality controls throughout the project either alone (blue, $N = 10$) or spiked into fractionated brain region digests (red, $N = 1{,}968$).
E Bar graph showing the number of unique peptides and protein groups identified from each brain region (Cerebellum (CBM), white matter cerebellum (CBw), cerebral cortex (CC), white matter cerebral cortex (CCw), claustrum (CL), caudate nucleus (CN), hippocampus (HPC), nucleus accumbens (NAC), red nucleus (NR), olives (OB), putamen (PUT) substantia nigra (SN), thalamus (THL)).
F Rank plot of proteins from either fresh-frozen (gray) or formalin-preserved (orange) cortex sorted by iBAQ rank. Dark marks within the rank line indicate the abundance of proteins only found in one of the two samples. The Venn diagram shows the overlap at protein group level.
G Results of an open modification search using FragPipe comparing the most common chemical modifications attributed to formaldehyde fixation (PSM, peptide spectrum match).

© 2023 The Authors

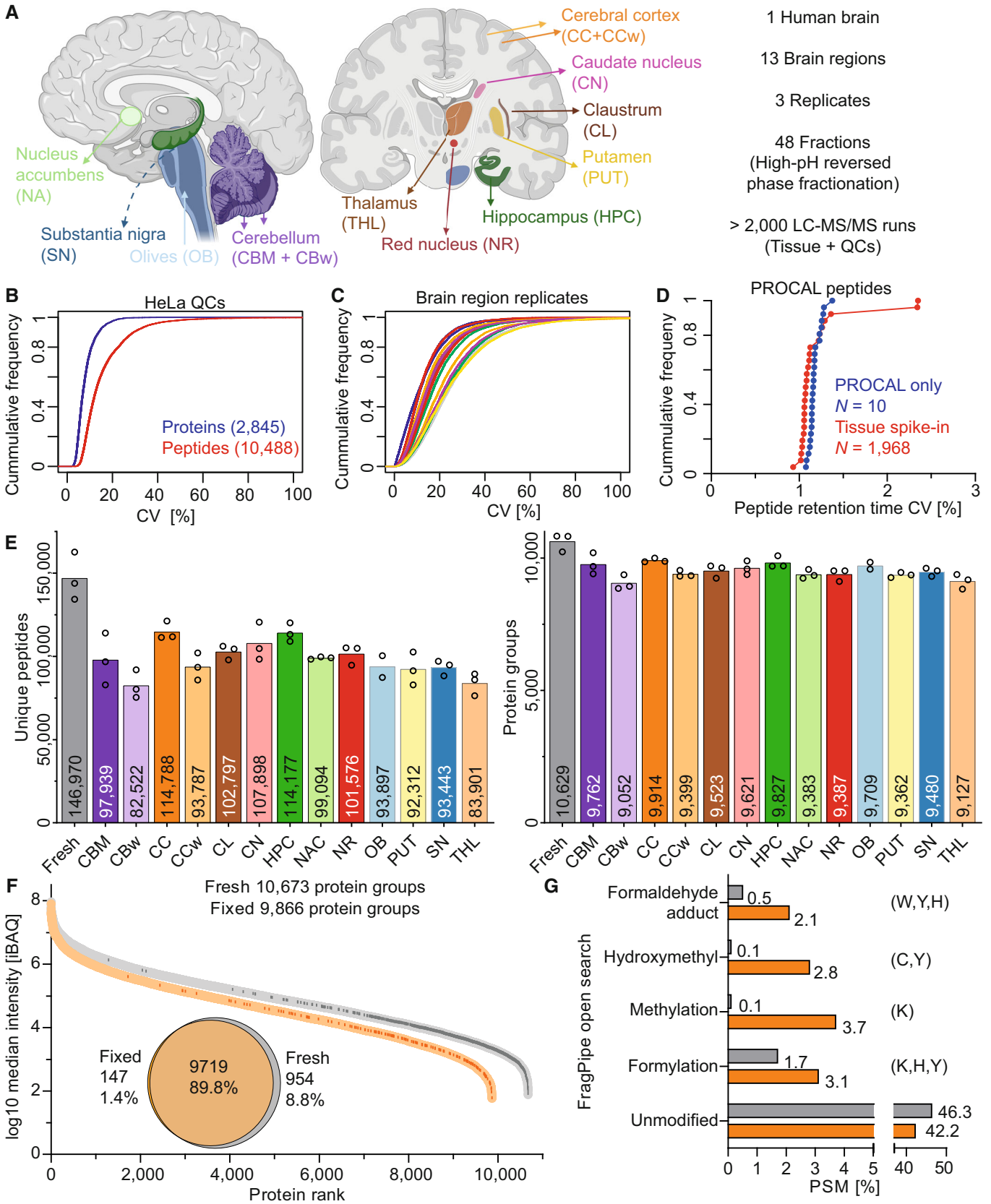

Figure 3.

© 2023 The Authors

biologically meaningful protein expression differences between regions. In total, 1,329 proteins of 9,719 (14%) fulfilled this criterion in any of the above four classes and are, from here on, referred to as brain region-enriched proteins (Dataset EV1).

A closer look at the brain region distribution within the different enrichment classes highlighted the predominance of CBM within class 1, 2 and 4. In contrast, class 3, which includes the group-enriched proteins, revealed a more even distribution between the

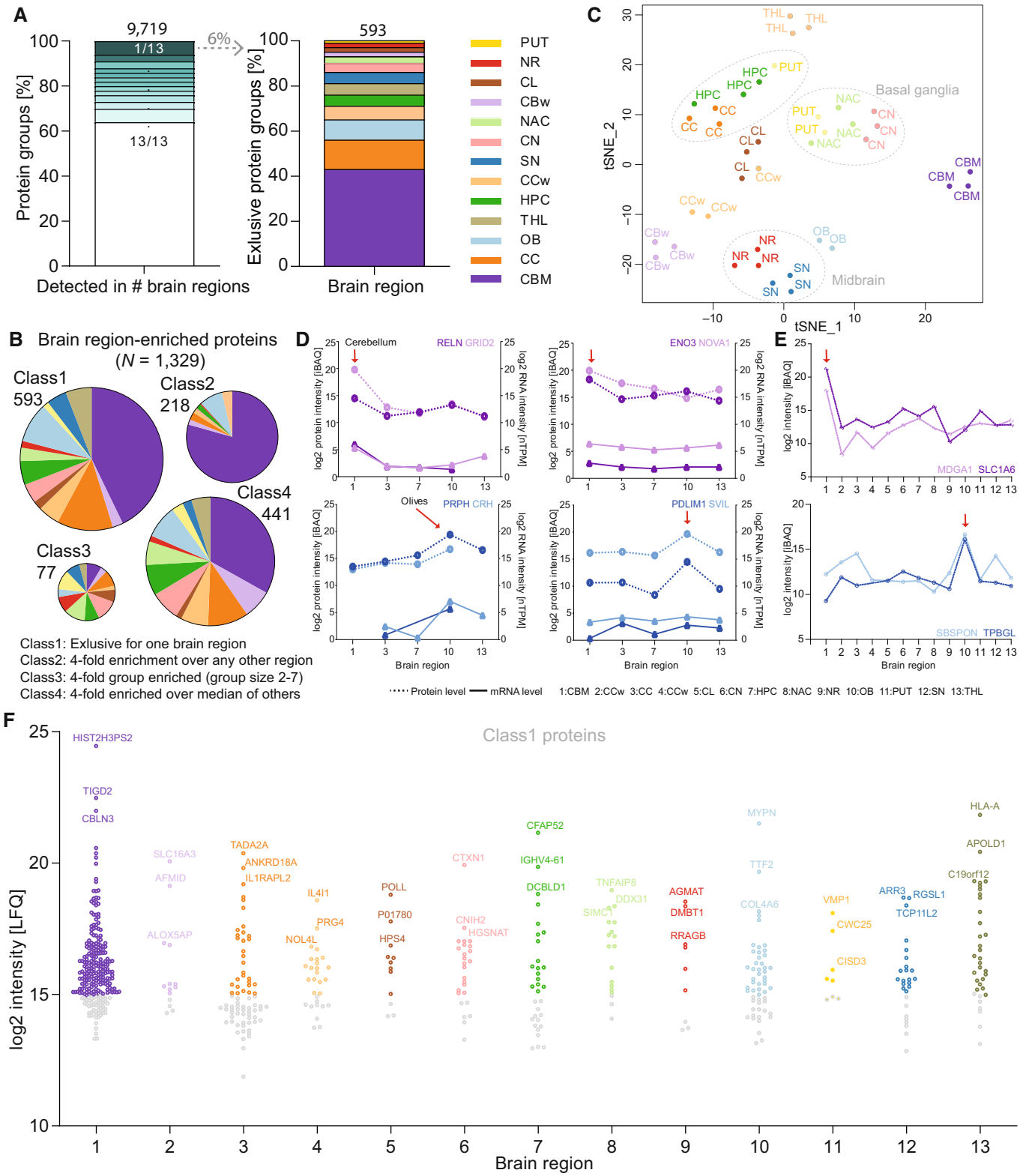

Figure 4.

Figure 4. Brain region-specific protein expression patterns.

A Percentage of proteins detected in at least two biological replicates and their distribution over 13 brain regions (left panel). Right panel: distribution of the 593 proteins exclusively identified in one brain region.
B Pie charts showing the distribution of proteins in four enrichment classes across the brain regions.
C tSNE plot of all replicate proteomes from all brain regions. Illustrating their proximity by considering all proteins. Proximity of basal ganglia (NAC, CN, PUT) and midbrain (NR, SN) samples are highlighted by dotted lines.
D Comparison of protein and mRNA levels of example proteins in five brain regions. The mRNA data was taken from the human protein atlas (HPA) (Sjöstedt *et al*, 2020). Red arrows point to brain regions in which the protein is statistically significantly higher expressed as in other brain regions.
E Protein expression profiles of four example proteins across all 13 brain regions. Red arrows as in panel (D).
F Swarm plot showing Class1 proteins (i.e., proteins exclusively identified in on region) sorted by LFQ intensity. Low abundant proteins (log2 LFQ < 15) are shown in gray.

brain regions. These include proteins such as BCL11B, CHAT, and SLC10A4, enriched in putamen (PUT), caudate nucleus (CN), and nucleus accumbens (NAC). These three regions are all part of the basal ganglia. The two brain regions of the midbrain, substantia nigra (SN), and red nucleus (RN), also share group-enriched proteins such as RTL1 and IL17RA. tSNE analysis as well as hierarchical clustering of all the data confirmed that anatomically close brain regions often share protein expression patterns (Figs 4C and EV4A, and Appendix Fig S3).

Next, we compared the proteomic and transcriptomic profiles of five brain regions for which mRNA data were available from the HPA project (Sjöstedt *et al*, 2020). As observed many times before, the overall correlation between mRNA and protein levels was low (Fig EV4B) (Carlyle *et al*, 2017; Wang *et al*, 2019). Similarly, while there are many cases where the trends in protein and mRNA levels were similar between brain regions (Fig 4D, left panels), there are also many cases for which mRNA levels were more stable than protein levels (Fig 4D, right panels). These observations yet again underscore the importance of measuring protein expression directly rather than relying on mRNA levels as a proxy, particularly when it comes to the identification of markers for certain brain regions. Class 1 proteins are the most likely source for such markers and many well-known cases were found within this class including MDGA1 and SLC1A6 in cerebellum as well as SBSPON and TPBGL in the medulla oblongata, the larger subregion of the olives (Fig 4E). Dozens of further candidates were identified for most brain regions and many of these are of high abundance ruling out the possibility that these are technical artifacts (Figs 4F and EV4B–D). Examples include the GPCR-associated signaling protein ARR3 in substantia nigra or the small (9 kDa) but poorly characterized protein CTXN1 in caudate nucleus.

A previously published transcriptome analyses of the major brain cell types (neurons, microglia, astrocytes, and oligodendrocytes) by the HPA has defined cell type-specific signatures based on mRNA expression and many of these have also been detected in this study (Figs 5A and EV5A). Of the 1,329 brain region-enriched proteins classified in the current work, 824 (62%) also fall into the cell type-specific mRNA HPA classification. These 824 regionally enriched and supposedly cell type-specific proteins split nearly evenly between neurons (46%) and glial cells (54% encompassing

Figure 5. High-level analysis of regionally enriched brain proteins.

A Bar graph showing the absolute and relative number of cell type-enhanced protein markers identified previously by the HPA project based on mRNA data (Sjöstedt *et al*, 2020) which either exhibit a regional enrichment in our proteome atlas (yellow) or not (blue). The four major brain cell types neurons (N) and the glial cell types astrocytes (A), microglia (M) and oligodendrocytes (O) are included in the analysis.
B Protein abundance profiles of cell type-enriched proteins of glial cells across the 13 brain regions.
C Same as panel (B) but for neuronal cells.
D Contribution of the different brain regions to the regionally-enriched, brain cell type-enhanced proteins (yellow proteins in panel A). The regional distribution of synaptic proteins according to UniProt is shown in addition.
E Protein abundance ratios of all synaptic proteins (according to UniProt) and oligodendrocyte enhanced proteins (according to HPA) in each specific brain region over the average of all other brain regions (dotted line) (Mean shown with SD, $N = 3$ biological).
F Mean ($N = 3$) log2 iBAQ intensities of the top 10 most abundant synaptic proteins in cerebral cortex (CC) plotted over the difference to the mean of all other brain regions.
G Absolute and relative number of proteins sorted by cellular compartment (according to UniProt) of regional enriched proteins (yellow) and non-regionally enriched proteins (blue) in the brain proteome atlas (cytoplasm (Cyto), membrane (Mem), transmembrane (TM), secreted (Sec), extracellular matrix (ECM), mitochondria (Mito), nuclear (Nuc)).
H Relative contribution of the 13 brain regions to the regionally-enriched proteins (yellow in panel G).
I Protein abundance ratios of all nuclear or secreted proteins (according to UniProt) in each specific brain region over the average of all other brain regions (dotted line) (Mean shown with SD, $N = 3$ biological).
J Examples of protein abundance profile of secreted proteins enriched in olives (OB).
K Mean ($N = 3$) log2 iBAQ intensities of the top 15 most abundant bromodomain-containing proteins in cerebellum (CBM) plotted over the difference to the mean of all other brain regions.
L Same as panel (K) but for zinc finger-containing proteins.
M Proportion of the different brain regions to the regionally-enriched proteins which are classified as drug targets of FDA-approved drugs (according to drugbank, https://go.drugbank.com/).
N Abundance of voltage-gated ion channels enriched in specific brain regions. The color code follows the brain regions and the yellow underlining indicates proteins, which are targets of FDA-approved drugs.
O Protein abundance profiles of brain region-enriched proteins that are targets of FDA-approved drugs.
P Further examples akin to panel (O).

 © 2023 The Authors

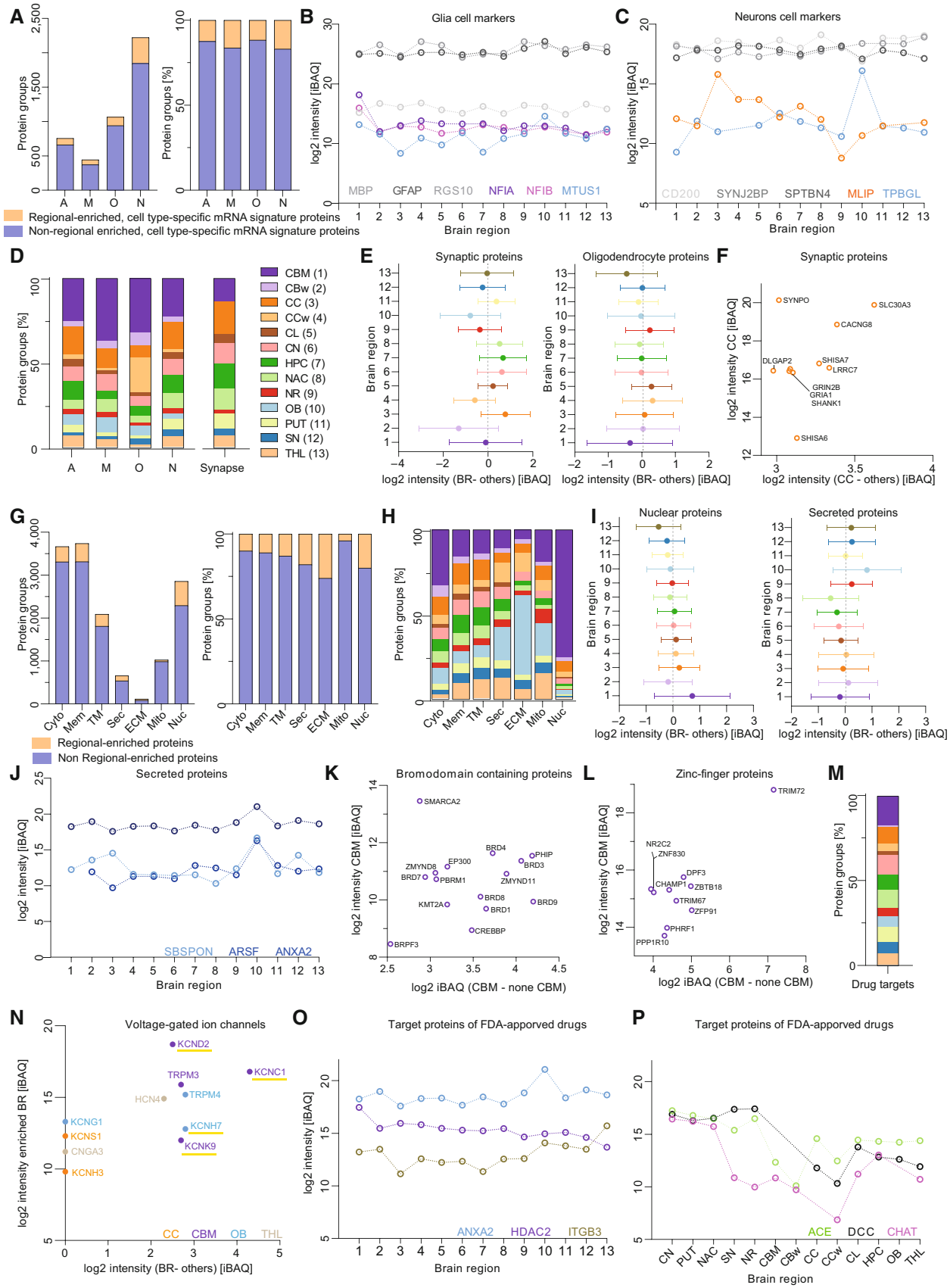
A
Protein groups
Protein groups [%]
A M O N
Regional-enriched, cell type-specific mRNA signature proteins
Non-regional enriched, cell type-specific mRNA signature proteins
B
Glia cell markers
log2 intensity [iBAQ]
MBP GFAP RGS10 NFIA NFIB MTUS1
Brain region
C
Neurons cell markers
CD200 SYNJ2BP SPTBN4 MLIP TPBGL
D
CBM (1)
CBw (2)
CC (3)
CCw (4)
CL (5)
CN (6)
HPC (7)
NAC (8)
NR (9)
OB (10)
PUT (11)
SN (12)
THL (13)
Synapse
E
Synaptic proteins
Oligodendrocyte proteins
log2 intensity (BR- others) [iBAQ]
F
Synaptic proteins
SYNPO
SLC30A3
CACNG8
DLGAP2
SHISA7
LRRC7
GRIN2B
GRIA1
SHANK1
SHISA6
log2 intensity CC [iBAQ]
log2 intensity (CC - others) [iBAQ]
G
Cyto Mem TM Sec ECM Mito Nuc
Regional-enriched proteins
Non Regional-enriched proteins
H
I
Nuclear proteins
Secreted proteins
J
Secreted proteins
SBSPON ARSF ANXA2
K
Bromodomain containing proteins
SMARCA2
BRD4
PHIP
BRD3
EP300
ZMYND8
BRD7
PBRM1
ZMYND11
BRD8
BRD1
BRD9
KMT2A
CREBBP
BRPF3
log2 intensity CBM [iBAQ]
log2 iBAQ (CBM - none CBM)
L
Zinc-finger proteins
TRIM72
NR2C2
ZNF830
DPF3
CHAMP1
ZBTB18
TRIM67
ZFP91
PHRF1
PPP1R10
M
Drug targets
N
Voltage-gated ion channels
KCND2
KCNC1
TRPM3
TRPM4
HCN4
KCNG1
KCNS1
CNGA3
KCNH3
KCNH7
KCNK9
CC CBM OB THL
log2 intensity enriched BR [iBAQ]
O
Target proteins of FDA-apporved drugs
ANXA2 HDAC2 ITGB3
P
Target proteins of FDA-apporved drugs
ACE DCC CHAT
CN PUT NAC SN NR CBM CBw CC CCw CL HPC OB THL

Figure 5.

microglia, astrocytes, and oligodendrocytes (Table 1)). Around 85% of the glial mRNA-based signature proteins were detected in all brain regions and in similar quantities, exemplified by the oligodendrocyte marker MBP, the astrocyte marker GFAP, and the microglia marker RGS10. The remaining 15% showed enrichment in at least one brain region including the oligodendrocyte protein MTUS1 in OB as well as the astrocytic proteins, NFIA and NFIB, in CBM (Fig 5B). Similarly, the majority of the neuronal mRNA-based signature proteins (83%) were not regionally enriched including the neuronal marker proteins CD200, SYNJ2BP, and SPTBN4 (Fig 5C) but around 17% were, including MLIP in CC and TPBGL in OB. A closer look at the brain region distribution of the 824 region-enriched mRNA signature proteins, showed that, for example, oligodendrocyte signature proteins are particularly prevalent in cortical white matter (CCw) (Fig 5D) with a 1.3-fold mean abundance ratio over the average of all other brain regions (Figs 5E and EV5B). This may be rationalized by the characteristic architecture of the white matter, which mainly consists of neuronal axons that are encapsulated by oligodendrocytes. In contrast, neuronal signature proteins were most prevalent in cortex (1.4-fold) (Fig EV5B) while the majority of synaptic signature proteins (as classified by UniProt) were detected in CC followed by HPC and CN (1.7, 1.6, and 1.5-fold, respectively) (Figs 5D and E, and EV5B). Among the top 10 most enriched synaptic proteins in the cerebral cortex are SHISA6 and SHISA7, which control synaptic transmission in excitatory neurons as well as SYNPO and LRRC7 involved in spine architecture (Fig 5F and Appendix Fig S4).

We next asked whether brain region-enriched proteins exhibit differences in cellular localization (as classified by UniProt) compared to all proteins detected in this study. While the majority of all detected proteins were localized to the cytoplasm or membrane-associated, most of the brain region-enriched proteins were annotated as nuclear, followed by membrane-associated, cytoplasmic, transmembrane, and secreted (Fig 5G). Interestingly, secreted and ECM-associated proteins were highly overrepresented in OB (log2 of 1.8-fold) (Fig 5H and I). These include the signaling cue proteins SBSPON, ARSF, and ANXA2 (Fig 5J). This observation is also reflected in the GO-term analysis of OB in which signal transmission and extracellular matrix are major functional annotations (Appendix Fig S5). Brain region-enriched proteins of nuclear localization were predominantly overrepresented in cerebellum (log2 of 1.8-fold) (Fig 5H and I) and included 15 members of the bromodomain and 163 members of the zinc-finger families of proteins (Top 10 enriched shown in Fig 5K and L). Again, this observation is backed by GO-term and KEGG analysis of CBM, which highlights DNA- and RNA-binding, transcription, and the spliceosome as major functional annotations (Appendix Fig S5). This relative overrepresentation of nuclear proteins within CBM is likely owing to the high density of neurons with enlarged nuclei within the gray matter of the cerebellum.

Perhaps anecdotally, we detected 12 neuropeptides of which three showed a regional enrichment. The endocrine hormones GAL and PNOC belong to class 1 enriched in OB and CC, respectively. In addition, PENK, a neuropeptide involved in pain perception (an opioid mimic) had high levels in midbrain (SN, NR) and basal ganglia (CN, PUT, NAC) (Fig EV5C). PENK was previously shown to be involved in glutamate release within the striatum (CN, PUT).

We then matched a list of proteins targeted by FDA-approved drugs (https://go.drugbank.com/) to the proteomic data and identified 470. Of these, 68 showed a brain region-specific enrichment covering all 13 brain regions (Fig 5M). Among the 68 were 11 voltage gated ion channels (Fig 5N), and ITGB3, ANXA2, and HDAC2 (Fig 5O). The latter is noteworthy because only HDAC2 showed a regional enrichment in cerebellum while all other detected HDAC family members were found with similar levels in all brain regions. Similarly, eight members of the annexin family were detected of which only ANXA2 and ANXA3 showed regional enrichment in CCw. In addition, several group enriched (Class 3) proteins classified as drug targets were overrepresented in basal ganglia (ACE, DCC, CHAT) and midbrain (ACE, DDC) (Fig 5P).

The above examples served to exemplify the types of questions one might pose to the data. In order to make the atlas easily accessible to the scientific community, we have created a web-based shiny app that enables visualizing the expression of any protein across any brain region (https://brain-region-atlas.proteomics.ls.tum.de/main_brainshinyapp/) and also offers tabular download of selected data and graphics.

Discussion

Tremendous efforts have been expended to map regional diversity in the brain mostly by using imaging and transcriptomic approaches that all attempt to help understand the highly organized substructures and diverse functions of the human brain. The current work makes a number of valuable contributions in this context. First, we developed a scalable, yet sensitive LC–MS/MS approach that paves the way for larger scale analysis, for example, by eventually analyzing all anatomically or functionally distinct regions of the brain or comparing brain structures of many individuals in terms of pathophysiology. Using this setup, the proteomic depth can be tuned seamlessly from 5,000 to 10,000 proteins by allocating either a few minutes or a few hours of time and requiring only single digit microgram quantities of total protein. Compared with the state of the art in micro-LC–MS/MS on Orbitrap instruments (Bian *et al*, 2020), the new setup achieves very similar proteome

Table 1. LC gradients run on the micro-flow LC timsTOF HT setup.

Lengths (incl loading & washing)	Sample loading B %	Gradient start B %	Gradient end B %	Column wash B %	Column equilibration B %
7.5 min	1	5	30	98	1
15 min	1	3.5	35	98	1
30 min	1	3	44	98	1
60 min	1	2	45	98	1

 © 2023 The Authors

coverage in 75% less time and using 90% less sample. An important advantage over traditional nanoLC–MS/MS is the noteworthy robustness of the setup as the entire brain region atlas project with > 2,200 consecutive samples was developed using the same LC column and ESI emitter. Second, we developed an efficient protocol for protein extraction from formalin-preserved brain material that enabled the collection of quantitative expression information for ~10,000 proteins from each brain region. This goes far beyond what was previously reported for (large-scale) FFPE material studies where proteome depth was limited to ~5,000 proteins (Coscia *et al*, 2020; Eckert *et al*, 2021; Bhatia *et al*, 2022). The fact that these data on FFPE material are nearly as deep as that for fresh-frozen holds enormous potential for the analysis of the millions of archived tissue specimen in human biobanks. Third, analysis of the 13 brain regions uncovered unique proteomic fingerprints including well-established regional marker proteins but also many new candidates. We found numerous cases where mRNA and protein profiles substantially deviated from each other underscoring the importance of measuring protein levels directly. There is also strong evidence for differences in cellular localization of proteins between brain regions as well as the expression of annotated drug targets. Forth, the web-based Shiny App along with the deposited MS raw files, protein identification and quantification data will serve the neuroscience community as a valuable data mining tool for many lines of investigations not covered in the current manuscript.

Despite the advances outlined above, many challenges remain. For instance, while we deconvoluted several substructures of, for example, the basal ganglia (caudate nucleus, putamen, and nucleus accumbens) and the midbrain (substantia nigra and red nucleus), many more exist but which were not covered here. Similarly, the atlas does not cover all brain cell types, so the contribution of rare cell types such as pericytes are likely overlooked. In addition, there is no component of spatial organization yet, for instance regarding differences in protein expression within a brain region or between the hemispheres. And, last, the atlas currently constitutes a static picture of a single human brain, so it does not shed any light on differences between individuals, dynamic changes during, for example, development or the onset and progression of disease. Still, the proteomic technology presented in this study has the potential to address most of these points because its throughput and robustness scales to the analysis of > 15,000 FFPE samples per year that may, for example, be deployed to mapping the brain or its regions at higher special resolution or focused parts of the brain across many biological or pathological conditions.

Materials and Methods

Reagents and Tools table

Reagent/Resource	Reference or Source	Identifier or Catalog Number
Chemicals, enzymes and other reagents		
Trypsin	Roche	Lab stock
SP3 beads	Hughes *et al* (2019)	Sera Mag A–B bead
2-Chloroacetamide (CCA)	Sigma-Aldrich	#C0267
PROCAL-Retention Time Standardization Kit	JPT	Zolg *et al* (2017)
Software		
MaxQuant	Cox and Mann (2008)	1.6.17.0
Spectronaut	Biognosys AG	17.1.221229.55965
FragPipe	Yu *et al* (2020a, 2020b), Chang *et al* (2020), Geiszler *et al* (2021)	FragPipe version 19.1, MSFragger version 3.7, IonQuant version 1.8.10, Philosopher version 4.8.1
GraphPad Prism	Dotmatics	9.5.1
Instrumentation		
timsTOF HT	Bruker	
VIP-HESI	Bruker	
Dionex UltiMate 3000 System	Thermo Fisher Scientific	
PepMap column	Thermo Fisher Scientific	#164711
XBridge BEH130 C18 column	Waters	#186003565
TissueLyser II	Qiagen	v2
Ultrasonication Device	Covaris	R230
Bravo Automated Liquid Handling Platform	Agilent	

© 2023 The Authors

Methods and Protocols

Brain tissue

The formalin-fixed brain of a 56-year-old Caucasian male was dissected coronally into 5-mm-thick slices. After no specific diagnostic observations were made upon macroscopic inspection and the diagnostic report was completed by a neuropathologist, 5 × 5 × 5 mm cubes of both hemispheres of the following 12 brain regions were collected: Nucleus accumbens, Substantia nigra, Thalamus, Red nucleus, Hippocampus, Putamen, Claustrum, Caudate nucleus, Cerebellum and Cerebral cortex (gray & white matter of the latter two). Additionally, two cubes of the olives were collected and stored in formalin until further processing. The investigation was approved by the local ethical committee of the Medical Faculty of the University Hospital rechts der Isar (MRI) of the Technical University Munich (project no. 176/21 S).

Sample preparation

HeLa cell lysate and plasma samples were diluted in 8 M urea buffer (80 mM Tris–HCl, pH 7.6) based on a protocol by Bian *et al* (2020)). A Bradford assay was used for protein concentration estimation. Proteins were denatured with 10 mM DTT for 60 min at 37°C while shaking at 700 rpm. Next, 55 mM 2-chloroacetamide (CCA) was added at 37°C while shaking for 60 min at 700 rpm. Five volumes of 50 mM Tris pH 8 was added and proteins were digested over night with trypsin (1:50) (Roche) at 37°C while shaking at 800 rpm. Formic acid was added to quench the digest (1% final concentration) and the peptides were at 5000 *g* for 15 min. Acidified peptides were desalted using Sep-Pak columns (HeLa) according to the user manual or using C18-based stage tips (plasma) (Rappsilber *et al*, 2007).

100 µl CSF per individual was digested using the SP3 approach (Hughes *et al*, 2019). In brief, CSF was denatured with 15 mM DTT in 50 mM ammonium bicarbonate (ABC) for 30 min at 45°C while shaking at 1,200 rpm, alkylated with 50 mM CAA in 50 mM ABC while incubating for 30 min at room temperature in the dark. Next, 15 mM DTT in 50 mM ABC was added. SP3 beads were washed with water and 5 µl beads (1:1) were used for protein binding in the presence of 70% ethanol. After washing with 80% ethanol, proteins were digested with trypsin (1:16) overnight at 37°C while shaking at 1,200 rpm. Peptides were desalted with stage tips (Rappsilber *et al*, 2007), dried in a Speed Vac and peptide concentrations were estimated using a Nanodrop.

Brain tissue cubes were incubated overnight in xylene followed by dichloromethane (DCM), washed three times with 100% ethanol, 96% ethanol, 70% ethanol and water. Next, the tissue was ground using a tissue lyser (Qiagen, 3 min at 300 s^{-1}) in 200 µl 500 mM Tris pH 9 by adding a metal ball (diameter 5 mm). Tissue lysis and decrosslinking was performed according to Eckert *et al* (2021) using 4% SDS, 10 mM DTT in 500 mM Tris pH 9. The samples were incubated for 90 min at 95°C while shaking at 1,200 rpm. In between, after 45 min of decrosslinking, samples were transferred to a 96 AFA-TUBE TPX Plate (520291) (100 µl per well) and sonicated for 5 min using the Covaris FPPE protocol using the R230 Focused Ultrasonication Device from Covaris (peak power: 350, duty factor: 25, cycles per burst 200, average power: 87.5). After an additional 45 min of crosslinking, protein concentration was measured using the PierceTM 660 nm protein assay kit according to the manufacture manual (with 0.25 g α-Cyclodextrin per 5 ml 660 nm assay solution). The pH of the lysate was adjusted to pH 7 using 8% formic acid. Next, 3 × 200 µg protein lysate (pH 7) of each sample was digested according to the SP3 protocol on a Bravo handling platform (Bian *et al*, 2020; Müller *et al*, 2020). In brief, 20 µl Sera Mag A–B bead mix (1:1) was used per 200 µg total protein lysate sample and proteins were bound at a final concentration of 70% ethanol. Beads were washed with 80% ethanol and acetonitrile (ACN). Proteins were denatured with 10 mM DTT (45 min, 37°C) followed by alkylation with 50 mM CAA in 40 mM Tris pH 7.6, CaCl2 2 mM (30 min, 25°C). Trypsin digestion was performed overnight (1:50). On the next day, peptides were collected, beads were washed with 2% formic acid, and desalted with Sep-Pak columns according to the user manual.

Off-Line High-pH Reversed-Phase Fractionation was performed using a Dionex Ultra 3000 HPLC system and a Waters XBridge BEH130 C18 column (3.5 µl 2.21 × 250 mm) based on Bekker-Jensen *et al* (2017)). 200 µg desalted peptides was reconstituted in 25 mM ABC pH 8 and separated with a linear gradient from 5% Buffer B to 40% Buffer B within 48 min at a flow rate of 200 µl/min (Buffer A: H_2O MS grade, Buffer B: ACN, Buffer C: 25 mM ABC pH 8 constant at 10%). Every 30 s, a fraction was collected, ending up in 96 fractions, which were pooled to 48 (combining fraction 1 and 49, etc.). Fractionated peptides were frozen, dried with a SpeedVac, and reconstituted with 0.1% FA plus 100 fmol PROCAL peptides. 50% of each fraction (around 2 µg) were injected for LC–MS/MS analysis.

Mass spectrometry

All proteomic data were acquired with a microflow LC coupled via a VIP HESI source to a timsTOF mass spectrometer. Liquid chromatography was performed on a Dionex UltiMate 3000 System including a WPS-3000TPL autosampler allowing direct sample pickup together with a NCS-3500RS Nano ProFlow equipped with a microflow selector within the flowmeter and standard nano pumps or a Vanquish Neo UHPLC-System in mircoflow mode (Thermo Scientific). All connections were closed via NanoViper capillaries with 50 µm ID (Thermo Scientific). A 20 µl sample loop was used. Peptides were separate on a Pepmap C18 column (1 mm ID, 15 cm lengths, 2 µm particle size) at a flow rate of 50 µl/min. Binary gradients (listed in Table 1) of buffer A and B were run (A: 0.1% FA in H_2O, B: 80% ACN 0.1%FA) including 3% DMSO, if not stated otherwise. The column oven temperature was set to 60°C. Sample loading, column washing, and equilibration were performed at maximum speed.

The source parameters were optimized in regard of signal stability as well as signal intensity and were kept constant over all measurements (Table 2). Different prototypes of emitters (IDs: 50, 80, 100 µm) and union connections between emitter and NanoViper capillary connected to the column were tested.

Table 2. VIP HESI source settings optimized for the micro-flow LC timsTOF HT setup using the 50 µm ID emitter.

End plate offset	500 V
Capillary	4,500 V
Nebulizer	2.5 bar
Dry gas	6 l/min
Dry Temp	240°C
Probe Gas Temp	300°C
Probe Gas Flow	5 l/min

 © 2023 The Authors

A timsTOF HT system was used for the majority of the measurements. Only the measurements depicted in Appendix Fig S1 were acquired on a timsTOF Pro2 system. DDA runs were acquired in PASEF mode in a mobility range of 0.85 and 1.3 with a ramp time of 100 ms with a duty cycle of 100%. The number of ramps varied according to the gradient lengths to control for sufficient data points for quantification (8 p per peak) between 4 and 10. Collision energy settings were optimized to 59 eV at 1/K0 of 1.6 Vs/cm^2 to 29 at 1/K0 of 0.6 Vs/cm^2, target intensity to 12,000 and intensity thresholds 1,600. MS2 scheduling was fastened from standard quadrupole switching time of 1.6 to 1.2 ms and MS2 acquisition time from 2.75 to 2 ms. DIA runs were acquired in PASEF mode in a mobility range of 0.64 to 1.45 with a ramp time of 100 ms with a duty cycle of 100%. Advanced collision energy settings were enabled. The DIA window scheme followed a 3 × 8 pattern of 25 m/z widths covering the whole mobility range and is shown in Table 3.

Raw data analysis

Data were search against a human, reviewed, canonical FASTA downloaded from UniProt (20,376 entries, date: 19.04.2022). PROCAL only runs as well as the tissue data of the brain resourced were additionally searched against a PROCAL entry. DDA data were processed with MaxQuant (Version 1.6.17.0). Standard settings were used with 1% false discovery rate (FDR). Additionally, trypsin was set as protease allowing up to two missed cleavages, methionine oxidation, and acetylation of N-termini were set as variable and carbamidomethylation of cysteine as fixed modification. Label-free quantification was activated with a minimal ratio count of 2. Match-between-runs was enabled for the tissue and body fluid data, but not the optimization and quality control data of the micro-flow timsTOF HT. The open search with FragPipe was performed using the "open" workflow (Version info: FragPipe version 19.1, MSFragger version 3.7, IonQuant version 1.8.10, Philosopher version 4.8.1). DIA data were analyzed with Spectronaut17 by Biognosys AG (version: 17.1.221229.55965). Standard setting for library-based searches were used without cross run normalization and matching to FASTA was enabled. Custom-made libraries based on DDA runs acquired on the optimized micro-flow LC timsTOF HT setup were used (Single-shot plasma: 12,133 precursors of 556 protein groups; Single-shot CSF: 23,098 precursors of 2,227 protein groups; Deep fractionated HeLa: 299,276 precursors of 12,356 protein groups). Further data analysis was performed with Microsoft Excel, R and Python. Data visualization was performed with R, GraphPad 9.5.1 and BioRender (Created with BioRender.com.).

Table 3. DIA window scheme used for all DIA micro-flow LC timsTOF HT runs.

#MS Type	Cycle Id	Start IM (1/K0)	End IM (1/K0)	Start Mass (m/z)	End Mass (m/z)
MS1	0	–	–	–	–
PASEF	1	1.01	1.37	800	825
PASEF	1	0.83	1.01	600	625
PASEF	1	0.64	0.83	400	425
PASEF	2	1.04	1.37	825	850
PASEF	2	0.85	1.04	625	650
PASEF	2	0.64	0.85	425	450
PASEF	3	1.06	1.37	850	875
PASEF	3	0.87	1.06	650	675
PASEF	3	0.64	0.87	450	475
PASEF	4	1.09	1.37	875	900
PASEF	4	0.9	1.09	675	700
PASEF	4	0.64	0.9	475	500
PASEF	5	1.11	1.37	900	925
PASEF	5	0.92	1.11	700	725
PASEF	5	0.64	0.92	500	525
PASEF	6	1.13	1.37	925	950
PASEF	6	0.94	1.13	725	750
PASEF	6	0.64	0.94	525	550
PASEF	7	1.16	1.37	950	975
PASEF	7	0.97	1.16	750	775
PASEF	7	0.64	0.97	550	575
PASEF	8	1.18	1.37	975	1,000
PASEF	8	0.99	1.18	775	800
PASEF	8	0.64	0.99	575	600

Data processing

For each sample, the medians were calculated from the subset of the proteins, which were expressed in at least 70% of the samples. Afterwards, the intensities were corrected by multiplying to the correction factor, which was obtained from dividing the average of medians by the median of that specific sample. To get an insight into the deferentially expressed proteins by pairwise comparisons across brain regions, the log-transformed intensities values were subjected to the imputation of missing values using down shift normal distribution (width = 0.3, downshift = 1.8). The resulting data for each protein were subjected to ANOVA followed by pairwise *t*-tests for multiple comparisons of independent groups (brain regions). The adjusted *P*-values were calculated by Benjamini/Hochberg method and the cutoff value of 0.05 was used to filter the proteins for each brain region. As reported in the Dataset EV1, a total number of 5,750 proteins showed to be expressed significantly across different brain regions.

To establish reasonable criteria for calling fold changes significant, we considered a worst case scenario where all proteins have a CV of 40% (the highest variance in the brain region replicates analyzed in this study). Note that this is a very conservative assumption, as the vast majority of the proteins have much lower CVs. For this, we simulated the intensities for two brain regions, each with three biological replicates, from two normal distributions with a shared mean mu (randomly sampled from a normal distribution with mean: 1,000.0, std: 200.0) and standard deviation mu*CV. This corresponds to the null hypothesis where there is no difference in means between the two brain regions. We repeated this for 10,000 proteins, resulting in a distribution of fold changes calculated by dividing the means of the intensities in each of the two regions (Appendix Fig S1D). This distribution of fold changes corresponds to an uncorrelated noncentral normal ratio (Hinkley, 1969), which has heavier tails than a log-normal distribution. This is a result of ratios from means drawn from opposite ends of the normal distribution. As expected, we observe that its analytical probability density

© 2023 The Authors

function closely overlays with the simulated fold changes (Appendix Fig S1E). We then applied numerical integration using the trapezoidal rule on the probability density distribution to obtain a cumulative density function (Appendix Fig S1F). From this cumulative density function, we obtain that the two-tailed 95% confidence interval corresponds to a fold change of 2.02 (log2-fold-change = 1.01). For the criterium of log2-fold-change > 2, the corresponding two-tailed *P*-value is 1.6e-3.

In order to pinpoint proteins enriched in a brain region, the classification scheme was adopted from (Sjöstedt *et al*, 2020). Class 1 encompasses proteins exclusively detected in one brain region within at least two biological replicates of that bran region and a maximum of one biological replicate in other brain regions. Class 2 includes proteins enriched in one brain region compared with all other brain regions in a pair-wise comparison by fourfold (log2 difference of 2) considering the average of each brain region. Class 3 include group enriched proteins which reveal fourfold enrichment in 2 to 7 brain regions compared the other brain regions. Class 4 encompasses brain region enhanced proteins which are fourfold enriched compared to the average of the other brain regions.

To make the analyzed data applicable to the front-end users, a shiny app was developed and dockerized in such a way that the user can explore the distribution of iBAQ or LFQ intensities of all identified proteins across different brain regions by searching for either gene name or UniProtID. In addition, an extra option was provided in the shiny app to make it possible for the selection and visualization of the annotated proteins according to their class categorizations or brain region enrichment.

To get an insight into the similarity of the biological replicates obtained from different brain regions, the log-transformed data for the most deferentially expressed proteins from the ANOVA results (as described earlier) were subjected to hierarchical cluster analysis using ward method and euclidean distance was selected as the metric before clustering. The resulting cluster-map is depicted in Appendix Fig S3. Moreover, t-SNE plots were conducted to visualize proximity between the sample using the Rtsne package considering either all proteins in the region-resolved brain resource detected in a tissue sample or the average over the biological replicates of a brain region.

GO terms related to biological processes for human were retrieved using the R package GO.db and the pathways related to KEGG pathways were downloaded from the comparative toxicogenomics database (http://ctdbase.org/). In both cases, the orthology mapping between ENTREZID and SYMBOL Ids was done using org.Hs.egGO. Afterwards, for each brain region, the proteins were primarily ranked according to their log2-fold change values and subsequently subjected to the fgsea package for fast gene set enrichment analysis. The pathways with positive normalized enrichment scores (NES) and *P*-adj < 0.05 were selected for further visualizations in Appendix Fig S5. Functional annotation clustering with the DAVID tool was applied to compare proteins over- and underrepresented in the fresh tissue compared with formalin-fixed tissue (Huang da *et al*, 2009; Sherman *et al*, 2022).

Data availability

Mass spectrometry raw data and search results performed with MaxQaunt or Spectronaut were deposited on MassIVE (MSV000091835) (Choi *et al*, 2020). Moreover, a Shinyapp is available to the public (https://brain-region-atlas.proteomics.ls.tum.de/main_brainshinyapp/) summarizing the quantitative differences on iBAQ or LFQ level of the proteins within the brain resource enabling a direct comparison of a protein of interest between the 13 brain regions.

Expanded View for this article is available online.

Acknowledgements
This work was partially supported by the Federal Ministry of Education and Research (CLINSPECT-M; FKZ161L0214A & 16LW0243K and DROP2AI; 031L0305A) and the SFB 1309 (project 401883058). We wish to thank the Bruker team involved in the project and all members of the Kuster group for their support. Study aim and brain region illustration was created with BioRender.com. Open Access funding enabled and organized by Projekt DEAL.

Author contributions
Johanna Tüshaus: Conceptualization; data curation; formal analysis; supervision; validation; investigation; visualization; methodology; writing – original draft; writing – review and editing. **Amirhossein Sakhteman:** Software; visualization. **Severin Lechner:** Data curation. **Matthew The:** Software. **Eike Mucha:** Methodology. **Christoph Krisp:** Methodology. **Jürgen Schlegel:** Resources. **Claire Delbridge:** Conceptualization; resources. **Bernhard Kuster:** Conceptualization; resources; formal analysis; supervision; funding acquisition; methodology; writing – original draft; project administration; writing – review and editing.

Disclosure and competing interests statement
BK is a founder and shareholder of MSAID and OmicScouts. He has no operational role in either company. EM and CK are employees of Bruker Daltonics. The other authors declare that they have no conflict of interest.

References

Akbarian S, Liu C, Knowles JA, Vaccarino FM, Farnham PJ, Crawford GE, Jaffe AE, Pinto D, Dracheva S, Geschwind DH *et al* (2015) The PsychENCODE project. *Nat Neurosci* 18: 1707 – 1712

Amunts K, Knoll AC, Lippert T, Pennartz CMA, Ryvlin P, Destexhe A, Jirsa VK, D'Angelo E, Bjaalie JG (2019) The Human Brain Project-Synergy between neuroscience, computing, informatics, and brain-inspired technologies. *PLoS Biol* 17: e3000344

Azevedo FA, Carvalho LR, Grinberg LT, Farfel JM, Ferretti RE, Leite RE, Jacob Filho W, Lent R, Herculano-Houzel S (2009) Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J Comp Neurol* 513: 532 – 541

Bader JM, Geyer PE, Müller JB, Strauss MT, Koch M, Leypoldt F, Koertvelyessy P, Bittner D, Schipke CG, Incesoy EI *et al* (2020) Proteome profiling in cerebrospinal fluid reveals novel biomarkers of Alzheimer's disease. *Mol Syst Biol* 16: e9356

Bekker-Jensen DB, Kelstrup CD, Batth TS, Larsen SC, Haldrup C, Bramsen JB, Sørensen KD, Høyer S, Ørntoft TF, Andersen CL *et al* (2017) An optimized shotgun strategy for the rapid generation of comprehensive human proteomes. *Cell Syst* 4: 587 – 599

Bhatia HS, Brunner AD, Öztürk F, Kapoor S, Rong Z, Mai H, Thielert M, Ali M, Al-Maskari R, Paetzold JC *et al* (2022) Spatial proteomics in three-dimensional intact specimens. *Cell* 185: 5040 – 5058

 © 2023 The Authors

Bian Y, Zheng R, Bayer FP, Wong C, Chang YC, Meng C, Zolg DP, Reinecke M, Zecha J, Wiechmann S *et al* (2020) Robust, reproducible and quantitative analysis of thousands of proteomes by micro-flow LC-MS/MS. *Nat Commun* 11: 157

Bian Y, Bayer FP, Chang YC, Meng C, Hoefer S, Deng N, Zheng R, Boychenko O, Kuster B (2021a) Robust microflow LC-MS/MS for proteome analysis: 38 000 runs and counting. *Anal Chem* 93: 3686 – 3690

Bian Y, The M, Giansanti P, Mergner J, Zheng R, Wilhelm M, Boychenko A, Kuster B (2021b) Identification of 7 000-9 000 proteins from cell lines and tissues by single-shot microflow LC-MS/MS. *Anal Chem* 93: 8687 – 8692

Biswas D, Shenoy SV, Chetanya C, Lachén-Montes M, Barpanda A, Athithyan AP, Ghosh S, Ausín K, Zelaya MV, Fernández-Irigoyen J *et al* (2021) Deciphering the interregional and interhemisphere proteome of the human brain in the context of the human proteome project. *J Proteome Res* 20: 5280 – 5293

Carlyle BC, Kitchen RR, Kanyo JE, Voss EZ, Pletikos M, Sousa AMM, Lam TT, Gerstein MB, Sestan N, Nairn AC (2017) A multiregional proteomic survey of the postnatal human brain. *Nat Neurosci* 20: 1787 – 1795

Chang HY, Kong AT, da Veiga Leprevost F, Avtonomov DM, Haynes SE, Nesvizhskii AI (2020) Crystal-C: a computational tool for refinement of open search results. *J Proteome Res* 19: 2511 – 2515

Choi M, Carver J, Chiva C, Tzouros M, Huang T, Tsai TH, Pullman B, Bernhardt OM, Hüttenhain R, Teo GC *et al* (2020) MassIVE.quant: a community resource of quantitative mass spectrometry-based proteomics datasets. *Nat Methods* 17: 981 – 984

Coscia F, Doll S, Bech JM, Schweizer L, Mund A, Lengyel E, Lindebjerg J, Madsen GI, Moreira JM, Mann M (2020) A streamlined mass spectrometry-based proteomics workflow for large-scale FFPE tissue analysis. *J Pathol* 251: 100 – 112

Covey TR, Thomson BA, Schneider BB (2009) Atmospheric pressure ion sources. *Mass Spectrom Rev* 28: 870 – 897

Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26: 1367 – 1372

Distler U, Schumann S, Kesseler HG, Pielot R, Smalla KH, Sielaff M, Schmeisser MJ, Tenzer S (2020) Proteomic analysis of brain region and sex-specific synaptic protein expression in the adult mouse brain. *Cells* 9: 313

Eckert S, Chang YC, Bayer FP, The M, Kuhn PH, Weichert W, Kuster B (2021) Evaluation of disposable trap column nanoLC-FAIMS-MS/MS for the proteomic analysis of FFPE tissue. *J Proteome Res* 20: 5402 – 5411

Geiszler DJ, Kong AT, Avtonomov DM, Yu F, Leprevost FDV, Nesvizhskii AI (2021) PTM-Shepherd: analysis and summarization of post-translational and chemical modifications from open search results. *Mol Cell Proteomics* 20: 100018

Geyer PE, Kulak NA, Pichler G, Holdt LM, Teupser D, Mann M (2016) Plasma proteome profiling to assess human health and disease. *Cell Syst* 2: 185 – 195

Green DJ, Rudd EA, Laugharn JA Jr (2014) Adaptive focused acoustics (AFA) improves the performance of microtiter plate ELISAs. *J Biomol Screen* 19: 1124 – 1130

Guo Z, Shao C, Zhang Y, Qiu W, Li W, Zhu W, Yang Q, Huang Y, Pan L, Dong Y *et al* (2022) A global multiregional proteomic map of the human cerebral cortex. *Genomics Proteomics Bioinformatics* 20: 614 – 632

Hahne H, Pachl F, Ruprecht B, Maier SK, Klaeger S, Helm D, Médard G, Wilm M, Lemeer S, Kuster B (2013) DMSO enhances electrospray response, boosting sensitivity of proteomic experiments. *Nat Methods* 10: 989 – 991

Hinkley DV (1969) On the ratio of two correlated normal random variables. *Biometrika* 56: 635 – 639

Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4: 44 – 57

Hughes CS, Moggridge S, Müller T, Sorensen PH, Morin GB, Krijgsveld J (2019) Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat Protoc* 14: 68 – 85

Karayel O, Virreira Winter S, Padmanabhan S, Kuras YI, Vu DT, Tuncali I, Merchant K, Wills AM, Scherzer CR, Mann M (2022) Proteome profiling of cerebrospinal fluid reveals biomarker candidates for Parkinson's disease. *Cell Rep Med* 3: 100661

Kong AT, Leprevost FV, Avtonomov DM, Mellacheruvu D, Nesvizhskii AI (2017) MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat Methods* 14: 513 – 520

Lechner S, Malgapo MIP, Grätz C, Steimbach RR, Baron A, Rüther P, Nadal S, Stumpf C, Loos C, Ku X *et al* (2022) Target deconvolution of HDAC pharmacopoeia reveals MBLAC2 as common off-target. *Nat Chem Biol* 18: 812 – 820

Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF, Boguski MS, Brockway KS, Byrnes EJ *et al* (2007) Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445: 168 – 176

Marchione DM, Ilieva I, Devins K, Sharpe D, Pappin DJ, Garcia BA, Wilson JP, Wojcik JB (2020) HYPERsol: high-quality data from archival FFPE tissue for clinical proteomics. *J Proteome Res* 19: 973 – 983

Meier F, Beck S, Grassl N, Lubeck M, Park MA, Raether O, Mann M (2015) Parallel accumulation-serial fragmentation (PASEF): multiplying sequencing speed and sensitivity by synchronized scans in a trapped ion mobility device. *J Proteome Res* 14: 5378 – 5387

Meier F, Brunner AD, Koch S, Koch H, Lubeck M, Krause M, Goedecke N, Decker J, Kosinski T, Park MA *et al* (2018) Online parallel accumulation-serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. *Mol Cell Proteomics* 17: 2534 – 2545

Meier F, Park MA, Mann M (2021) Trapped ion mobility spectrometry and parallel accumulation-serial fragmentation in proteomics. *Mol Cell Proteomics* 20: 100138

Melliou S, Sangster KT, Kao J, Zarrei M, Lam KHB, Howe J, Papaioannou MD, Tsang QPL, Borhani OA, Sajid RS *et al* (2022) Regionally defined proteomic profiles of human cerebral tissue and organoids reveal conserved molecular modules of neurodevelopment. *Cell Rep* 39: 110846

Messner CB, Demichev V, Bloomfield N, Yu JSL, White M, Kreidl M, Egger AS, Freiwald A, Ivosev G, Wasim F *et al* (2021) Ultra-fast proteomics with scanning SWATH. *Nat Biotechnol* 39: 846 – 854

Metz B, Kersten GF, Hoogerhout P, Brugghe HF, Timmermans HA, de Jong A, Meiring H, ten Hove J, Hennink WE, Crommelin DJ *et al* (2004) Identification of formaldehyde-induced modifications in proteins: reactions with model peptides. *J Biol Chem* 279: 6235 – 6243

Molbay M, Kolabas ZI, Todorov MI, Ohn TL, Ertürk A (2021) A guidebook for DISCO tissue clearing. *Mol Syst Biol* 17: e9807

Müller T, Kalxdorf M, Longuespée R, Kazdal DN, Stenzinger A, Krijgsveld J (2020) Automated sample preparation with SP3 for low-input clinical proteomics. *Mol Syst Biol* 16: e9111

Rappsilber J, Mann M, Ishihama Y (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* 2: 1896 – 1906

Sharma K, Schmitt S, Bergner CG, Tyanova S, Kannaiyan N, Manrique-Hoyos N, Kongi K, Cantuti L, Hanisch UK, Philips MA *et al* (2015) Cell type- and brain region-resolved mouse brain proteome. *Nat Neurosci* 18: 1819 – 1831

© 2023 The Authors

Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, Imamichi T, Chang W (2022) DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res* 50: W216 – W221

Sjöstedt E, Zhong W, Fagerberg L, Karlsson M, Mitsios N, Adori C, Oksvold P, Edfors F, Limiszewska A, Hikmet F *et al* (2020) An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science* 367: eaay5947

Wang D, Eraslan B, Wieland T, Hallström B, Hopf T, Zolg DP, Zecha J, Asplund A, Li LH, Meng C *et al* (2019) A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol Syst Biol* 15: e8503

Yu F, Haynes SE, Teo GC, Avtonomov DM, Polasky DA, Nesvizhskii AI (2020a) Fast quantitative analysis of timsTOF PASEF data with MSFragger and IonQuant. *Mol Cell Proteomics* 19: 1575 – 1585

Yu F, Teo GC, Kong AT, Haynes SE, Avtonomov DM, Geiszler DJ, Nesvizhskii AI (2020b) Identification of modified peptides using localization-aware open search. *Nat Commun* 11: 4065

Zolg DP, Wilhelm M, Yu P, Knaute T, Zerweck J, Wenschuh H, Reimer U, Schnatbaum K, Kuster B (2017) PROCAL: a set of 40 peptide standards for retention time indexing, column performance monitoring, and collision energy calibration. *Proteomics* 17: 1700263

License: This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

 © 2023 The Authors