



IN VIVO LARGE SCALE MAPPING OF PROTEIN TURNOVER IN THE HUMAN CEREBROSPINAL FLUID

Sylvain Lehmann, Christophe Hirtz, Jérôme Vialaret, Maxence Ory, Guillaume Gras Combes, Marine Le Corre, Stéphanie Badiou, Jean-Paul Cristol, Olivier Hanon, Emmanuel Cornillot, et al.

► To cite this version:

Sylvain Lehmann, Christophe Hirtz, Jérôme Vialaret, Maxence Ory, Guillaume Gras Combes, et al.. IN VIVO LARGE SCALE MAPPING OF PROTEIN TURNOVER IN THE HUMAN CEREBROSPINAL FLUID. *Analytical Chemistry*, 2019, 91 (24), pp.15500-15508. 10.1021/acs.analchem.9b03328 . hal-02368099

HAL Id: hal-02368099

<https://hal.science/hal-02368099>

Submitted on 9 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

IN VIVO LARGE SCALE MAPPING OF PROTEIN TURNOVER IN THE HUMAN CEREBROSPINAL FLUID

Sylvain Lehmann^{*,†,‡}, Christophe Hirtz^{†,‡}, Jérôme Vialaret[†], Maxence Ory[□], Guillaume Gras Combes^{‡,‡}, Marine Le Corre^{‡,‡}, Stéphanie Badiou^{‡,‡}, Jean-Paul Cristol^{‡,‡}, Olivier Hanon[°], Emmanuel Cornillot^{‡,□}, Luc Bauchet^{‡,‡}, Audrey Gabelle^{‡,‡}, Jacques Colinge^{*,‡,□,□}

^{*}CHU de Montpellier, 34295 Montpellier, France ; IRMB, INSERM, Laboratoire de Biochimie Protéomique Clinique, 34295 Montpellier, France. [†]Université de Montpellier, 34090 Montpellier, France. [□]Institut de Recherche en Cancérologie de Montpellier, INSERM, 34298 Montpellier, France. [‡]CHU de Montpellier, Hôpital Gui de Chauliac, Service de Neurochirurgie, 34295 Montpellier, France ; INSERM U1051, 34295 Montpellier, France. [§]Département de Biochimie et Hormonologie, CHU de Montpellier, 34295 Montpellier, France ; PhyMedExp, Université de Montpellier, INSERM, CNRS, 34090 Montpellier, France. [°]AP-HP, Hôpital Broca, Service de Gériatrie, 75013 Paris, France ; Université Paris Descartes, Sorbonne Paris Cité, 75006 Paris, France. ^{*}Centre Mémoire de Ressources et de Recherche Languedoc-Roussillon, 34295 Montpellier, France ; CHU de Montpellier, Hôpital Gui de Chauliac, 34295 Montpellier, France. [□]Institut régional du Cancer de Montpellier, 34298 Montpellier, France.

ABSTRACT: The extraction of accurate physiological parameters from clinical samples provides a unique perspective to understand disease etiology and evolution, including under therapy. We introduce a new methodologic framework to map patient proteome dynamics *in vivo*, either proteome-wide or in large targeted panels. We applied it to ventricular cerebrospinal fluid (CSF) and could determine the turnover parameters of almost 200 proteins, whereas a handful was known previously. We covered a large number of neuron biology- and immune system-related proteins including many biomarkers and drug targets. This first large data set unraveled a significant relationship between turnover and protein origin that relates to our ability to investigate organ physiology with protein labeling strategy specifics. Our data constitute the first draft of CSF proteome dynamics as well as a repertoire of peptides for the community to design new analyses. The disclosed methods apply to other fluids or tissues provided sequential sample collection can be performed. We show that the proposed mathematical modeling applies to other analytical methods in the field.

Clinical proteomics mostly relies on the absolute quantification of targeted proteins or global proteome quantification^{1,2}. Although highly successful, this type of analysis does not reveal the synthesis and clearance rates behind the observed abundance. Detailed and tissue-specific knowledge of individual protein turnover constitutes a complementary perspective that provides a unique insight into protein regulation. For instance, protein turnover abnormalities were related to disease pathophysiology in certain cases: amyloid- β (A β) and Tau in Alzheimer disease^{3,4} (AD), the superoxide dismutase [Cu-Zn] (SOD1) in amyotrophic lateral sclerosis⁵, or the retinol-binding protein 4 (RBP4) in type 1 diabetes⁶. Other authors employed large-scale turnover mapping to identify the regulated processes involved in zebrafish heart morphogenesis⁷ or tissue remodeling during early-stage human heart failure⁸.

Turnover data are commonly obtained by mass spectrometry (MS) and hence isotopic tracers are employed to label the newly synthesized proteins. The ratio of labeled *versus* unlabeled protein peptide abundances (Figure 1a) is called the relative isotope abundance (RIA). Usually, a time course is realized to acquire RIA curves from which synthesis and/or degradation rates can be learned through mathematical modeling. The tracer is typically introduced in animal or patient diet, or cell culture media. Commonly used tracers⁹ are ¹³C₆-labeled amino acids,

e.g. leucine (Leu) or phenylalanine, [²H₂]O, or ¹⁵N. Different tracing protocols exist and a common choice consists of delivering the tracer continuously over a rather long time (weeks). Labeling thus reaches saturation in most proteins but the longest-lived ones^{10,11}. Other protocols, which could be regarded as pulse-chase experiments, provide the tracer over a limited period and keep collecting samples afterwards¹². To reduce analytical work it is possible to multiplex samples prior to MS as was demonstrated in HeLa cells¹³. In every case, performing *in vivo* experiments in tissues causes the measured turnover to be the net result of multiple phenomena and not just a physical property of the proteins. The observed turnover results from local synthesis and degradation (molecular biology) as well as passive and active transport across tissues (physiology). Moreover, for a given protein, its turnover may vary in different tissues. Depending on the tracing protocol, different abilities to separate local *versus* remote contributions might be achieved.

Limited *in vivo* data are available in human. Initial efforts aimed at characterizing the total protein dynamics upon uptake of isotopically labeled aminoacids¹⁴. More recent work taking advantage of progress in proteomics determined protein turnover at the single protein level^{5,6,12,15}, in small- to medium-size panels^{16,17}, or proteome wide⁸ with a strong bias towards plasma. The protocol called stable isotope labeling kinetics¹²

(SILK) was introduced to follow A β in the cerebrospinal fluid (CSF). A $^{13}\text{C}_6$ -Leu tracer was intravenously injected for 9 hours while CSF was sequentially collected over 24 hours or more. It

is a thus pulse-chase protocol. SILK allowed its authors to determine the turnover of AD molecular actors such as A β

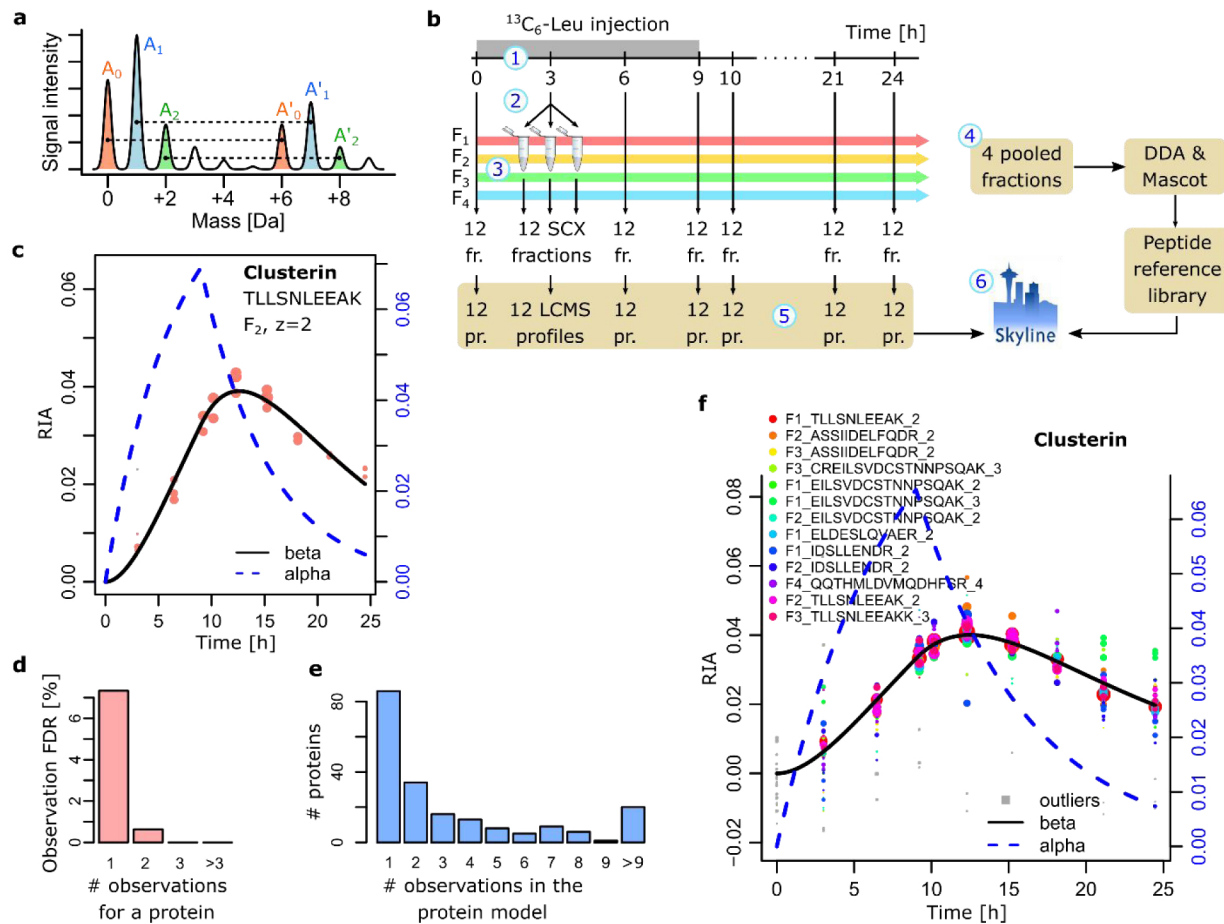


Figure 1. SILK principle. a) Principle of RIA calculation. b) Proteomics workflow: 1, tracer injection for 9 hours; 2, split of each sample in 3 aliquots to increase MS coverage; 3, peptide SCX separation; 5, LC-MS profiles of each fraction of each aliquot at every time point; 6, extraction of MS peaks according to the peptide library. c) Typical example of an RIA curve featuring gradual incorporation before clearance for a clusterin peptide. The size of the dots represents the labeled peptide MS signal intensity (arbitrary scale). d) False discovery rate estimation of the selected observations (one peptide in a specific fraction and at a specific charge state). e) Number of proteins whose mathematical model was built on 1, 2, etc. observations. f) Example of a protein mathematical model for clusterin (observations are denoted fraction_peptide_charge).

peptides¹², tau proteins⁴, and ApoE¹⁵. One major finding was AD association with a reduced CSF clearance of A β . This observation drew attention to the glymphatic system impacting CSF circulation¹⁸ and was essential to model the kinetics of amyloid accumulation over the evolution of the disease¹⁹. SILK was also used to confirm the on-target effect of an anti-AD drug²⁰.

In this work, we introduce *whole proteome* SILK (wpSILK), a novel proteomics framework to perform SILK-like experiments on a large scale. In the first instance, we eliminated any affinity purification step and integrated peptide liquid chromatography (LC) to deal with sample complexity and to perform analyses proteome-wide by high-resolution MS (HRMS) profiling. In the second instance, we followed targeted peptides by multiple reaction monitoring² (MRM). A sophisticated bioinformatics pipeline including mathematical

modeling was developed to process the complex wpSILK data. We demonstrated the potential of wpSILK on ventricular CSF, a fluid of paramount clinical relevance, obtaining the turnover parameters of ~200 proteins. Our data constitute the first draft of the CSF dynamic proteome. They provide a vast repertoire of easily MS-detectable peptides amenable to turnover experiments that cover a broad range of central nervous system (CNS) functions, known biomarkers, and drug targets. Finally, we tried to dissect the relationship between observed turnover and local *versus* remote protein origin. Our results have the potential to speed up pharmacokinetic/pharmacodynamic (PK/PD) modeling of therapeutic targets related to neurological diseases.

EXPERIMENTAL SECTION

Human samples

Samples were generated following the clinical protocol “In Vivo Alzheimer Proteomics (PROMARA)” (ClinicalTrials Identifier: NCT02263235), which was authorized by the French ethical committee CPP Sud-Méditerranée IV (#2011-003926-28) and by the ANSM agency (#121457A-11). Patients hospitalized in neurosurgery unit having in place a temporary ventricular derivation of the CSF were enrolled. To facilitate ethical approval (see above), result comparisons, and data sharing we decided to follow the original SILK $^{13}\text{C}_6$ -Leu infusion protocol strictly¹². Briefly, $^{13}\text{C}_6$ -Leu prepared per the European Pharmacopeia [19] was intravenously administered. After a 10min initial bolus at 2 mg/kg, an 8h50 infusion at 2 mg/kg/h was performed. Three to 4 mL of ventricular CSF samples were collected at times: 0, 3.02, 6.47, 9.22, 10.17, 12.3, 15.22, 18.13, 21.13, and 24.47 hours. At times 0 and 9, 3 mL of plasma EDTA were also collected. CSF and blood were aliquoted in polypropylene tubes of 1.5 mL and stored at -80°C until further analysis.

Sample analysis and data generation

Ventricular CSF and blood were collected from three post subarachnoid hemorrhage patients over a 24 to 36 hours period with intravenous injection of $^{13}\text{C}_6$ -Leu. The experiment started 8 to 19 days after initial, medical ventricular drainage and normalization of CSF clinical chemistry analysis (normal CSF protein content lies in the 0.2-0.4 g/L range²¹). Patient data are reported in Table S1 and patient 1, who was chosen for HRMS profiling, had the largest post drainage delay. The other two patient samples were used for the MRM variant of wpSILK to confirm 26 selected proteins. CSF and plasma free $^{13}\text{C}_6$ -Leu concentrations were estimated by UPLC-MS/MS at different time points (Table S2) and matched previously reported values¹².

Patient 1 CSF samples were submitted to the wpSILK HRMS workflow (Figure 1b). Briefly, each sample was split into 3 aliquots and digested (LysC and trypsin). Strong cation exchange (SCX) LC was used to obtain 6 fractions for each. Fractions 5 and 6 were regrouped and eventually discarded due to very limited peptide content (Figure S1). The other 4 fractions were submitted to LC-MS profiling by first pooling the 3 identical SCX fractions of each aliquot and injecting this pool 3 times for better proteome coverage. A few injections were repeated due to poor LC (Figure S2). In parallel, we generated a peptide reference library by pooling identical SCX fractions over all the time points (Figure 1b), which were submitted 6 times each to LC-MS-MS/MS. Mascot peptide identifications ($<1\%$ FDR) were imported in Skyline software²². In total, we found 4,558 peptides present in 1,001 proteins and 3,196 peptides from 860 proteins were Leu-containing. Skyline was applied to extract primary quantitative data from LC-MS, i.e. peptide identity, SCX fraction number, retention time, charge state, nominal and $^{13}\text{C}_6$ -Leu-shifted masses, and peak areas resulting in a large tabular export of 3,044,088 rows (Table S3). Time-dependent incorporation of $^{13}\text{C}_6$ -Leu in a peptide could be followed by computing RIAs at each time point.

Full details of the chromatography and mass spectrometry procedures are provided as Supplemental Information for the sake of concision.

Peptide selection and protein model construction

The limited duration and amounts of $^{13}\text{C}_6$ -Leu injection due to obvious patient safety considerations precluded the full

labeling of proteins. Consequently, heavy ($^{13}\text{C}_6$ -Leu labeled) peptide signals were 10 to 100 times weaker than their light (unlabeled) counterparts. Moreover, untargeted HRMS profiling and subsequent peptide identification against a reference peptide library may yield wrong assignments. Data processing hence had to ensure that true RIA curves were extracted from Skyline export. Our algorithm follows the general line of bottom-up quantitative proteomics with peptide-level filters and protein-level modeling based on unique – non-shared – peptides only. Considering occasional low heavy peptide signals and Skyline peak misassignments, we found that RIAs were more accurately computed using the most intense isotopic peak only, e.g. $A'_1/(A'_1 + A_1)$ referring to Figure 1a. Initial filters were applied to eliminate incomplete or too weak signals. RIA curves from the same peptide in distinct fractions or at different charge states were treated separately, as *independent observations*. While some peptides gave rise to convincing RIA curves, e.g. Figure 1c, many adopted dubious shapes (Figure S3). We thus reasoned that a mathematical model of $^{13}\text{C}_6$ -Leu incorporation might help to eliminate ill-shaped curves and noisy data. Moreover, such a model is necessary to extract turnover parameters.

The so-called 2-compartment model, where tracers from a free pool integrate a protein-bound pool, has been shown to fit turnover dynamics data accurately^{23,24}. Considering the specifics of SILK, i.e. tracer injection over a limited time and free tracer concentration in the range 10-15% as opposed to saturation, we generalized this model. We also wrote the model in a novel fashion, better separating the role of the parameters. Mathematical details, differences with respect to previous literature, and the overall data processing algorithm are provided in Supporting Information (SI). We only introduce the general principle here. A function $f(t)$ representing $^{13}\text{C}_6$ -Leu injection is introduced, its value is 0 (no injection) or 1 (injection), i.e. $f(t) = 0$ if $t \leq 0$ or $t > 9$, and $f(t) = 1$ otherwise. Denoting $A = A_L + A_H$ the sum of light (unlabeled) and heavy (labeled) free Leu abundance, we can define the free ratio $\alpha = \frac{A_H}{A_L + A_H}$. Similarly, the protein-bound total Leu abundance is $P = P_L + P_H$, with bound ratio $\beta = \frac{P_H}{P_L + P_H}$. Assuming a rate λ of free $^{13}\text{C}_6$ -Leu availability and protein appearance/clearance rate k_c , we obtain the dynamical system:

$$\begin{cases} \alpha' = (\lambda f(t) - \alpha)k_c \\ \beta' = (\alpha - \beta)k_c \end{cases} \quad (1)$$

The absence of $^{13}\text{C}_6$ -Leu at $t = 0$ imposes $\alpha(0) = 0 = \beta(0)$. The first equation features the entrance of new tracers in the free pool and exit into the bound pool, while the second one features the corresponding entrance in the bound pool and protein degradation. Note that according to the introduction, we talk about appearance and clearance rates instead of the sole synthesis and degradation rates. Also, the model is written assuming a steady-state, which imposes that appearance and clearance occur at the same rate k_c . Model parameters were fit by minimizing squared-differences between RIA values and $\beta(t)$.

The application of the model to Skyline output allowed us to impose obvious morphological filters, e.g. $\beta(t)$ and $\alpha(t)$ both increasing initially, maximum of $\beta(t)$ not reached before 9h, and sufficient correlation between $\beta(t)$ and experimental RIAs.

To estimate the FDR associated with peptide observations we processed peptides devoid of Leu identically. We estimated a global FDR of 7.3%, which dropped to 0.6%, respectively 0%, when 2, respectively 3 or more, observations were available for a single protein (Figure 1d). At this FDR we selected 965 observations, covering 579 distinct peptides and 235 proteins. Out of these 579 peptides, 532 were unique, i.e. not shared by multiple proteins, and served as the basis of protein turnover calculations for 196 distinct proteins. Peptides from precursors

of complement C4 isotypes A and B were found with a large number of selected observations (23), none of which were unique and neither C4A nor C4B were counted. We added C4A to our list as a representative of both isotypes, which raised the total of distinct proteins to 197. Among these 197 proteins, 86 were detected with a single observation at 7.3% FDR, an acceptable risk given the unique nature of the data. The remaining 111 proteins were associated with high-confidence turnover information (Figure 1e).

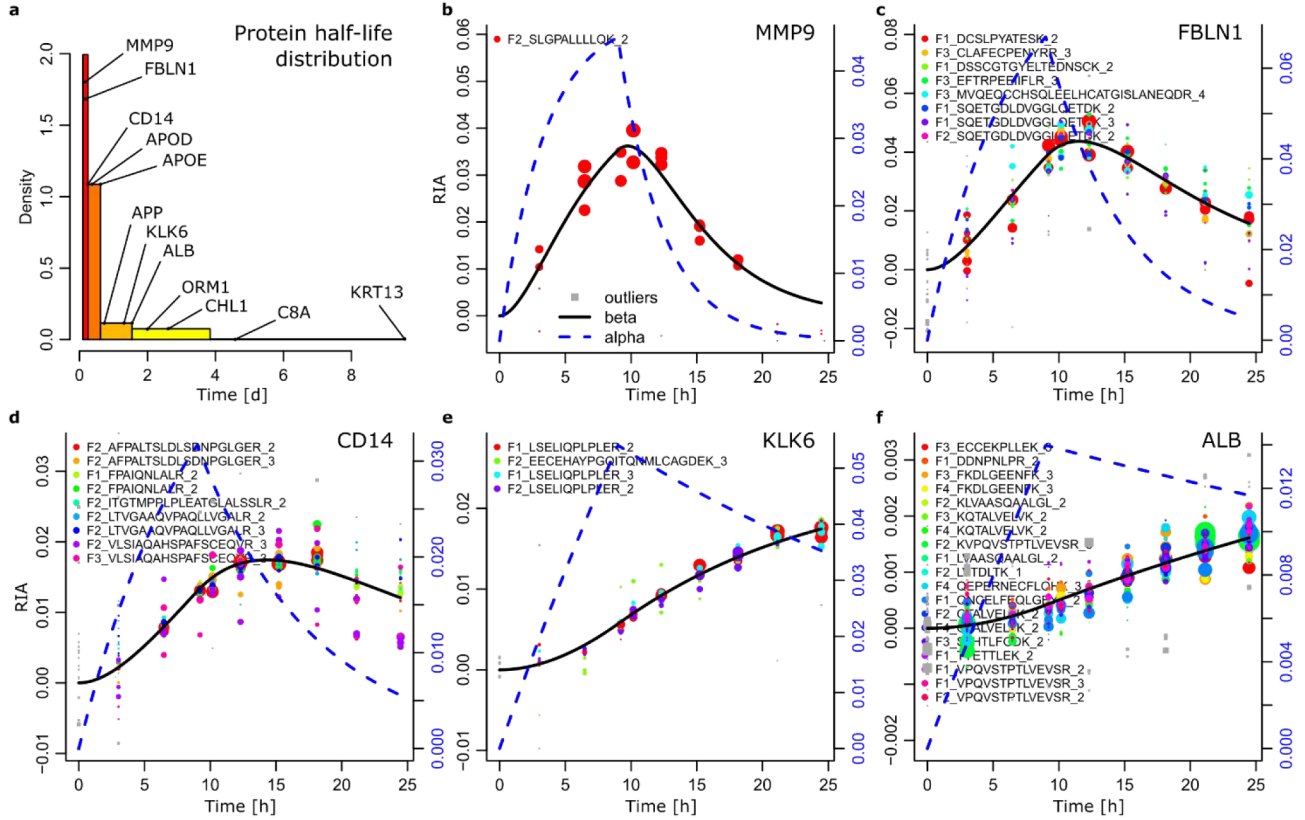


Figure 2. Overview of the observed dynamics. a) Histogram of observed half-lives with selected examples indicated. b)-f) Representative examples of distinct dynamics: Matrix metalloproteinase-9 (MMP9); Fibulin-1 (FBLN1), Monocyte differentiation antigen CD14; Kallikrein-6 (KLK6); and Serum albumin (ALB).

When more than one observation was available, we integrated them. Although isoforms or chains might display different turnovers^{15,25}, peptides that were specific to an isoform did not feature distinct dynamics except for a few cases that are discussed below. Most of the observed variability was thus expected to be experimental. We implemented a robust algorithm, where all the observations of unique peptides were first combined and fit with our model (eq 1). Then, outlier observations were eventually discarded and the integrated model recomputed. A typical example is featured in Figure 1f. Finally, to obtain estimates of model parameter standard deviations, including 1-observation proteins, we employed a bootstrap procedure²⁶. Details of the overall procedure, parameter tables as well as plots for all the selected proteins are provided in SI.

The found protein CSF dynamics were very diverse (Figure 2). The distribution of half-lives consisted of bulk of rather short-lived proteins followed by other proteins harboring slower clearance rates (Figure 2a). Many proteins displayed an

increase of their labeled proportion over the whole experiment duration, e.g. KLK6 or ALB (Figure 2e-f) despite a tracer injection stop at 9h.

Validation of selected proteins in additional patients

To confirm wpSILK HRMS results we exploited the MRM variant of our framework on 26 chosen proteins. MRM was executed on the CSF samples directly, without prior SCX peptide separation (SI) following a standard protocol²⁷ that we restricted to proteotypic peptides²⁸. The mathematical model was applied successfully to MRM data (Figure 3a-b) and the various peptide selection filters applied identically. Starting with patient 1, we found MRM RIA curves highly similar to HRMS data (Figure 3c-e) for all the proteins but two, C9 and GC (Figure 3f-g). The amplitude of the curves was not preserved, reflecting differences in the MS technology, but the shape of the curve governed by the clearance rate k_c was well-preserved. The new parametrization of the 2-compartment model we introduced in eq 1 facilitated the estimation of close k_c from similar curves but different amplitudes, yielding

reproducible estimations of the turnover parameter k_c (Figure 3h). The parameter λ , more related to the amplitude was nonetheless correlated (Figure 3i). Protein dynamics from patients 2 and 4 (Figure 3c-g), were also close and we could estimate inter-individual turnover parameter k_c variability in the range 10-20% for most proteins (Figure 3j), λ being more variable (Figure 3k) and reflecting potential variation in $^{13}\text{C}_6$ -Leu patient labeling efficiencies. The 26 protein model parameters are reported in Table S4.

RESULTS AND DISCUSSION

Dynamic proteome composition

Among the 197 proteins for which we obtained turnover parameters, 185 were known CSF proteins and 190 known

plasma protein. We used as reference plasma proteome the union of the plasma proteome database²⁹ (PPD) and proteins annotated as plasma in UniProtKB/Swissprot. Similarly, we defined a reference CSF proteome by taking the union of CSF proteins reported in two comprehensive lists^{30,31}. Potentially novel CSF proteins were (HUGO symbols for short) CEP290, FBXW10, KSR2, LOX, SH2D3A, SIK2, SPEN, TMEM212 and immunoglobulins (IGHV3-43, IGHV3-74, IGKV3D-15, IGKV4-1). FBXW10, SH2D3A, and TMEM212 were also absent from our reference plasma proteome. By compiling annotations from Reactome, KEGG, and GOBP, we classified our 197 proteins in general categories and more detailed functions (Figure 4a-b). We covered mainly neuron biology-associated proteins (56), the immune system (44), the extracellular matrix (ECM) (37),

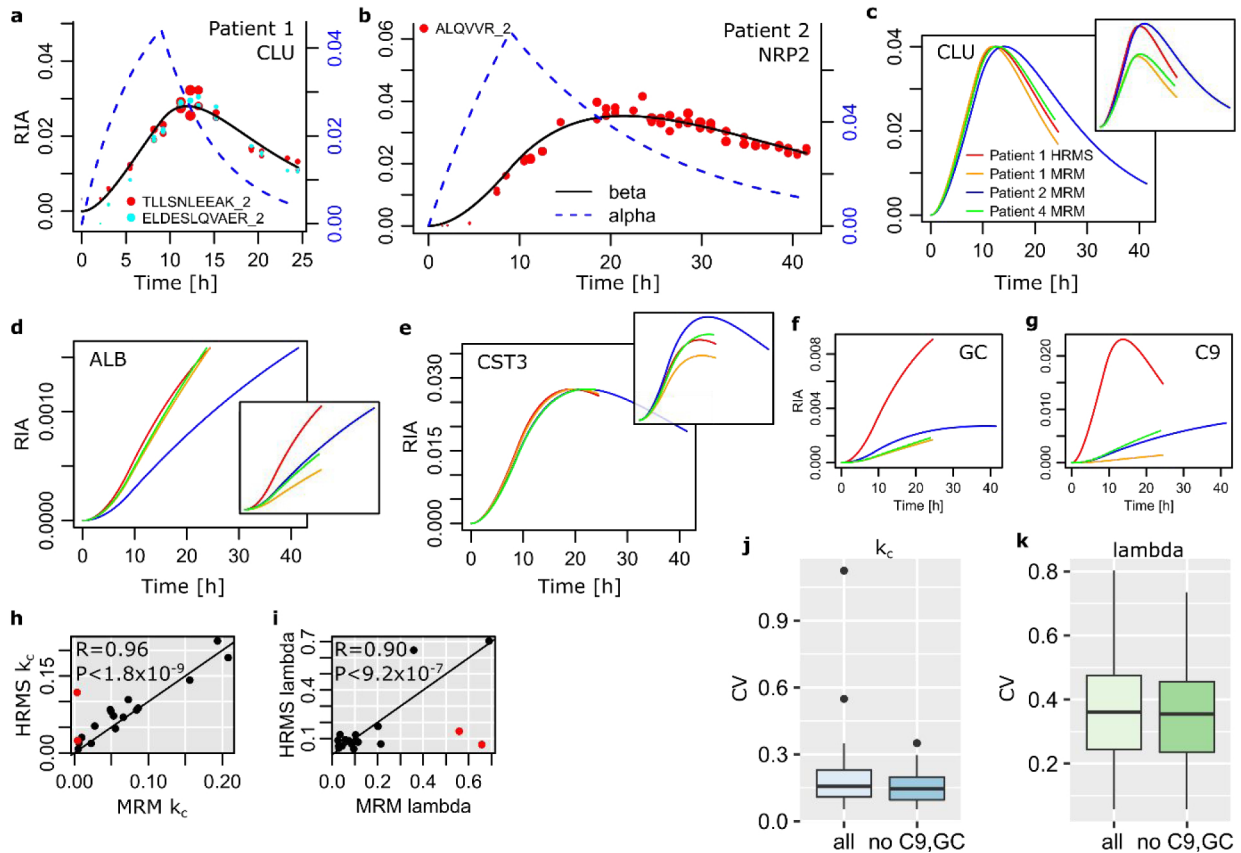


Figure 3. Validation of selected proteins by triple-quad MRM. a) Two clusterin peptides targeted in MRM. b) Neuropilin 2 MRM data in patient 2 who was followed over 41.5 hours. c) Clusterin dynamics in 3 patients, including HRMS and MRM data in patient 1. Original, non-aligned RIA reported in the insert. d) Serum albumin. e) Cystatin-C. f,g) Only 2 cases of strong disagreement between HRMS and TQ data (vitamin D-binding protein, GC, and complement component C9). h,i) Reproducibility of patient 1 turnover parameters between HRMS and MRM data. Pearson correlation, GC and C9 data excluded (red dots). Pearson's correlation coefficient and P-value by transforming to a Student's t distribution. j,k) Relative variability of turnover parameters between the 3 patient MRM data (left, all proteins), and omitting GC and C9 (right).

intracellular proteins (34) likely reflecting cell leakage, and hemostasis (26).

Neural tissue development and renewal proteins were comprised of different subcategories, e.g. apolipoproteins involved in lipid delivery, growth factors such as IGF-2 and proteins related to IGF transport and uptake, secretogranins, or proteins responsible for small molecule transport such as albumin, transferrin, transthyretin, vitamin D binding protein,

and alpha-1-acid glycoprotein 1 involved in drug uptake for instance. Multiple neuropeptides were also profiled, among which the proprotein convertase subtilisin/kexin type 1 inhibitor (PCSK1N), proenkephalin, prosaposin (PSAP), out at first homolog, neural EGFL like 2, chromogranins A and B, etc. Neuronal tissue homeostasis was represented by the amyloid- β precursor protein (APP), cystatin C, follistatin like 1, and calstn1 among others. Intracellular and extracellular

matrix proteins (ECMs, fibulins, VTN) were likely waste or debris of dead or damaged cells in majority. The dynamics of their clearance informs us of the ability of the CSF – and the CNS - to normally eliminate them, deviation from these rates might reflect pathological states.

Interestingly, we could identify distinct dynamics of the precursor and processed peptides of the myelinotrophic and neurotrophic factor PSAP involved in the lysosomal degradation of sphingolipids (Figure 4c-d) and the prohormone convertase 1 inhibitor PCSK1N (Figure S4). In some cases, peptides from the same protein display very different dynamics as exemplified by SPP1 (Figure S5). Its peptides segregate according to primary sequence positions coherently, which

might suggest the existence of undescribed post-translational processing of this protein that plays a major role in tumor progression and inflammation³². We also obtained different turnover parameters for the components of the complement system, which might reflect overall control through the shortest-lived components (C1QB, C1QC, C1R, C1S, and C9). Similarly, at the level of a protein complex, FGA has a much higher clearance rate than FGG and might control the fibrinogen assembly in a rate-limiting fashion (SI).

A repertoire of CSF proteotypic turnover biomarkers

Our data provide a large repertoire of CSF peptides that yielded high-quality MS signals compatible with dynamics pro

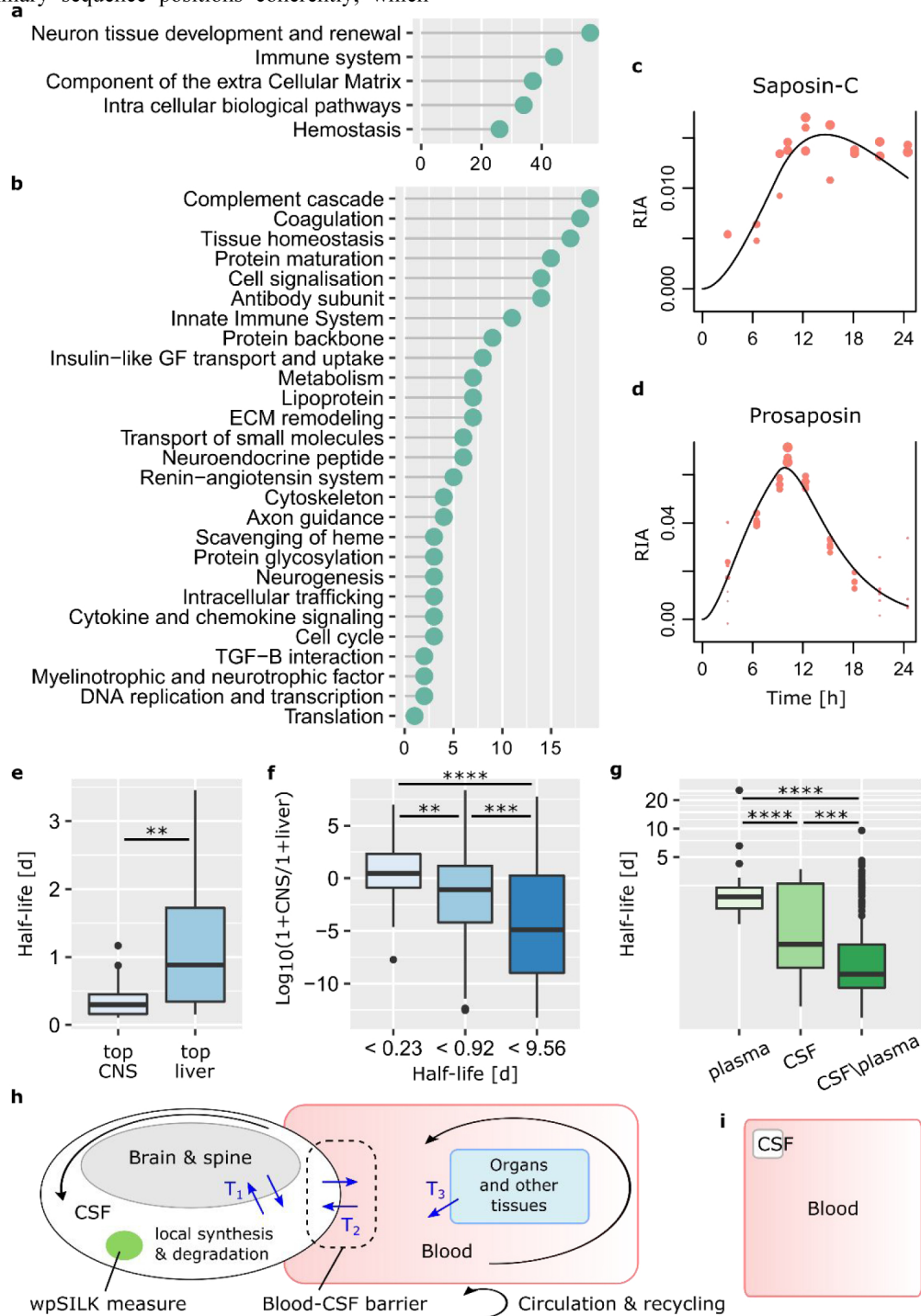


Figure 4. Dynamic proteome composition and relationship with synthesis location. a),b) General and more detailed categories of proteins for which turnover was determined. c),d) Mature chain Saposin-C versus propeptide Prosaposin dynamics. e) Top 22 proteins expressed in the CNS and the liver (GTEx database³³) display very different stability (**P<0.01, Kolmogorov-Smirnov (KS) test). f) Significant trend over the entire data set relating the CNS/liver TPM ratios³³ to the first (<0.23), second and third (<0.92), and fourth (<9.56) quartiles of half-lives. (***P<0.005, ****P<10⁻⁵, KS test). g) Price *et al.*, human plasma¹⁷ half-lives (left) *versus* the same protein in our CSF data (middle) and our CSF protein not present in the plasma data (right, KS test). h) The dynamics measured by wpSILK (green dot) are the result of transfers (T₁, T₂, T₃) occurring at different speeds, most of which transit through the blood which acts as a large reservoir compared to CSF. i) CSF (white) and blood (pink) respective volumes.

filing. More than 100 of the corresponding proteins harbor mutations or display altered expression in one or several diseases (Table S5). Among them, neurological diseases including Alzheimer (APOE, APP), spinocerebellar ataxia (FAT2), leukoencephalopathy (CSF1R, HTRA1, RNASET2, PSAP) and various disorders (AHSG, ATP6AP2, CEP152, CHI3L1, CTSD, DST, RELN, RPS23) are well represented. We also found proteins responsible for amyloidosis (FGA, GSN, LYZ, APOA2, APP, B2M, CST3), a situation where the investigation of the protein fate, i.e. the balance between production, aggregation, and clearance, is essential to understand the underneath pathophysiology¹². We noticed 16 proteins (ADAM9, CEP290, C3, C4a, C9, CFB, CFH, CFI, CST3, DCN, EFEMP1, HTRA1, IGFBP7, LTBP2, RBP4, TGFBI), which are linked to eye diseases, and in particular macular degeneration. We also surveyed the clinical trial database (clinicaltrials.gov) and found 23 proteins that are the direct targets of tested therapies (Table S6), 4 of them for diseases affecting the CNS and targets for future AD therapy (APP, CD14, MMP9, SERPING1).

Further AD proteins (52 in total) were found using the “Diseases and Functions Annotation” Ingenuity® tool. For instance, MMP-9 (Figure 2b), involved in neuronal plasticity, acts as α -secretase, therefore, reducing A β production³⁴, or ApoE, ALB (Figure 2f), and A2M that are known A β carrier. Another example is CD14 (Figure 2d), which is a receptor for TREM2, a recently identified genetic risk factor of AD and is a marker of microglial cells that mediate the phagocytosis of the amyloid component³⁵ and contribute to neuroinflammation.

A general relationship between turnover and protein origin

The CSF is relatively poor in cells and most proteins are imported from the blood or the CNS with a double origin for many of them. We continued our analysis by investigating the relationship between the tissue of origin of a protein and its turnover. Due to protein transport in the body, we reasoned that mRNA expression was a better locator of synthesis compared to tissue protein abundance. The Genotype-Tissue Expression (GTEx) project³³ served as reference. A majority of our 197 proteins were of CNS or liver origin. We grouped GTEx tissues to obtain CNS, liver, and other tissues median transcripts per million (TPM) for each gene (SI). Selecting the top 25 CNS and liver transcripts, removing common selections, left us with 22 CNS and 22 liver genes. The corresponding CNS proteins displayed significantly shorter half-lives (Figure 4e) suggesting a potential general trend. We hence grouped the whole set of 197 proteins after their half-life in 3 bins: quartiles 1, 4, and 2&3. Comparing the ratio of CNS over liver transcript abundance in these 3 groups revealed a significant association with the half-life (Figure 4f). To complete this analysis we reprocessed published plasma data¹⁷ with our mathematical model and algorithm (SI). For the 41 proteins that were shared

by our CSF data and at least 2 out of 3 patients of the plasma study, we observed shorter half-life in the CSF (Figure 4g). This significant trend was even stronger for CSF proteins absent from the plasma study, in agreement with Figure 4f.

We believe that the relationship between half-life and protein origin results from the combined effect of CSF renewal on the one hand, and plasma and other tissues acting as a reservoir on the other hand (Figure 4h). The CSF is mostly produced by the choroid plexus and reabsorbed by arachnoid granulations, its renewal occurs 3-4 times/day. Proteins entering the CSF, from the CNS or the plasma, experience this flow which adds to protein degradation. Remarkably, the median half-life of our 197 proteins is 7.4 hours, a value compatible with the CSF renewal rate. Nonetheless, a dissymmetry is created with proteins not produced in the CNS due to the aforementioned reservoir effect, which might even be amplified by frequent higher protein plasma concentrations and the larger blood volume compared to CSF (Figure 4i). This reservoir effect may thus change CNS clearance rates of proteins that are not present in the CNS predominantly or not blocked at the CSF-blood barrier, resulting in an apparent continuous supply of labeled proteins. One could hypothesize that the stability of proteins produced in the liver could be higher due to the necessity to travel a long way. This nonetheless contradicts murine data, where brain proteins were found more stable than liver proteins¹¹, an observation we could confirm reanalyzing these data and considering proteins detected in brain, blood, and liver to limit the contribution of intracellular proteins (Figure S6).

CONCLUSIONS

To be able to link molecular biology and physiology in a global picture is a very desirable goal. This is the foundation of many past and current analytical efforts and tremendous progresses have been achieved. Nevertheless, to fully embrace the complexity of the relationship between locally occurring molecular interactions and the interplay of several tissues remains an immense challenge. In this report, we contributed what we believe could be an important tool to address that task. We propose a methodological framework (wpSILK) to *in vivo* interrogate clinical samples for protein turnover in a highly multiplexed, targeted fashion employing MRM or proteome-wide through HRMS. It required the development of innovative bioinformatics and mathematical modeling, which were shown to be applicable to already published data sets obtained with different proteomics choices as well. Recent MS instrument developments such as applications ion mobility or further improvements might ultimately alleviate the need for peptide SCX chromatographic separation or allow to analyze all the time points at once exploiting isobaric tagging. In such a case, wpSILK implementation would be similar to the MRM variant we presented and mathematical modeling would remain identical.

We illustrated our methods by characterizing the human ventricular CSF *in vivo*. We could obtain CSF turnover parameters for 197 proteins, which is a significant increase over previously existing data covering a handful of proteins^{4,5,12,15} (A β , APOE, SOD1, tau) to the best of our knowledge. We further showed result reproducibility across a panel of 26 proteins in the same patient and two additional individuals with an orthogonal MS technology (MRM). This panel provided the first estimate of inter-individual turnover variability (10-20%) for late-sixty females. Among previous data, only APOE was also present in our study, we found a turnover parameter k_c in the range 3-5.5%/day for APOE, whereas previous values in younger (22-49 years old) males were in the 1.5-2%/day range. Overall, the inter-individual variability we observed in turnover is compatible with what was reported for APOE, SOD1, and tau (17-40%). Admittedly, the use of material from patients who suffered subarachnoid hemorrhage might bias our data. It should nonetheless be considered that wpSILK was performed after a rather long time following this trauma, integration of ¹³C₆-Leu only starts at this late moment, and CSF protein concentrations and cell counts returned to normal ranges (Table S1). Accordingly, turnover data as reported here constitute a first draft of the human CSF dynamics. Additional studies with healthy individuals – that are difficult to organize for obvious ethical reasons – will have to corroborate these numbers.

The opportunity to analyze the first large CSF turnover data set unraveled a significant correlation between CSF clearance rates and the origin of proteins present in this fluid, i.e. CNS *versus* peripheral. We could support this discovery by integrating published human^{23,36} and murine¹¹ data. CNS synthesized proteins tended to be cleared faster, which we believe is related to CSF renewal and peripheral tissues (blood, liver, etc.) acting as a reservoir. CSF renewal impacts the clearance of every protein present in the CSF, whereas protein entering the CSF through its barrier with blood experience a concomitant *reservoir effect*. The latter mechanism provides labeled proteins to CSF for a longer time due to peripheral tissues larger volumes, frequent higher concentrations, and delays metabolizing free ¹³C₆-Leu and exporting the labeled proteins towards the CSF. That is, the obtained turnover data accurately reflect CSF physiology but access to CNS turnover parameters might be limited for proteins that are not CNS-specific, especially if they have a longer half-life. This observation has consequences on the choice of the tracing protocol suggesting that continuous tracing protocols might tend to bias *in vivo* CSF-inferred CNS dynamics with average body dynamics. This is especially true for biomarkers with multiple indications in the CNS and at other organs, e.g. transthyretin (TTR) for AD³⁷ and malnutrition or protein metabolic impairment³⁸. Pulse-chase tracing should be preferred to more accurately approach CNS physiology.

Our complete data set constitutes a broad and diverse repertoire of proteotypic peptides amenable to turnover analyses from which the community could compose panels of MRM peptides to conduct research on CNS pathologies or biology. In some cases (PSAP and PCSK1N), we were able to distinguish between propeptides and active chain dynamics to follow enzymatic processing and maturation. This knowledge represents a veritable asset for the community, notably when the metabolism of a specific protein will be at the center of the investigation. This is the case in therapeutic research to evaluate

the on-target effect of a pharmaceutical agent as exemplified on a secretase inhibitor acting on A β production²⁰.

The wpSILK method applies to other fluids, e.g. lumbar CSF¹² or blood, or tissues provided sequential sample collection can be performed.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifiers PXD012462 (MS/MS and Mascot identifications for the reference peptide library) and PXD012502 (HRMS profiles). MRM data were submitted to PeptideAtlas with the identifier PASS01312. Software is available upon request for nonprofit organizations and academic users.

AUTHOR INFORMATION

Corresponding Author

* E-mails: s-lehmann@chu-montpellier.fr ; jacques.colinge@inserm.fr.

Author Contributions

SL, CH, JV, JC conceived the study. SL, CH, JV, EC, MO, JC analyzed data and wrote the manuscript. GGC, MLC, SB, JPC, OH, LB, AG provided patient material and clinical data. All authors have given approval to the final version of the manuscript.

ACKNOWLEDGMENT

This work was supported by the National French Alzheimer effort (“Plan Alzheimer”, ProMara PHRC 2010). JC was supported by a Fondation ARC grant PJA 20141201975 for the computer infrastructure. We thank Randall J. Bateman (St Louis, MO, USA) for his help realizing this project.

REFERENCES

- (1) Doherty, M. K.; Whitfield, P. D. Proteomics Moves from Expression to Turnover: Update and Future Perspective. *Expert Rev Proteomics* **2011**, *8* (3), 325–334. <https://doi.org/10.1586/epr.11.19>.
- (2) Meyer, J. G.; Schilling, B. Clinical Applications of Quantitative Proteomics Using Targeted and Untargeted Data-Independent Acquisition Techniques. *Expert Rev Proteomics* **2017**, *14* (5), 419–429. <https://doi.org/10.1080/14789450.2017.1322904>.
- (3) Mawuenyega, K. G.; Sigurdson, W.; Ovod, V.; Munsell, L.; Kasten, T.; Morris, J. C.; Yarasheski, K. E.; Bateman, R. J. Decreased Clearance of CNS Beta-Amyloid in Alzheimer’s Disease. *Science* **2010**, *330* (6012), 1774. <https://doi.org/10.1126/science.1197623>.
- (4) Sato, C.; Barthélemy, N. R.; Mawuenyega, K. G.; Patterson, B. W.; Gordon, B. A.; Jockel-Balsarotti, J.; Sullivan, M.; Crisp, M. J.; Kasten, T.; Kirmess, K. M.; et al. Tau Kinetics in Neurons and the Human Central Nervous System. *Neuron* **2018**, *98* (4), 861–864. <https://doi.org/10.1016/j.neuron.2018.04.035>.

- (5) Crisp, M. J.; Mawuenyega, K. G.; Patterson, B. W.; Reddy, N. C.; Chott, R.; Self, W. K.; Wehl, C. C.; Jockel-Balsarotti, J.; Varadhachary, A. S.; Buccelli, R. C.; et al. In Vivo Kinetic Approach Reveals Slow SOD1 Turnover in the CNS. *J. Clin. Invest.* **2015**, *125* (7), 2772–2780. <https://doi.org/10.1172/JCI80705>.
- (6) Jourdan, M.; Jaleel, A.; Karakelides, H.; Ford, G. C.; Kahn, B. B.; Nair, K. S. Impact of Type 1 Diabetes and Insulin Treatment on Plasma Levels and Fractional Synthesis Rate of Retinol-Binding Protein 4. *J. Clin. Endocrinol. Metab.* **2009**, *94* (12), 5125–5130. <https://doi.org/10.1210/jc.2009-1331>.
- (7) Konzer, A.; Ruhs, A.; Braun, H.; Jungblut, B.; Braun, T.; Krüger, M. Stable Isotope Labeling in Zebrafish Allows in Vivo Monitoring of Cardiac Morphogenesis. *Mol. Cell Proteomics* **2013**, *12* (6), 1502–1512. <https://doi.org/10.1074/mcp.M111.015594>.
- (8) Lam, M. P. Y.; Wang, D.; Lau, E.; Liem, D. A.; Kim, A. K.; Ng, D. C. M.; Liang, X.; Bleakley, B. J.; Liu, C.; Tabaraki, J. D.; et al. Protein Kinetic Signatures of the Remodeling Heart Following Isoproterenol Stimulation. *J. Clin. Invest.* **2014**, *124* (4), 1734–1744. <https://doi.org/10.1172/JCI73787>.
- (9) Claydon, A. J.; Beynon, R. Proteome Dynamics: Revisiting Turnover with a Global Perspective. *Mol. Cell Proteomics* **2012**, *11* (12), 1551–1565. <https://doi.org/10.1074/mcp.O112.022186>.
- (10) Hammond, D. E.; Claydon, A. J.; Simpson, D. M.; Edward, D.; Stockley, P.; Hurst, J. L.; Beynon, R. J. Proteome Dynamics: Tissue Variation in the Kinetics of Proteostasis in Intact Animals. *Mol. Cell Proteomics* **2016**, *15* (4), 1204–1219. <https://doi.org/10.1074/mcp.M115.053488>.
- (11) Price, J. C.; Guan, S.; Burlingame, A.; Prusiner, S. B.; Ghaemmaghami, S. Analysis of Proteome Dynamics in the Mouse Brain. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107* (32), 14508–14513. <https://doi.org/10.1073/pnas.1006551107>.
- (12) Bateman, R. J.; Munsell, L. Y.; Morris, J. C.; Swarm, R.; Yarasheski, K. E.; Holtzman, D. M. Human Amyloid-Beta Synthesis and Clearance Rates as Measured in Cerebrospinal Fluid in Vivo. *Nat. Med.* **2006**, *12* (7), 856–861. <https://doi.org/10.1038/nm1438>.
- (13) Wang, F.; Cheng, K.; Wei, X.; Qin, H.; Chen, R.; Liu, J.; Zou, H. A Six-Plex Proteome Quantification Strategy Reveals the Dynamics of Protein Turnover. *Sci Rep* **2013**, *3*, 1827. <https://doi.org/10.1038/srep01827>.
- (14) Waterlow, J. C. Whole-Body Protein Turnover in Humans--Past, Present, and Future. *Annu. Rev. Nutr.* **1995**, *15*, 57–92. <https://doi.org/10.1146/annurev.nu.15.070195.000421>.
- (15) Wildsmith, K. R.; Basak, J. M.; Patterson, B. W.; Pyatkovskyy, Y.; Kim, J.; Yarasheski, K. E.; Wang, J. X.; Mawuenyega, K. G.; Jiang, H.; Parsadanian, M.; et al. In Vivo Human Apolipoprotein E Isoform Fractional Turnover Rates in the CNS. *PLoS ONE* **2012**, *7* (6), e38013. <https://doi.org/10.1371/journal.pone.0038013>.
- (16) Jaleel, A.; Nehra, V.; Persson, X.-M. T.; Boirie, Y.; Bigelow, M.; Nair, K. S. In Vivo Measurement of Synthesis Rate of Multiple Plasma Proteins in Humans. *Am. J. Physiol. Endocrinol. Metab.* **2006**, *291* (1), E190–197. <https://doi.org/10.1152/ajpendo.00390.2005>.
- (17) Price, J. C.; Holmes, W. E.; Li, K. W.; Floreani, N. A.; Neese, R. A.; Turner, S. M.; Hellerstein, M. K. Measurement of Human Plasma Proteome Dynamics with (2)H(2)O and Liquid Chromatography Tandem Mass Spectrometry. *Anal. Biochem.* **2012**, *420* (1), 73–83. <https://doi.org/10.1016/j.ab.2011.09.007>.
- (18) Boespflug, E. L.; Iliff, J. J. The Emerging Relationship Between Interstitial Fluid-Cerebrospinal Fluid Exchange, Amyloid- β , and Sleep. *Biol. Psychiatry* **2018**, *83* (4), 328–336. <https://doi.org/10.1016/j.biopsych.2017.11.031>.
- (19) Jack, C. R.; Holtzman, D. M. Biomarker Modeling of Alzheimer's Disease. *Neuron* **2013**, *80* (6), 1347–1358. <https://doi.org/10.1016/j.neuron.2013.12.003>.
- (20) Bateman, R. J.; Siemers, E. R.; Mawuenyega, K. G.; Wen, G.; Browning, K. R.; Sigurdson, W. C.; Yarasheski, K. E.; Friedrich, S. W.; Demattos, R. B.; May, P. C.; et al. A Gamma-Secretase Inhibitor Decreases Amyloid-Beta Production in the Central Nervous System. *Ann. Neurol.* **2009**, *66* (1), 48–54. <https://doi.org/10.1002/ana.21623>.
- (21) Roche, S.; Gabelle, A.; Lehmann, S. Clinical Proteomics of the Cerebrospinal Fluid: Towards the Discovery of New Biomarkers. *Proteomics Clin Appl* **2008**, *2* (3), 428–436. <https://doi.org/10.1002/prca.200780040>.
- (22) Pino, L. K.; Searle, B. C.; Bollinger, J. G.; Nunn, B.; MacLean, B.; MacCoss, M. J. The Skyline Ecosystem: Informatics for Quantitative Mass Spectrometry Proteomics. *Mass Spectrom Rev* **2017**. <https://doi.org/10.1002/mas.21540>.
- (23) Guan, S.; Price, J. C.; Ghaemmaghami, S.; Prusiner, S. B.; Burlingame, A. L. Compartment Modeling for Mammalian Protein Turnover Studies by Stable Isotope Metabolic Labeling. *Anal. Chem.* **2012**, *84* (9), 4014–4021. <https://doi.org/10.1021/ac203330z>.
- (24) Rahman, M.; Previs, S. F.; Kasumov, T.; Sadygov, R. G. Gaussian Process Modeling of Protein Turnover. *J. Proteome Res.* **2016**, *15* (7), 2115–2122. <https://doi.org/10.1021/acs.jproteome.5b00990>.
- (25) Doherty, M. K.; Brownridge, P.; Owen, M. A. G.; Davies, S. J.; Young, I. S.; Whitfield, P. D. A Proteomics Strategy for Determining the Synthesis and Degradation Rates of Individual Proteins in Fish. *J. Proteomics* **2012**, *75* (14), 4471–4477. <https://doi.org/10.1016/j.jprot.2012.03.025>.
- (26) Davison, A. C.; Hinkley, D. V. *Bootstrap Methods and Their Applications*; Cambridge Series in Statistical and Probabilistic Mathematics; Cambridge University Press, 1997.
- (27) Percy, A. J.; Chambers, A. G.; Yang, J.; Jackson, A. M.; Domanski, D.; Burkhart, J.; Sickmann, A.; Borchers, C. H. Method and Platform Standardization in MRM-Based Quantitative Plasma Proteomics. *J. Proteomics* **2013**, *95*, 66–76. <https://doi.org/10.1016/j.jprot.2013.07.026>.
- (28) Hirtz, C.; Bros, P.; Brede, C.; Lescuyer, P.; Maceski, A. M.; Vialaret, J.; Delatour, V.; Lehmann, S.; IFCC Working Group on Clinical Quantitative Mass Spectrometry Proteomics (WG-cMSP). Regulatory Context and Validation of Assays for Clinical Mass Spectrometry Proteomics (CMSP) Methods. *Crit Rev Clin Lab Sci* **2018**, *55* (5), 346–358. <https://doi.org/10.1080/10408363.2018.1470159>.
- (29) Nanjappa, V.; Thomas, J. K.; Marimuthu, A.; Muthusamy, B.; Radhakrishnan, A.; Sharma, R.; Ahmad Khan, A.; Balakrishnan, L.; Sahasrabudhe, N. A.; Kumar, S.; et al. Plasma Proteome Database as a Resource for Proteomics Research: 2014 Update. *Nucleic Acids Res* **2014**, *42* (Database issue), D959–D965. <https://doi.org/10.1093/nar/gkt1251>.
- (30) Fernandez-Irigoyen, J.; Labarga, A.; Zabaleta, A.; de Morentin, X. M.; Perez-Valderrama, E.; Zelaya, M. V.; Santamaria, E. Toward Defining the Anatomic-Proteomic Puzzle of the Human Brain: An Integrative Analysis. *Proteomics Clin Appl* **2015**, *9* (9–10), 796–807. <https://doi.org/10.1002/prca.201400127>.
- (31) Barkovits, K.; Linden, A.; Galozzi, S.; Schilde, L.; Pacharra, S.; Mollenhauer, B.; Stoepel, N.; Steinbach, S.; May, C.; Uszkoreit, J.; et al. Characterization of Cerebrospinal Fluid via Data-Independent Acquisition Mass Spectrometry. *J. Proteome Res.* **2018**, *17* (10), 3418–3430. <https://doi.org/10.1021/acs.jproteome.8b00308>.
- (32) Castello, A.; Fischer, B.; Eichelbaum, K.; Horos, R.; Beckmann, B. M.; Strein, C.; Davey, N. E.; Humphreys, D. T.; Preiss, T.; Steinmetz, L. M.; et al. Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell* **2012**, *149* (6), 1393–1406. <https://doi.org/10.1016/j.cell.2012.04.031>.
- (33) GTEx Consortium. The Genotype-Tissue Expression (GTEx)

- Project. *Nat. Genet.* **2013**, 45 (6), 580–585. <https://doi.org/10.1038/ng.2653>.
- (34) Talamagas, A. A.; Efthimiopoulos, S.; Tsilibary, E. C.; Figueiredo-Pereira, M. E.; Tzinia, A. K. Abeta(1-40)-Induced Secretion of Matrix Metalloproteinase-9 Results in SAPPalpha Release by Association with Cell Surface APP. *Neurobiol. Dis.* **2007**, 28 (3), 304–315. <https://doi.org/10.1016/j.nbd.2007.07.016>.
- (35) Bate, C.; Veerhuis, R.; Eikelenboom, P.; Williams, A. Microglia Kill Amyloid-Beta1-42 Damaged Neurons by a CD14-Dependent Process. *Neuroreport* **2004**, 15 (9), 1427–1430.
- (36) Wang, D.; Liem, D. A.; Lau, E.; Ng, D. C. M.; Bleakley, B. J.; Cadeiras, M.; Deng, M. C.; Lam, M. P. Y.; Ping, P. Characterization of Human Plasma Proteome Dynamics Using Deuterium Oxide. *Proteomics Clin Appl* **2014**, 8 (7–8), 610–619. <https://doi.org/10.1002/prca.201400038>.
- (37) Velayudhan, L.; Killick, R.; Hye, A.; Kinsey, A.; Güntert, A.; Lynham, S.; Ward, M.; Leung, R.; Lourdasamy, A.; To, A. W. M.; et al. Plasma Transthyretin as a Candidate Marker for Alzheimer's Disease. *J. Alzheimers Dis.* **2012**, 28 (2), 369–375. <https://doi.org/10.3233/JAD-2011-110611>.
- (38) Ingenbleek, Y.; Bernstein, L. H. Plasma Transthyretin as a Biomarker of Lean Body Mass and Catabolic States¹². *Adv Nutr* **2015**, 6 (5), 572–580. <https://doi.org/10.3945/an.115.008508>.

Authors are required to submit a graphic entry for the Table of Contents (TOC) that, in conjunction with the manuscript title, should give the reader a representative idea of one of the following: A key structure, reaction, equation, concept, or theorem, etc., that is discussed in the manuscript. Consult the journal's Instructions for Authors for TOC graphic specifications.

