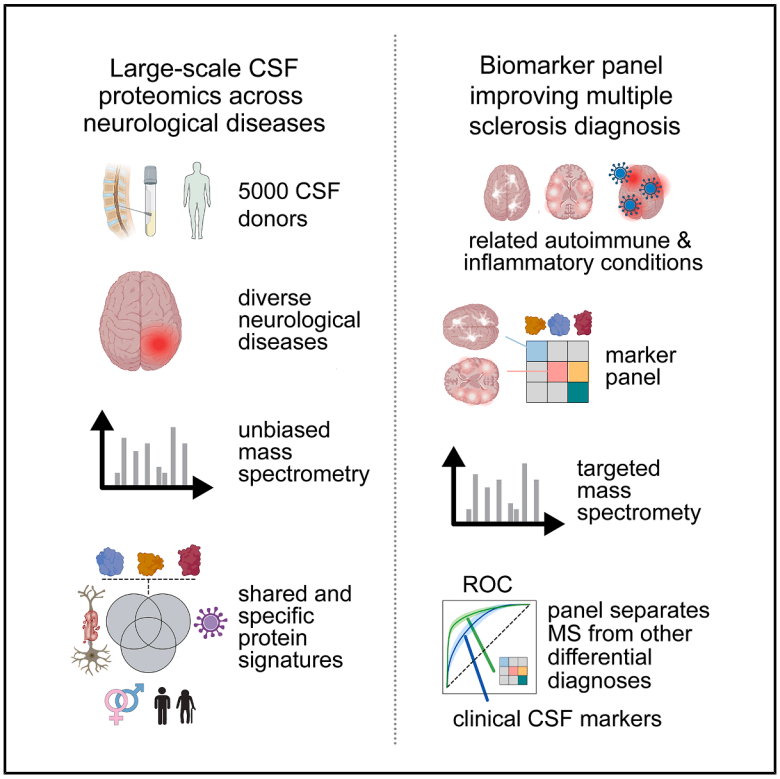
Cell

Resource

Large-scale proteomics across neurological disorders uncovers biomarker panel and targets in multiple sclerosis

Graphical abstract



Highlights

- CSF proteomics of 5,000 samples across neurological disorders quantifies 1,500 proteins per sample
- Dominant signatures for blood-CSF barrier impairment, age, and sex
- CSF proteome-based donor staging correlates with MS disability and progression
- Targeted assay for 22 CSF markers aiding differential diagnosis of OCB-negative MS

Authors

Jakob Maximilian Bader, Christine Makarov, Sabrina Richter, ..., Christiane Gasperi, Bernhard Hemmer, Matthias Mann

Correspondence

c.gasperi@tum.de (C.G.), hemmer@tum.de (B.H.), mmann@biochem.mpg.de (M.M.)

In brief

Deep proteome profiling of over 5,000 cerebrospinal fluid samples by mass spectrometry maps protein alterations across major neurological disorders, resolving key sources of variation as well as shared and disease-specific signatures. This framework yields a 22-protein assay that improves the differential diagnosis of multiple sclerosis from other inflammatory conditions, particularly in diagnostically challenging oligoclonal band-negative individuals.

Bader et al., 2026, Cell *189*, 2128–2147
April 2, 2026
https://doi.org/10.1016/j.cell.2026.01.017

CellPress

CellPress
OPEN ACCESS

Cell

Resource

Large-scale proteomics across neurological disorders uncovers biomarker panel and targets in multiple sclerosis

Jakob Maximilian Bader,[1,10] Christine Makarov,[2,10] Sabrina Richter,[3] Maximilian Thomas Strauss,[4,5] Friederike Held,[2,6] Maria Wahle,[1] Michael Baggio Lorenz,[1] Lara Pöschl,[2] Patricia Skowronek,[1] Marvin Thielert,[1] Achim Berthele,[2] Wen-Feng Zeng,[1,7] Constantin Ammar,[1] Isabell Bludau,[1,8] Benjamin Schubert,[3] Fabian J. Theis,[3] Christiane Gasperi,[2,11,*] Bernhard Hemmer,[2,9,11,*] and Matthias Mann[1,4,11,12,*]

[1]Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany
[2]Department of Neurology, School of Medicine and Health, Klinikum rechts der Isar University Hospital, Technical University of Munich, Munich, Germany
[3]Institute of Computational Biology, Helmholtz Center Munich, Munich, Germany
[4]NNF Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark
[5]OmicVision Biosciences, Nørrebro, Copenhagen, Denmark
[6]Institute for Immunity, Transplantation, and Infection, Stanford University, Stanford, CA, USA
[7]School of Engineering and School of Life Sciences, Westlake University, Hangzhou, China
[8]Institute of Neuropathology, University Hospital Heidelberg, Heidelberg, Germany
[9]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany
[10]These authors contributed equally
[11]These authors contributed equally
[12]Lead contact
*Correspondence: c.gasperi@tum.de (C.G.), hemmer@tum.de (B.H.), mmann@biochem.mpg.de (M.M.)
https://doi.org/10.1016/j.cell.2026.01.017

SUMMARY

Cerebrospinal fluid (CSF) is central to neurological diagnostics, yet biomarkers are lacking for many clinical needs. To enable its large-scale proteomic characterization, we developed a high-throughput mass spectrometry workflow quantifying approximately 1,500 proteins per CSF sample across 5,000 individuals, covering a spectrum of neurological disorders. This revealed proteomic alterations associated with blood-CSF barrier impairment, age, and sex, enabling deconvolution of shared and disease-specific signatures. We then focused on multiple sclerosis (MS), using an improved analytical technology that quantified 2,100 proteins per sample. From these data, we derived a 22-protein panel that distinguished MS from related inflammatory diseases and outperformed established markers in challenging cases. A targeted mass spectrometry assay using isotope-labeled standards validated this panel in an independent cohort, offering a clinically compatible format. Additionally, we highlight proteins of therapeutic interest and demonstrate proteome-based staging of individuals along the relapsing-progressive MS spectrum, which correlates with clinical outcomes.

INTRODUCTION

Cerebrospinal fluid (CSF) is the mainstay for molecular diagnostics in neurology, yet its diagnostic potential is far from fully realized in clinical practice. There is a pressing need for improved biomarkers to support diagnosis, prognosis, and treatment decisions. In multiple sclerosis (MS), intrathecal immunoglobulin synthesis, detected as CSF-specific oligoclonal bands (OCBs), complements radiological and clinical evidence within the McDonald diagnostic criteria.[1] Yet, despite this elaborate framework, misdiagnosis and underdiagnosis rates remain substantial, affecting roughly one in seven and one in three people with MS, respectively, with correct diagnosis often needing more than a year.[2] This delays early initiation of disease-modifying therapy (DMT), which is crucial for reducing disability progression in MS. Conversely, misdiagnosis leads to unnecessary DMT use in over half the cases, exposing individuals to potentially serious side-effects.[3,4] These problems are especially acute in OCB-negative individuals, who account for 10% of MS cases and experience double the overall misdiagnosis rate.[5]

There is a strong case for improved MS CSF biomarkers,[6,7] mirrored by the proposed update of the McDonald criteria, which includes kappa free light-chain (kFLC) index positivity as an alternative to OCB in detecting immunoglobulins.[8,9] Still, OCB

and kFLC have similar diagnostic accuracy overall,[10] and kFLC is not sensitive enough to detect MS in OCB-negative individuals at cutoffs retaining MS specificity, meaning that their use will not solve the issue.[11,12] More broadly, classical markers of inflammation or neurodegeneration like NEFL are clinically useful for disease activity management but not diagnosis.[8,13–15] Further, predicting long-term MS progression and disease course evolution to guide therapeutic decisions remains a critical unmet need.

This warrants the unbiased exploration at scale of proteins as biomarkers. Indeed, there are nearly 700 PubMed-listed publications when filtering for "MS," "biomarker," and "CSF." Biomarkers have been proposed for diverse needs, including differential diagnosis, disease activity monitoring, prognosis, and disease subtype classification. In any case, for clinical adoption, new markers must demonstrate clear benefit compared with established ones.[11] With hundreds of candidates emerging from discovery studies, it often remains unclear which should be prioritized for costly large-scale validation.

Cross-study integration to delineate final markers remains challenging due to varying study goals and limited overlap in proteins between them. Using the best technologies available at the time, studies detected only between 100 and 1,000 proteins,[16–23] with a few recent exceptions.[22,24] Furthermore, the use of healthy control groups,[16,20,25,26] or other broadly defined and often non-inflammatory neurological conditions,[13,18,22,27–31] may not align with the clinical need for differential diagnosis of a given patient. MS is most difficult to demarcate from other inflammatory diseases, especially demyelinating (DM) autoimmunity (AIM), followed by other CNS or peripheral nervous system (PNS) AIM, and infections, which were seldom included as well-defined entities.[11,17,24,32] Although these studies typically only encompass 50–200 samples,[30,33] they nevertheless highlight that novel markers should be in reach of proteomic analysis, sidestepping lengthy candidate-based approaches.

Larger and more diverse cohorts improve biomarker generalizability and enable machine-learning approaches that prioritize proteins by their predictive power instead of relying only on differential expression. Although single-analyte markers remain attractive in theory, few new candidates have translated into clinical use. Instead, discovery studies increasingly yield multimarker panels.[30,33] Generally, effectively assembling such protein panels benefits from scaling up both proteins and samples measured. Notably, a recently approved 18-protein panel outperforms NEFL in predicting MS activity, illustrating that panel-based approaches may surpass current standards.[34,35]

Using the latest advances in mass spectrometry-based proteomics, we here developed a high-throughput CSF workflow that addresses the above key constraints in sample throughput and protein coverage in this challenging matrix. Applied to over 5,000 CSF samples across a broad spectrum of neurological diseases, we identified and validated a 22-protein panel that demonstrates superior performance in OCB-negative MS compared with established clinical markers. Additionally, it performs at least equally across all OCB statuses, opening it up to all-purpose use. We convert discovery findings to a targeted assay with quantification to spike-in standards, bridging the gap between discovery and routine assays for clinical laboratories. Finally, we demonstrate how CSF proteomics can stage individuals and inform disease progression, offering a foundation for future longitudinal and interventional studies.

RESULTS

Workflow development and characterization

We studied the CSF proteome across 5,000 samples, an order of magnitude more than in typical CSF studies. This required a drastically enhanced sample throughput compared with previous studies such as our Alzheimer's disease CSF study, which measured 12 samples per day (SPD) for a total of 200 samples.[36] To retain deep and uniform proteome sampling, we integrated robotic sample preparation and a robust analytical setup, including the Evosep One for liquid chromatography and the Bruker timsTOF Pro 2 mass spectrometer for label-free data-independent acquisition (DIA). We optimized virtually every step of the workflow (Figures 1A and S1A–S1K; STAR Methods). The spectral library was based on the consensus UniProt proteome but elaborately tailored to our CSF measurement through optional depletion of abundant proteins, high-pH fractionation of tryptic peptides, as well as complementary prediction and pruning of entries observed across cohort measurements (Figures S1D–S1H; STAR Methods). Altogether, at the quintupled throughput of 60 SPD, measurement of the cohort within 3 months became feasible. Moreover, this raised the number of detected proteins from about 1,200 to 1,450 per sample, 150 of which were isoforms resolved by specific peptides (Figures 1B and S1L).

As our study progressed, the powerful Orbitrap Astral mass spectrometer was released. With this at hand, we again boosted proteome sampling to about 2,100 proteins per sample and, at the same time, increased throughput to 100 SPD (Figure S1M). Thanks to the higher dynamic range of the analyzer, samples with low numbers of proteins benefited disproportionally more, which is particularly useful as it increases the number of proteins with consistent observation across samples (Figure S1N). Further, the Astral's narrow ion windows and resulting high specificity enabled spectral analysis against a very large human proteome library predicted by AlphaPeptDeep.[37] This mitigates the need for future experimental library generation while further improving identifications (Figure S1O; STAR Methods). In a later part of this study, we used this improved setup to (re-)measure about 2,600 samples to study MS disease course types and find diagnostic biomarkers for the subgroup of individuals with MS without CSF-specific OCBs.

Study collective and proteomic dataset

The discovery cohort comprised 5,045 CSF samples, each from a different donor, in a cross-sectional design. We focused on MS and also included samples from various other neurological areas, such as AIM, neurodegeneration, infection, neoplasia, and vascular as well as other diseases, including neurological controls without evidence of structural alterations (Figure 1A; Table S1). The cohort reflected the age and sex distributions characteristic of these diseases (Figure S2A). Likewise, measures of blood-CSF barrier (BCB) impairment, e.g., albumin ratio

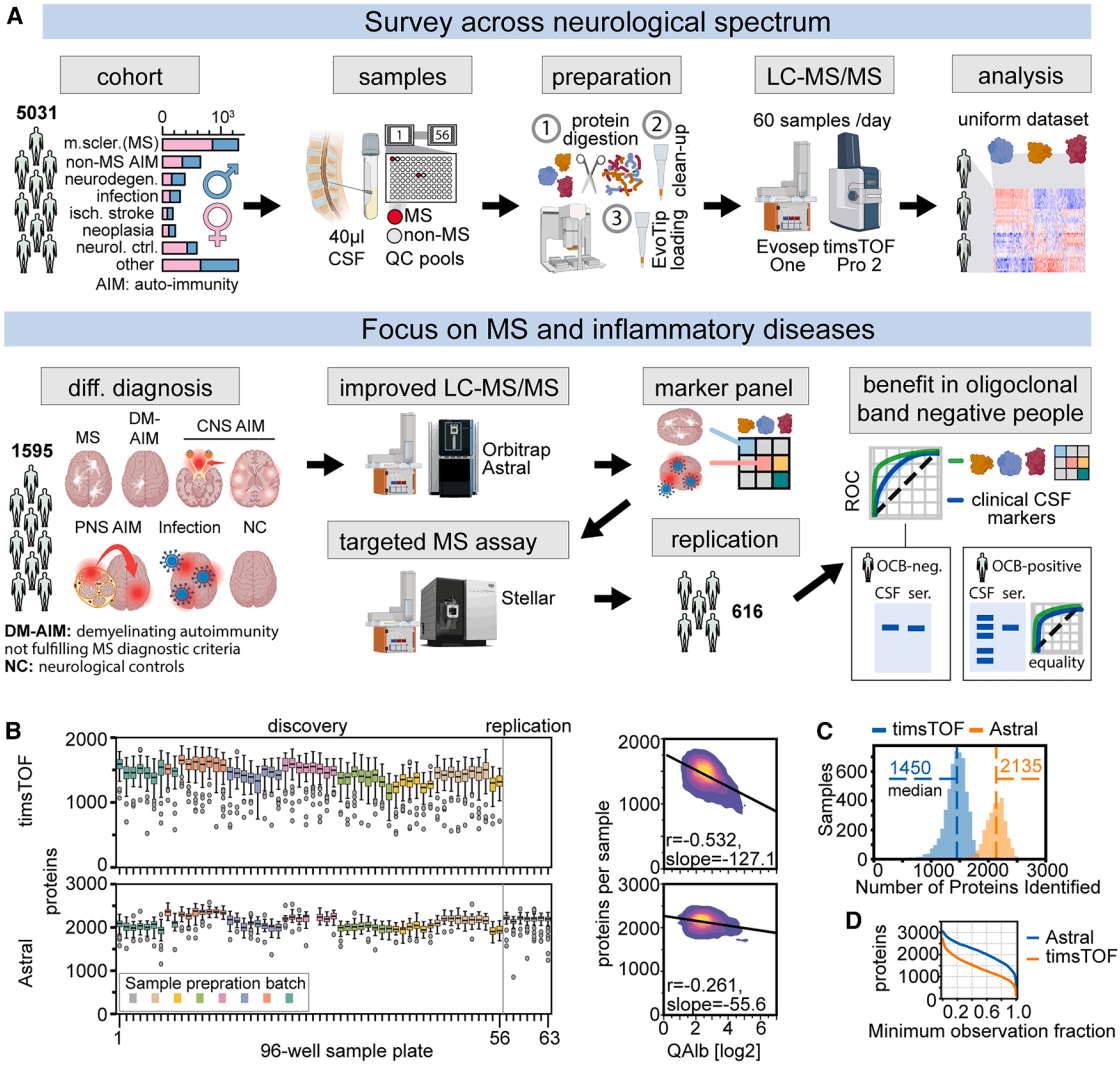


Figure 1. Study design and proteomic dataset characteristics

(A) Features of this study, related to analytical and cohort design, sample arrangement and preparation, proteomic data acquisition, and analysis.

(B) Number of proteins measured per sample along the sample plates of study (left). Dependence on the albumin ratio (right). timsTOF (top) and Astral (bottom) measurement. One Tukey boxplot shows the distribution of proteins measured across samples of a 96-well plate.

(C) Aggregated number of proteins measured per sample.

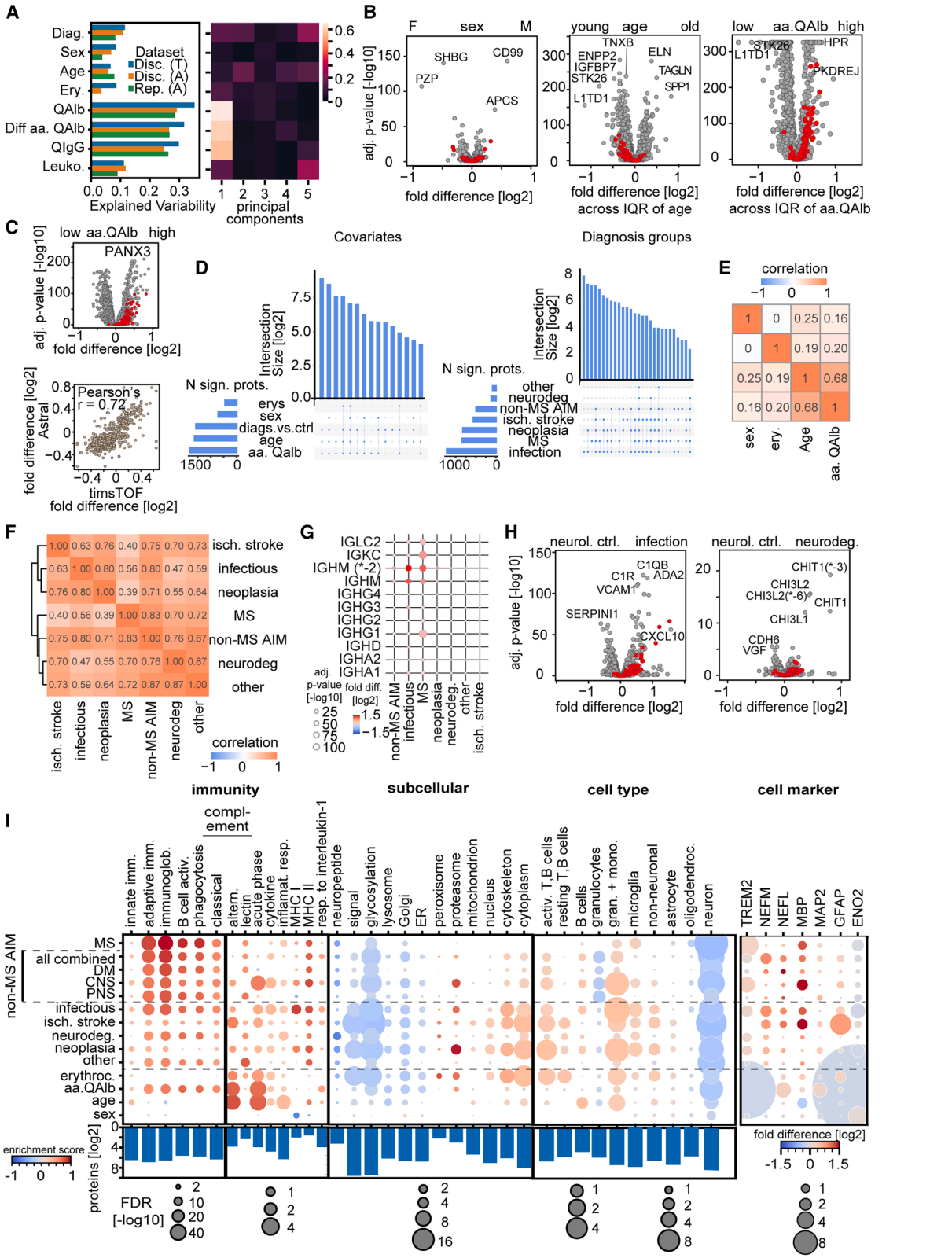
(D) Completeness of protein quantifications. Number of proteins retained (*y*) when filtering for the minimum fraction of cohort samples with protein quantification (*x*).

See also Figures S1 and S2.

(QAlb), CSF immunoglobulin concentration, number of immune cells, and presence of erythrocytes, exhibited largely overlapping but disease-inherent distributions (Figures S2B–S2E).

Large-scale cohort measurements pose challenges for maintaining uniform data quality. Our setup largely eliminated analytical decay, particularly in chromatography (Figures S1I–S1K). In the timsTOF workflow, variability in protein detection instead stemmed from sample properties, namely BCB impairment, giving CSF samples a plasma-like matrix (Figures 1B and 1C). Explaining this effect, mass spectrometry detected three to four times fewer proteins in plasma than CSF[36,38–48] (Figure S1P) due to the even stronger dominance of highly abundant proteins such as albumin (Figure S1Q). This is despite the roughly 100-fold higher total protein concentration in plasma (60–80 g/L vs. 0.1–1 g/L; Figure S1R). However, the Astral workflow maintains high protein coverage, even in samples with BCB impairment, effectively overcoming its impact on protein detection in CSF (Figure 1C).

Note that label-free mass spectrometry detects and quantifies proteins relative to the proteome rather than the sample volume. To normalize across the proteome, we use DirectLFQ, a proteomics-tailored algorithm that shifts samples to explain apparent differences between them. By design, this algorithm reports minimal regulation for as many proteins as possible, which also



(legend on next page)

makes it robust toward variable numbers of identified proteins across samples.[49] Furthermore, highly abundant proteins such as albumin do not have more leverage on normalization than lower-abundance ones.

Detection of proteins was consistent across the cohort, with more than a thousand observed in at least 80% of samples (Figure 1D). In our data, biological effects dominated the first principal component, technical batches the second, and batch correction increased biological effects (Figures S1S and S1T). About 800 proteins had a coefficient of variation (CV) of less than 20% for the entire workflow; when using the Astral mass spectrometer, this number rose to 1,000, which is about twice that observed in neat plasma (Figure S1U).[50] Variability generally increased with decreasing protein abundance but overall derived mostly from sample preparation (Figure S1U).

We conclude that this large and uniform dataset is a strong basis for the exploration of CSF proteome dynamics across a diverse set of neurological diseases.

BCB impairment dominates the CSF proteome

In our data, BCB impairment accounted for three times the variability explained by the clinical diagnosis or other covariates. This trend was consistent across our timsTOF and Astral instruments and is supported by a recent affinity proteomics study using the OLINK platform (Figures 2A and S2F).[24] From a proteomic perspective, CSF is an approximately 100-fold more dilute plasma variant, with about 70% overlapping proteins that largely retain the abundance order they have in plasma, aligning with CSF production by ultrafiltration in the choroid plexus (Figures S2G and S2H). The accumulation of plasma proteins via both physiological ultrafiltration and pathological BCB impairment is the primary determinant of total CSF protein concentration (Pearson's r = 0.95). Accordingly, differences between plasma and "BCB-intact" CSF proteomes correlate well with the QAlb effects observed across cohort samples (Figure S2I; STAR Methods).

Such QAlb effects across individuals correlated between our timsTOF and Astral measurements and further with independent OLINK data, although with a general down-shift for QAlb effects in mass spectrometric data (Figures 2B, 2C, and S2J). This is because BCB impairment increases a large portion of the proteome, but normalization is relative to the proteome in mass spectrometry. Consequently, proteins that do not originate in plasma appear reduced by higher QAlb.

In any case, BCB impairment in CSF should be accounted for regardless of the technology. Linear regression, a simple and widespread approach, readily permits this by integrating QAlb as a covariate. We note, however, that this is an approximation that captures the dominant component but may miss more subtle or complex barrier-related behavior. Nevertheless, inclusion of QAlb as a regression covariate equally corrects for true increases and apparent reductions. Accordingly, normalizing mass spectrometric data not to the entire proteome but to proteins that were found to be unaffected by QAlb in the literature[51] did not change disease-associated effects obtained through QAlb-correcting regression (Figure S2K). Generally, however, proteomics studies do not correct for QAlb, presumably because they often lack the data. We find that CSF abundance of the plasma proteins ITIH1 or C5 was highly correlated with BCB impairment. When used as proxies, they yielded highly correlated BCB-impairment and disease effects across the proteome when replacing QAlb (Figures S2L and S2M).

Effects of demographic and clinical covariates on the CSF proteome

Covariates such as age or leukocyte count in CSF explained about as much variability as disease and regulated hundreds to thousands of proteins (Figures 2A–2D and S2N). Covariate effects were predominantly orthogonal to each other. Yet, BCB impairment increases with age, as expected, yielding correlated effects[52] (Figures 2E and S2O). Across diseases, covariate effects were consistent, indicating little interaction with specific diseases (Figures S2P and S2Q). Age stands out because healthy aging diverges from disease-associated neurodegeneration. However, the fact that individuals were diagnosed directly after CSF collection limits such interactions. Indeed, we found only a few proteins with diverging age effects, including NEFL, as expected (Figure S2R). Correction strategies such as interaction terms or subtraction of healthy aging effects produced overall concordant results (Figures S3A and S3B; STAR Methods).

Figure 2. Effect of diagnoses and covariates on the CSF proteome

Differential expression analyses of protein intensities were corrected for age, sex, difference to age-adjusted albumin ratio (aa. QAlb), erythrocyte presence, and sample plate. Fold difference and significance values were obtained from the regression model (see STAR Methods).

(A) Variability of the CSF proteome explained by disease and covariates across the timsTOF (T) and Astral (A) discovery and replication cohorts (left). Contribution of these variables to the first five principal components of timsTOF data (right). Data after correction for sample plate batch effects.

(B) Protein regulation by covariates. timsTOF data. Immunoglobulin open reading frames (ORFs) are shown in red.

(C) QAlb effects in Astral data (top) and comparison with timsTOF data (bottom).

(D) Number of proteins regulated by covariates (left) in the multivariate regression model, see STAR Methods. Aggregation of proteins regulated in any disease ("diags vs. ctrl") and breakdown by disease (right).

(E) Correlation of protein regulation across clinical covariates.

(F) Correlation of protein regulation across diseases.

(G) Disease effects of immunoglobulin class-defining constant regions. Neurological controls as reference.

(H) Effects of infection or neurodegeneration. Other diseases are shown in Figures 3 and S3. Immunoglobulin ORFs are shown in red.

(I) Enrichment of annotation terms in the fold differences of individual variables in the multivariate regression. Inflammation and immunity (left), subcellular location (center left), and cell type-enriched proteomes (center right); see also Figures S3I and S3J. Regulation of individual cell-type markers (right, values from the regression model). Non-MS AIM as pooled autoimmune diseases other than MS. Enrichment score and significance from PERSEUS enrichment tool (see STAR Methods).

See also Figures S2 and S3.

However, we found that fitting and subtracting across diagnosis groups that have different age distributions is problematic because effects are not uniform across the age range (Figure S3C). Generally, covariate-regulated protein outliers included several positive controls (Figures 2B and S3D; Table S3).

Disease-associated perturbations of the CSF proteome

A major finding from our analyses is that individual neurological conditions regulated fewer proteins than confounders. The number of proteins regulated by disease status ranged from about a hundred in neurodegenerative diseases to up to 1,200 in infectious conditions (Figures 2D and S3E). Many disease effects, especially if small, only became apparent at large sample sizes (Figure S3F). Although the magnitude of changes differed across diseases, disease effects compared with neurological controls were typically well correlated, in particular for related diseases such as MS and other AIM, pointing to strong shared alterations ($r = 0.83$, Figures 2F, S3G, and S3H). Outliers associated with individual diseases contained several positive controls (Table S4; Figures 2G, 2H, and S3E).

Inflammation and immunity shaped many disease signatures. MS and other autoimmune diseases showed the strongest enrichment of immunoglobulins, whereas innate immune pathways were more prominent in infection and neoplasia (Figure 2I). Major histocompatibility complex class I (MHC class I) proteins were selectively increased in these two groups, aligning with upregulation in tissue cells induced by pathogens and tumor inflammation. Complement activation also differed, with the classical pathway dominant in MS and infection and the alternative pathway in ischemic stroke.

The release of cellular proteins emerged as a shared damage signature that characterized stroke, neoplasia, and infection, partially explaining their surprisingly correlated alterations (Figures 2F and 2I). Cell-type mapping using literature proteomes indicated immune cells as dominant sources of cellular material, likely reflecting their ability to enter CSF[53,54] (Figures S3I and S3J). Concordantly, microglial TREM2 was increased across diseases. Neurofilament proteins NEFM and NEFL were also elevated, indicating axonal injury.

MS-regulated proteins

MS is characterized by intrathecal immunoglobulin synthesis. In our data, the class-defining heavy chains were most strongly associated with MS (Figure 2G). Additionally, about 140 open reading frames present in every human encode the variable regions of immunoglobulin heavy and light chains, including kappa light chains, whose free forms have emerged as MS biomarkers. Using peptides from non-variable sequence portions, we quantified the products of these genes in hundreds to thousands of individuals (Figures S3K and S3L). These immunoglobulins accounted for many of the most strongly MS-associated proteins, including compared with other forms of AIM, which we further subdivided into DM AIM not fulfilling MS criteria and other CNS and PNS AIM (Figures 3A and 3B).

To put our findings in context with other proteomic technologies, we compared our results with a recent study that compared about 140 individuals with MS with 40 healthy controls using OLINK's proteomic technology.[16] About 450 proteins were measured by both technologies, about 500 only by OLINK's platform, and about 1,300 only by our timsTOF workflow, highlighting that mass spectrometry is competitive for CSF protein detection (Figure S3M). The fold changes of overlapping and significantly regulated proteins between MS and controls across the two study cohorts correlated well (Pearson's $r = 0.73$, Figure S3N).

A strength of our dataset is that it enables comparison of MS to a wide range of other neurological conditions. Across all these comparisons, 77 proteins were significant, and nearly all showed their highest abundance in MS CSF (Figure 3C). About three-quarters were immunoglobulins, typically of the G isotype. Notably, immunoglobulin M (IgM) proteins were even more abundant in infectious diseases, consistent with the lower rate of intrathecal IgM synthesis in MS (Figure 3D).[55] The approximately 20 non-immunoglobulin proteins included several with reported MS associations, although prior evidence often came from studies that used non-inflammatory CSF controls or studied MS associations outside CSF (Figure 3E; Table S4), making retention compared with diverse controls noteworthy. We also found additional associations of mechanistic or therapeutic interest, such as RYR2, CSF1R, and CTSW, which we put into perspective in the discussion section. Further proteins were ARSA and RECK, which are involved in myelin lipid metabolism and extracellular matrix remodeling via protease inhibition, relevant to demyelination and BCB impairment, respectively.[56,57]

Deconvolution of specific disease signatures

To explore proteomic features relevant to disease, we applied multi-class classification using XGBoost (STAR Methods). Identifying individual diseases from the combined rest of our cohort showed generally high performance for diseases other than AIM, with averaged area under the receiver operator curve (ROC AUC) values ranging from 0.80 to 0.91. Pairwise classification performance separating two given diseases was likewise high but dropped when AIM was involved: for instance, to 0.76 for separating MS from DM AIM (Figures 4A, 4B, and S4A).

To explain the model, we extracted SHAP values, which quantify the contribution of individual protein features to overall prediction of disease status (STAR Methods). To additionally study relationships of disease and covariates, we further analyzed principal components of the SHAP matrix. Separation of diseases was performed by the first three components, and, for MS predominantly, by components one and three (Figures 4C and S4B). Notably, BCB impairment influenced the second component, underlining its limited disease specificity, which, however, separated some CNS conditions, such as infections from controls. Age and leukocyte count were more discriminatory, including as features themselves, which aligns with differences inherent to the diverse diseases we covered (Figure 4D). Note that the inclusion of covariates means that remaining contributions of proteins are independent of or additional to these covariates.

Next, we investigated the decision-driving features in greater detail. Out of the entire proteome, the top 250 most informative proteins in each disease accounted for about 70% of the model's information (Figures S4C and S4D). At that cutoff,

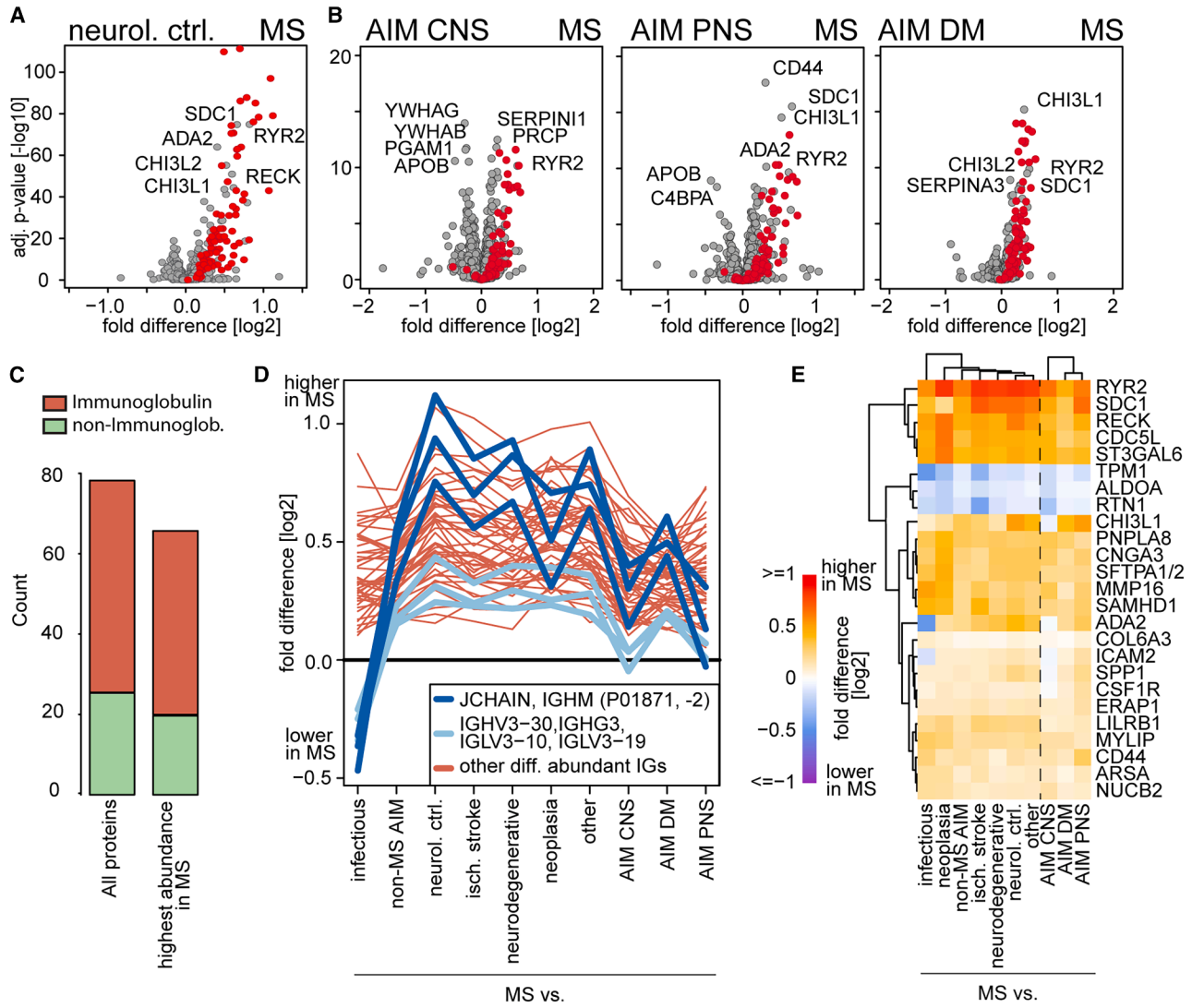


Figure 3. MS-associated CSF proteome alterations

Fold difference and significance values were obtained from the regression model (see STAR Methods).

(A) Protein regulation between MS and neurological controls. Immunoglobulins are shown in red.

(B) Protein regulation between MS and subtypes of non-MS autoimmunity (AIM), including demyelinating (DM) AIM. Immunoglobulins ORFs are shown in red.

(C) Number of proteins that are differentially abundant between MS and all other neurological conditions. AIM is pooled as one condition.

(D) Abundance of immunoglobulins, relative to MS. Immunoglobulins are included if present in (C). IgM-related proteins are shown in dark blue, IgG proteins with a similar profile are shown in light blue. Other Ig proteins are shown in red.

(E) Regulation of non-immunoglobulin proteins that are significant, comparing any other disease to MS.

See also Figure S3.

most of these proteins informed about one or two diseases. The top 20 features overall and those for individual diseases included strong positive controls but also additional associations of therapeutic or mechanistic interest, which we again list in more detail in Table S4 (Figures 4D, 4E, and S4E). For MS, this, for instance, included MBP, CHIT1, or the aforementioned potential drug target CTSW. SHAP-based clustering increased sensitivity beyond the top 20 proteins, additionally revealing CD44, RYR2, and IGHG1 (Figures 4E–4G). For neoplasia, this approach produced surprisingly many oncogenes, linked to brain or other types of cancer, including the strong positive controls IDH1 and EGFR, as well as tumor suppressors predicting against the disease (Figures 4E and S4F). Further expected or insightful associations were NEFM for neurodegeneration but also CNS AIM, synuclein beta (SNCB) for stroke, CXCL10 for infection, and that controls were defined by the absence of other diseases' markers.

At the pathway level, adaptive immunity predicted for MS, as expected, whereas innate immunity predicted against it, consistent with stronger innate responses in infection (Figure 4H).

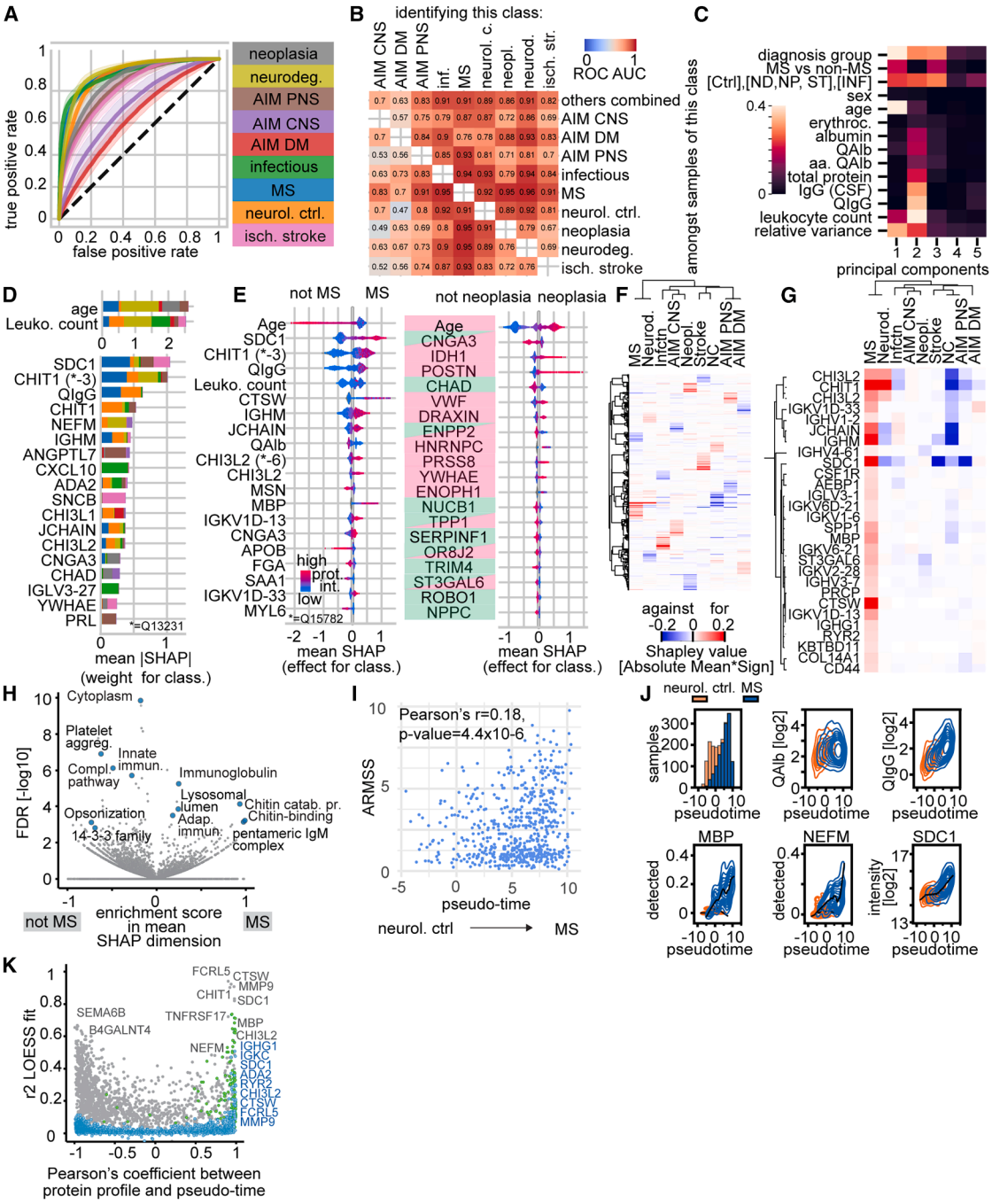


Figure 4. Survey of multi-class disease prediction and pseudo-time analysis of MS vs. neurological controls

(A) ROC curves. Individual neurological conditions vs. all other conditions combined. AIM, autoimmunity; DM AIM, demyelinating AIM not fulfilling MS criteria.

(B) ROC AUC matrix. Model identifying given disease (*x*) from all others combined or another individual disease (*y*). Multi-class model with potential error of predicting neither disease *x* nor *y*; therefore, performance for identification of *x* is stated (STAR Methods). Thus, non-equivalence across diagonal.

(C) Relationship of model feature weights, clinical covariates, and separation of diseases. Principal-component analysis of the model's SHAP values. Explained variability (color code) in component projections (*x*) accounted for by $\log_2$-transformed covariates or disease status (*y*).

(D) Most influential features for the entire model ranked by their mean absolute SHAP values. Contribution to class-specific decision-making is color coded.

(E) Twenty most influential features for the classification of MS (left) and neoplasia (right). Samples are datapoints, the intensity of a given protein in a sample relative to the intensity in other samples is color coded; red denotes high protein intensity and blue is low intensity. The SHAP value measures a protein's contribution to the model's decision-making and is shown on the *x* axes. A positive SHAP value means that the protein's specific intensity in that sample (e.g., high or low; see color code) predicts for a given disease. In contrast, a negative SHAP value corresponds to prediction against that disease, making other diseases more likely in that sample. For neoplasia, oncogenes are shown in red, tumor suppressors in green, and contextual role in both colors.

(legend continued on next page)

Hypoxia pathways marked neoplasia and glycolysis/fibrinolysis ischemic stroke (Figure S4G). Thus, our large-scale analysis efficiently disentangled pathological mechanisms.

Proteome-based staging of individuals with MS

Donors in cross-sectional cohorts reflect a range of disease severity, both regarding symptoms and molecular phenotypes. To capture this, we calculated sample positions along a pseudo-time axis, analogous to those in single-cell RNA sequencing but here reflecting proteomic perturbation rather than chronology (STAR Methods). As expected, neurological controls clustered at early pseudo-times, persons with MS at later ones, and pseudo-time staging correlated with the age-related multiple sclerosis severity score (ARMSS) at CSF sampling (Figures 4I, 4J, and S4H). Pseudo-time staging was not biased by age, sex, erythrocyte contamination, or protein identification depth, and the MS disease courses were evenly distributed on this scale in our cohort (Figure S4H). Although immunoglobulin load and BCB impairment are generally correlated, only immunoglobulins increased along the pseudo-time (Figures 4J and S2E). Detection rates of MBP and neurofilaments, particularly NEFM, also rose with pseudo-time, underscoring the sensitivity of this analysis. Neurofilament proteins were elevated in MS compared with controls, reaching statistical significance in the Astral data. As clinical severity in MS is associated with immunoglobulins and neurofilaments, but not with BCB impairment, our pseudo-time results separate meaningful biology from confounders.[58–60]

Next, we extended the protein profile concept to the entire proteome, calculating both intensity and detection rate profiles and ranking them based on their explained variability and linearity (STAR Methods; Figures 4K and S4I). This yielded strong positive controls, for instance, TNFRS17/BCM, known as B cell maturation antigen, as well as MBP, NEFM, CHIT1, SDC1, and various immunoglobulins. MMP9, FCRL5, and the aforementioned CTSW had the strongest MS association. MMP9 correlates with MS disease activity[61]; the role of the other two in MS is not known.

FCRL5, an immunomodulatory Fc receptor and a marker of atypical memory B cells (atMBCs),[62–66] is associated with MS susceptibility. atMBCs accumulate in chronic infection and various autoimmune diseases.[62,67] Generally, atMBCs are induced by interferon (IFN)-γ, which is abundant in MS lesions and increased during active disease.[62,68,69] Although IFN-γ was too low in abundance to detect, we observed a concordant profile of CXCL10, which B cells secrete in response to IFN-γ (Figure S4I).[70] Ectopic lymphoid follicles are common in MS, located near the CSF in brain tissue, are rich in B cells, and may further promote atMBC generation.[71,72] B cell-depleting therapies in MS should therefore be applied with careful monitoring of immune competence and infection risk to balance their therapeutic impact on autoimmune pathology with the need to prevent infections.[73]

Biomarker panel for OCB-positive and -negative MS

For a clinical MS diagnosis, other inflammatory conditions are the most challenging to separate,[74] which aligns with our classification of CSF proteomes (Figure 4B). OCBs, the established diagnostic CSF biomarker for MS, are not present in all MS cases, and identifying MS in symptomatic individuals with negative OCB status is complicated.[5] In our cohort of >1,300 MS patients, only 72% had CSF-specific OCBs, and in another 13% the OCB testing result was borderline, highlighting the prevalence of the issue.

To delineate markers that address the above two challenges, we selected an approximately 1,600-sample-strong, covariate-stratified subset of MS, inflammatory conditions, and controls and re-measured them with our improved Astral workflow (Figure 5A). MS CSF had a higher abundance of immunoglobulins and, in particular, B cell markers when OCB positive (Figure 5B). With more than 2,000 proteins and subject age and sex available as features for cross-disease machine learning, MS classification performance saturated at around 20–30 features (Figure S5A; STAR Methods). Across all samples, a 30-feature proteomic panel performed equally to the reference set of clinically established markers and additional covariates, including OCB status, IgG ratio, albumin ratio, leukocyte count, age, and sex (Figures 5C–5E and S5B–S5D). However, among OCB-negative samples, the proteomic panel outperformed the clinical markers by about 0.1–0.2 ROC AUC in discerning MS and CNS or PNS autoimmune diseases, including in a 600-strong replication cohort (Figure 5E). As a control, classification of OCB-negative samples by any model did not benefit from limiting the training set to OCB negatives (Figure S5E).

The marker panel contained both previously discussed marker proteins for MS and some without prior prominence in MS, such as FCRL5 or CTSW (Table S4). Additional proteins reflected pathways including myelin pathology (MBP), axonal degeneration (NEFM), microglia/innate immunity (C1QB), immunoglobulins (JCHAIN and IGKC), T (CD7) and memory B (CD27) cell load, overall cellular content (ACTG), or BCB impairment

(F) SHAP values across proteins (*y*) and diseases (*x*). Directionality interpretation as in (E).
(G) SHAP subcluster of MS-predicting features. Zoom of (F).
(H) Enrichment of annotation terms in SHAP values informing about MS.
(I) Correlation of pseudo-time staging of individuals from neurological controls to MS, with the neurological ARMSS score measured at CSF sampling.
(J) Characterization of MS pseudo-time. Top: distribution of MS and neurological control samples, associated QAlb and QIgG distributions along pseudo-time. Bottom: protein distributions of MBP, NEFM, and SDC1 along the pseudo-time, based on either the measured protein intensity or the protein detection rate local to the pseudo-time. Further, distributions are averaged into a protein profile, shown by a black line, using LOESS regression (STAR Methods).
(K) Prioritization of protein profiles according to linearity along pseudo-time (*x*) and explained variance (*y*). Profiles based on intensities (light blue) or detection rate (gray). Immunoglobulins shown in green, regardless of profile type. Trajectories with a positive Pearson's coefficient (*x* axis) correspond to proteins with high abundance in MS, and those with negative coefficients to proteins with low abundance in MS compared to neurological controls, relative to the proteome.
See also Figure S4.

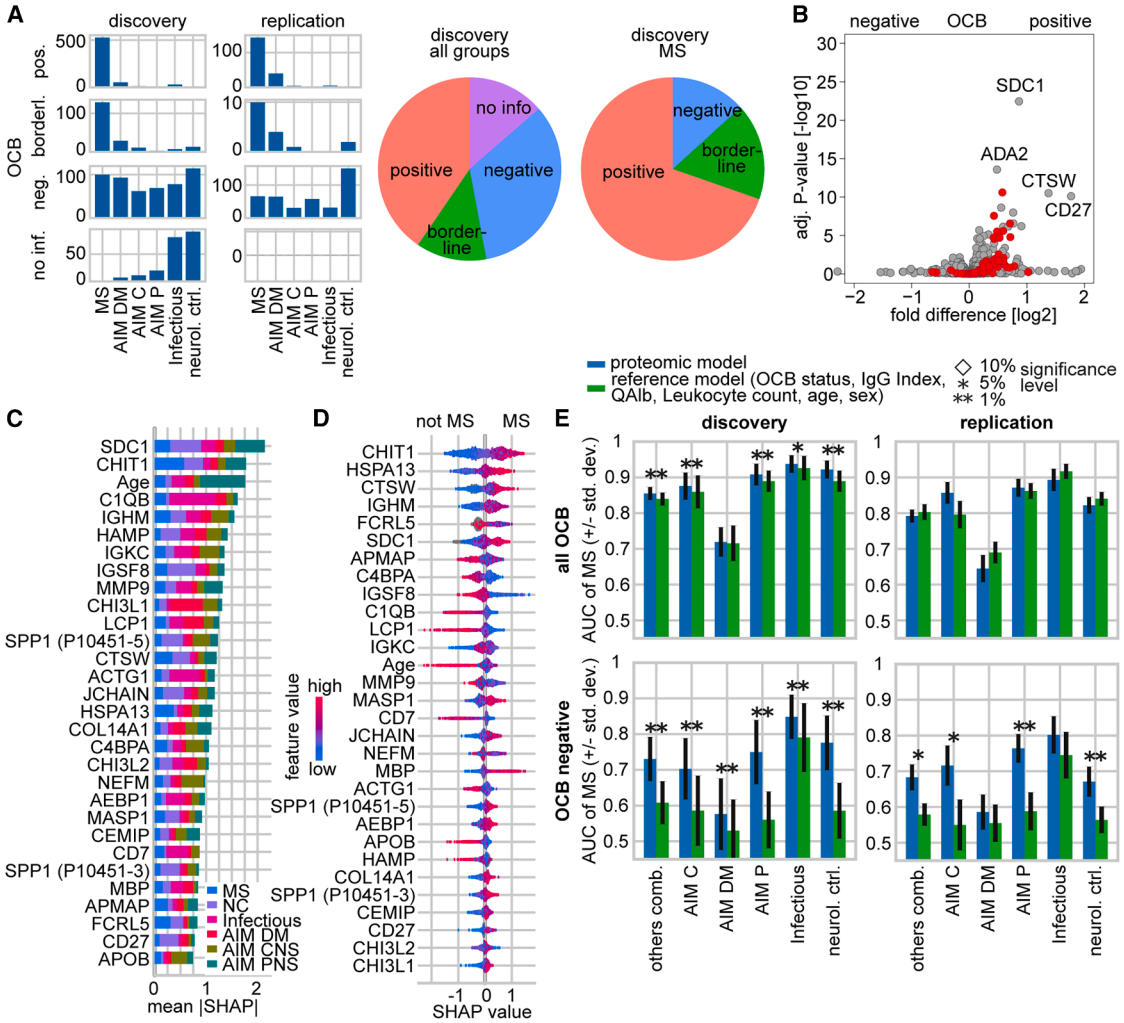


Figure 5. Discovery of a proteomics protein panel to separate oligoclonal band (OCB)-pos. and OCB-neg. MS from CNS inflammation

(A) Sample statistics of the stratified discovery subset and replication cohort.

(B) Protein regulation between OCB-pos. and OCB-neg. MS. Immunoglobulins are shown in red.

(C) Importance of the 30 most important features by mean absolute SHAP values. Contribution to disease-specific decision-making is color coded.

(D) Information content of features for prediction of MS across samples. Samples are shown as datapoints; intensity of the feature (protein) in the sample is color coded. SHAP value as a metric for contribution to decision-making on the *x* axes.

(E) ROC AUC for MS in the proteomic model (blue) and in the reference model of clinical CSF markers (green) in samples from all OCB groups (top) and in the OCB-negative group only (bottom). Discovery (left) and replication (right) cohorts. Standard deviation error bars are shown. ROC AUC datapoints derive from cross-validation splits (discovery) or bootstraps of the replication cohort predictions (replication). Significance from paired analysis, proteomic vs. reference model (STAR Methods).

See also Figure S5.

(APOB). This illustrates that biomarker panels can excel through quantitative pathway strength rather than single high-specificity markers.

Targeted assay for routine clinical application

Adoption of biomarkers requires assays for routine measurements, which typically creates a translational hurdle for novel ones from exploratory proteomic studies. We built a targeted assay for the 29 proteins on the recently introduced Stellar mass spectrometer (Figure S6A; STAR Methods). Analyte peptides were quantified as a ratio to 20 fm of isotope-labeled standards, which overall correlated with DIA quantities but improved data for low-abundance proteins (Figures S6B–S6E). We again measured the stratified discovery and replication cohort and selected 22 peptides, one for each protein, based on signal quality and workflow CV in discovery (STAR Methods). The biological effect across donor samples was at least twice, but typically 5- to 15-fold, greater than the workflow CV (Figures S6F and S6G). Classification of diseases based on targeted measurements overall mirrored results based on discovery proteomics (Figures 6A–6D, S6H, and S6I). Notably, improved quantification by targeting further boosted the performance lead of proteomic

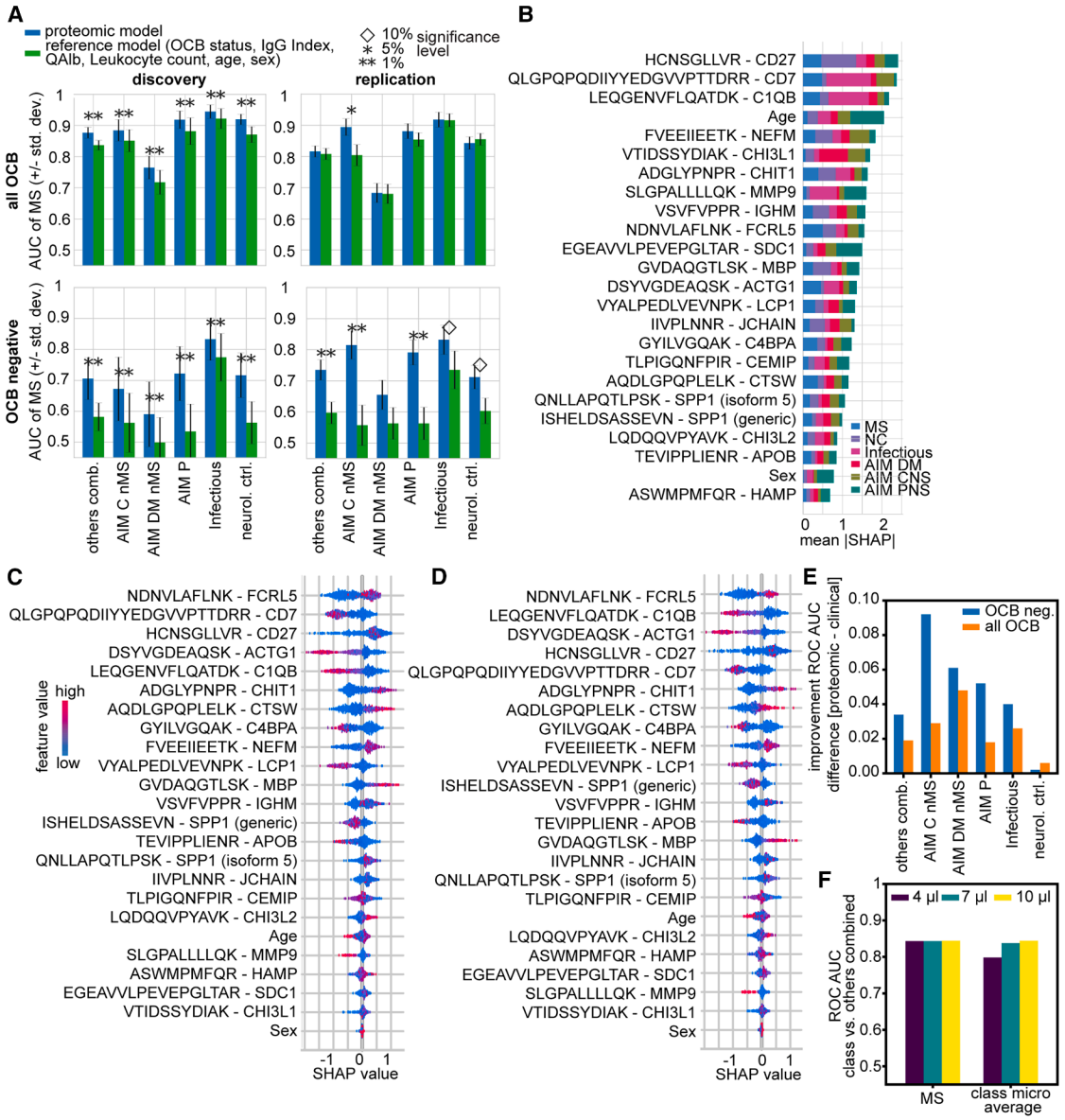


Figure 6. Targeted assay for a 22-protein panel to separate OCB-pos. and OCB-neg. MS from CNS inflammation

(A) ROC AUC for MS in the proteomic model (blue) and in the reference model of markers used in clinical practice (green). Performance in OCB-neg. samples (bottom) and samples across all OCB statuses (top). Discovery cohort (left) and replication cohort (right). Standard deviation error bars are shown. ROC AUC datapoints derive from cross-validation splits (discovery) or bootstraps of the replication cohort predictions (replication). Significance from paired analysis, proteomic vs. reference model (STAR Methods).

(B) Feature importance summary. Importance as mean absolute SHAP values. Contribution to class-specific decision-making is color coded. Discovery data.

(C) Feature importance for MS prediction in samples of all OCB states. Samples are shown as datapoints; intensity of the protein in the sample is color coded. SHAP value as a metric for contribution to decision-making on the *x* axes, with contributions to predictions for the diagnosis group as positive values and contributions against the prediction of that disease as negative values.

(D) Feature importance for MS prediction in OCB-neg. samples. Otherwise, equivalent to (C).

(E) Performance improvement by targeted measurement. Double normalization: first, proteomic model minus clinical model performance; second, targeted minus discovery proteomics (DIA) differences.

(F) Performance robustness toward variable CSF amounts loaded onto the mass spectrometer in testing data, namely two 96-well plates of the replication cohort. Training data remained the initial discovery cohort data, where 10 µL of CSF peptides were measured.

See also Figure S6.

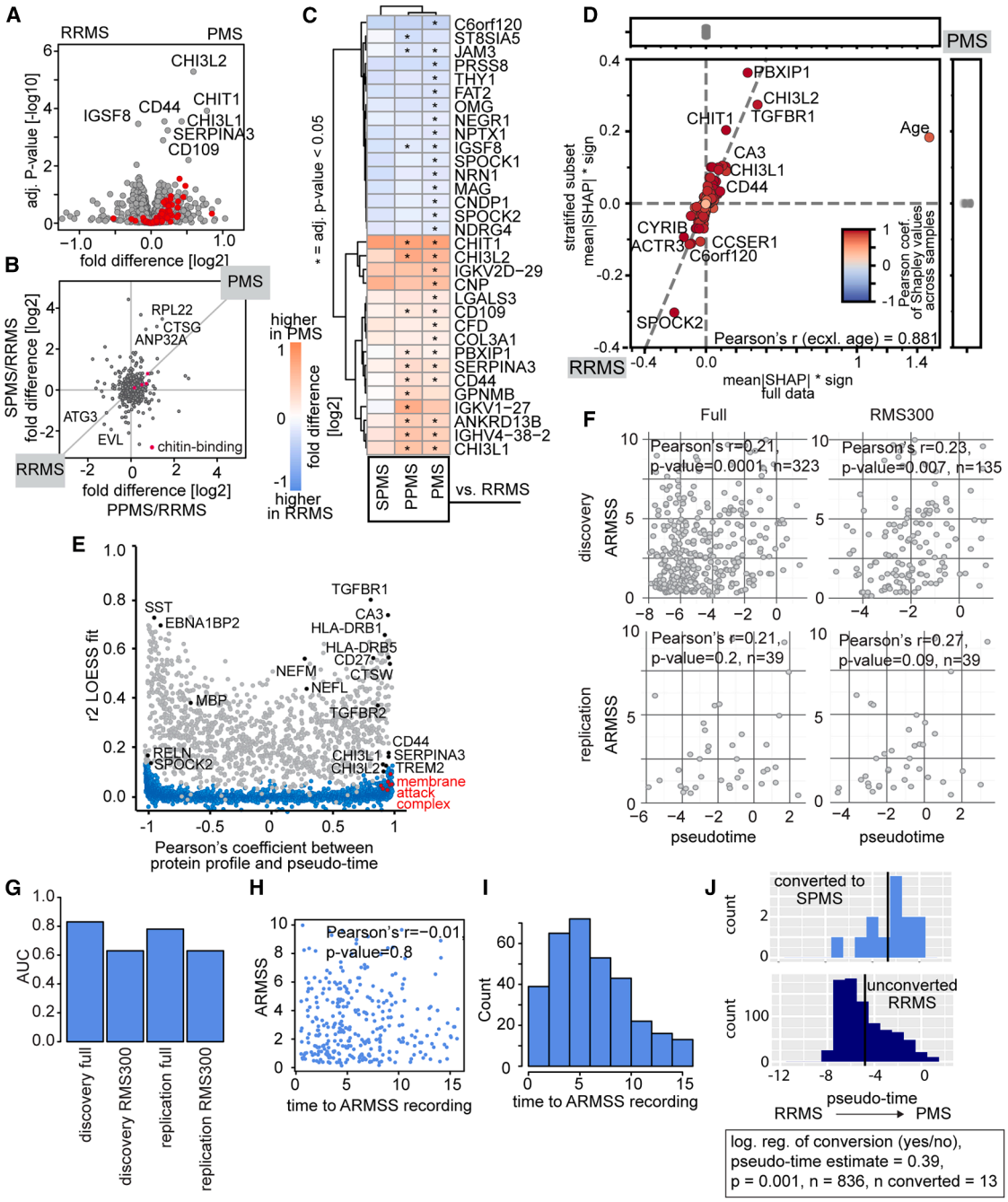


Figure 7. CSF alterations in relapsing and PMS, and disease course conversion

(A) Protein regulation between relapsing-remitting MS (RRMS) and progressive MS (PMS). Immunoglobulin ORFs are shown in red.

(B) Effect sizes of protein regulation between RRMS and PMS subtypes. Primary PMS (PPMS) and secondary PMS (SPMS).

(C) Effect sizes for all proteins that are significantly regulated (adj. $p < 0.05$, marked by *) in at least one of the subtype comparisons. RRMS, relapsing-remitting MS; PPMS, primary PMS; SPMS, secondary PMS; PMS, progressive MS combined.

(D) Comparison of SHAP values of models for classification of progressive and remitting MS (PMS and RRMS) in the full dataset (*x* axis) or in the subset stratified for age and leukocyte count (*y* axis). Protein features with non-zero SHAP values in both models in the center plot, proteins with non-zero SHAP values in only one model on either axis. Aggregated SHAP values are shown on the axes. Proteins are color coded by the correlation of SHAP values between the two models across samples.

(E) Prioritization of trajectories according to linearity (*x* axis) along pseudo-time and variability explained by the trajectory (*y* axis). Trajectories are color coded, with intensity trajectories in blue and local detection rate trajectories in gray. Outliers and selected proteins of interest are highlighted. Trajectories with a positive Pearson's coefficient (*x* axis) correspond to proteins with high abundance in PMS, and those with negative coefficients to proteins with high abundance in RMS.

(legend continued on next page)

markers relative to the clinical reference model by 0.02–0.1 ROC AUC (Figure 6E). Further, the biological information was stored in the abundance relationships of panel proteins to one another, not their ratios to the fixed 20 fm standard. Accordingly, injecting 30% or 60% less CSF specifically in testing data, as expected, did not alter performance, an inherent robustness advantage of panels compared with single-analyte markers (Figure 6F).

CSF signatures of relapsing-remitting MS and PMS

Diverging pathological mechanisms between relapsing-remitting MS (RRMS) and progressive MS (PMS) limit the effectiveness of standard immunosuppression, underscoring the need for mechanistic insight for therapeutic innovation. We pooled primary and secondary PMS (PPMS and SPMS) to increase statistical power in comparisons with RRMS due to the naturally lower PMS prevalence and the challenge posed by the gradual RRMS-to-SPMS conversion. Nevertheless, PPMS and SPMS exhibited similar proteomic alterations, consistent with their clinical similarity (Figures 7A–7C and S7A–S7D).

We prioritized discriminant proteins for RRMS and PMS by machine learning (Figure 7D). Because RRMS usually manifests at a younger age than PPMS and is associated with more immune cells in the CSF, we addressed this imbalance by creating a stratified data subset that we analyzed alongside the full dataset. Reassuringly, the balanced subset specifically reduced the weight of age, but proteins behaved concordantly with the full dataset. Protein associations combining machine learning and differential expression included several unexpected findings, including CNP and NPTX1, as well as positive controls, for instance, CHI3L1, CHI3L2, CHIT1, and CD44 (Table S4). Consistent with current models of MS pathology, markers of adaptive immunity were more predictive for RRMS, whereas innate immune signatures predominated in PMS (Figures 7D, 7E, and S7E). SERPINA3, an emerging marker of glial exhaustion, was also associated with PMS. Glial exhaustion is increasingly implicated as a contributing mechanism in disease progression. Further, transforming growth factor (TGF) signaling and, in particular, TGFBR1 likewise predicted for PMS. In our classification surveying neurological conditions, this pathway predicted for neurodegenerative diseases (Figure S4G), consistent with the neurodegenerative pathology of PMS in MS.

Proteome-based staging of MS disease course evolution

To model the molecular continuum between relapses and progression and to explore its clinical relevance, we applied our pseudo-time analysis to stage individuals between RRMS and PMS. This correlated consistently with the ARMSS coefficient, across full, age-, and leukocyte-stratified data as well as in the replication cohort (Figures 7F–7I). We used the latest available disability scores, typically 5 but up to 15 years after CSF sampling, which did not confound the pseudo-time correlation but highlighted predictive potential.

Additionally, individuals with RRMS who later converted to SPMS were significantly shifted to higher pseudo-times than those without observed conversion (Figure 7J). Median time to conversion after CSF sampling was 3 years, and this delay was inversely associated with the pseudo-time (Pearson's $r = -0.73$, linear regression: $p = 0.004$, slope $= -0.46$, $R^2 = 0.5$, $N = 13$ converters). These findings highlight the potential utility of proteome-based staging to inform long-term treatment decisions.

Given this clinical relevance, we explored protein profiles along the pseudo-time as before. Outliers included several of the proteins observed in our above analyses on MS generally (Figures 7E and S7F–S7J; Table S4). Among RRMS-associated proteins, we detected a human interactor of the Epstein-Barr virus (EBV) protein EBNA1, which itself influences MS pathogenesis.[75–79] Further, EBV infection generally is a strong MS risk factor.[80] Conversely, PMS associations were linked to innate immunity, the complement system, and human leukocyte antigen class II molecules (HLA-DRB1 and HLA-DRB5), as well as CTSW and FCRL5, which are part of our MS biomarker panel (Figures 7E, S7I, and S7J). Notably, the HLA-DRB1 allele 15:01 is associated with increased susceptibility to MS and worse health outcomes.[81–83] In summary, CSF proteomics captures the molecular continuum from RRMS to SPMS, with insightful protein associations.

DISCUSSION

We have developed an in-depth high-throughput workflow for mass-spectrometry-based proteomics to analyze CSF. Using the timsTOF Pro 2 instrument, we detected 1,500 proteins per sample across 5,000 samples, a scale competitive with binder-based CSF proteomic technologies, which also differ in properties.[16,24,84,85] Using the Orbitrap Astral, we subsequently raised protein identifications to 2,100 per sample, throughput to 100 SPD, and made protein identification largely independent of BCB impairment, addressing key limitations of earlier workflows. These technological advances now enable proteomic studies of CSF at scale, with high completeness and robustness.

Together with cohort diversity, this enabled the systematic assessment of disease and covariate effects and their interplay. So far, sex and age had received most attention[26,86]; however, our study revealed BCB impairment to be the dominant source of variability, explaining roughly three times more variance than disease or any other covariate. Subsequent analysis revealed this to be true in OLINK data as well. This underscores the need to assess and correct for BCB impairment, which is not yet routine practice. For studies lacking QAlb information, which

(F) Relation of pseudo-time separating RMS (enriched at low pseudo-time) and PMS (high) with the latest available ARMSS score. Either all available MS samples in the discovery cohort were used (full) or only the age- and leukocyte-count-stratified subset (RMS300). Machine learning classifiers of RMS/PMS status from these two dataset variants of the discovery cohort were applied to the replication cohort.
(G) ROC AUC for the settings represented in (F).
(H) Distribution of ARMSS scores in relation to the time difference between CSF collection and ARMSS scoring.
(I) Distribution of the time difference between CSF collection and ARMSS scoring.
(J) Pseudo-time distribution of individuals with RRMS that converted to SPMS (top) or that did not convert (bottom).
See also Figure S7.

is the standard metric of BCB impairment, we provide proteome-derived substitutes. Although disease effects varied in magnitude, they correlated across conditions, reflecting common processes such as inflammation or release of intracellular material. Accordingly, we found only about 20 proteins differentially abundant in MS vs. other neurological diseases. Unsurprisingly, many more proteins are found to be MS-associated across literature studies that compare MS to fewer or less-related diseases.[21,87–90] This emphasizes the importance of focusing on all clinically relevant control conditions to identify specific alterations.

Our study reproduced many associations between MS and individual proteins from past investigations of CSF or tissues or from genetic studies. These include CHIT1, CHI3L1, CHI3L2, MBP, GFAP, SPP1, MMP9, SDC1, and neurofilaments, which serve as positive controls. Several of these had only been shown in comparisons with healthy or non-inflammatory controls, whereas our data extend this to other inflammatory diseases (Table S4). For PMS, we identified emerging candidates, such as SERPINA3, which are present in the brain and point to glial exhaustion[91] as a pathological factor in progression. Regarding MS generally, we found some protein associations of therapeutic interest. For example, RYR2 and CSF1R inhibition ameliorates experimental AIM, and drugs targeting these proteins already exist and are approved for cardiac conditions or under evaluation in oncology.[92–98] CTSW, another candidate, modulates mucosal inflammation.[99] More broadly, because most drug targets are proteins, large-scale human proteomics offers a rational basis for prioritizing candidates. With improved coverage and scalability, CSF proteomics is now positioned to support this translational step.

Biomarker development remains the primary clinical application of CSF analysis. Regarding MS, differential diagnosis has persistent challenges, especially in OCB-negative individuals. We delineated a 22-protein panel that outperformed current markers in this subgroup, while maintaining comparable performance across OCB statuses, suggesting broader applicability. The panel achieves discriminatory power by integrating the levels of fairly disease-specific proteins and common processes such as inflammation, indicating that uniquely specific and sensitive single markers may be aspirational but are neither easily found nor actually required. In a related development, neither GFAP nor NEFL are specific to MS, but their ratio has been discussed as a biomarker separating MS from neuromyelitis optica spectrum disorder (NMOSD).[100]

Biomarker translation is often hindered by fragmentation of evidence across studies.[101,102] Here, we demonstrated how technological advances have enabled an integrative approach that combines the key aspects in a single study: large-scale protein and sample analysis, inclusion of diverse and clinically relevant controls, multi-protein panel assembly, direct benchmarking against established markers, assay transfer to a routine-compatible format, and verification of clinical benefit in a replication cohort. In terms of clinical applicability, targeted mass spectrometry is already established in clinical laboratories for small molecules and some proteins. The technology offers low marginal cost per sample and efficiently covers multiple biomarkers of a panel, offering seamless assay translation.

Treatment options in MS range from mild immunomodulators to strong immunosuppressive drugs that reduce acute inflammation. The latter are most effective at slowing disability accumulation when initiated early, although they carry greater side-effect risks. The emergence of therapies that target progression itself, such as tolebrutinib, makes it valuable to identify patients with elevated progression independent of relapse activity (PIRA) early in the disease course.

However, current clinical indicators such as relapse history, MRI lesion burden, and disability scores offer only modest prognostic value and often require longitudinal observation, limiting their utility for guiding timely treatment decisions.[103,104] Emerging CSF biomarkers such as NEFL, GFAP, and CHIT1 reflect distinct pathological processes and are differentially useful for acute and long-term assessment.[8,11,105,106] Multi-protein panels integrate such complexity to improve performance, exemplified by the Octave test for acute MS disease activity, though large-scale validation is still needed.[34,35]

Additionally, CSF proteomic profiling can broadly predict levels of disability years after CSF sampling.[16] With our pseudo-time staging, we also expand the prognostic range: this single proteome-derived metric simultaneously correlates with multiple clinical outcomes, namely disability, RRMS-to-SPMS conversion risk, and time to conversion. Another major advantage is that training the pseudo-time prediction is independent of clinical outcomes. Accordingly, samples lacking outcome data can still contribute to training. Outcome heterogeneity, such as varying time to the available outcome data or rare events such as SPMS conversion, is difficult for direct outcome prediction but not for pseudo-time staging. This is a particular boon for cross-sectional cohorts, which predominate in practice. Our study showcases that even cross-sectional cohorts of CSF sampled at diagnosis can reveal long-term prognostic patterns, which are especially relevant for guiding emerging PIRA-targeted treatment strategies. Analysis of controlled longitudinal cohorts will be crucial for clinical translation. Use of the integrative approach presented here for MS differential diagnosis is a promising advance toward routine-compatible assays.

Limitations of the study

Mass-spectrometry-based proteomics generally quantifies proteins relative to the overall proteome rather than providing absolute concentrations of individual analytes. Protein detectability can depend on the sample matrix and may vary across individuals. As with other proteomic techniques, not all proteins are consistently observed across all samples. Nevertheless, our study shows that mass spectrometry is well suited for discovery in large-scale cohorts, and interpretation is best focused on identified associations rather than missing signals.

In this study, we addressed key translational barriers by converting discovery data into a targeted assay with spike-in standards, which enabled successful replication in an independent cohort and lays the groundwork for routine laboratory workflows. For clinical qualification, however, further steps will be required, such as establishing absolute quantification using standards with defined concentrations. Quantification boundaries, linearity, and measurement precision across the dynamic range need to be benchmarked across instruments and sites. Validating

clinical benefit will require large, well-characterized cohorts, ideally through prospective, multi-center studies. Such studies will also need to identify how to best integrate advanced CSF biomarkers with MRI findings.

Determining diagnostic performance depends on independent ground truth, such as post-mortem confirmation or long-term clinical follow-up, both difficult to obtain for CSF. In practice, many studies, particularly at the discovery and early validation stage, are limited to well-curated cross-sectional cohorts like ours. This makes it essential to extract maximal value from such cohorts through rigorous design and comprehensive analysis as we have done here.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Matthias Mann (mmann@biochem.mpg.de).

Materials availability

This study did not generate new, unique reagents.

Data and code availability

The mass spectrometry data have been deposited at the ProteomeXchange Consortium via the PRIDE partner repository in three datasets. Dataset PRIDE: PXD045058 pertains to the full discovery cohort measured on the timsTOF Pro 2, as well as the pilot cohort measured on the Orbitrap setups (Q Exactive HF-X, Exploris 480). Dataset PRIDE: PXD064570 contains the Astral-measured data of the stratified subset of the discovery cohort and the replication cohort. Dataset PRIDE: PXD064513 contains the targeted mass spectrometry data.

All software used to collect and analyze data for this work was either published previously or is commercially available. The code generated to analyze the data of this study is deposited on the OSF platform (DOI: https://doi.org/10.17605/OSF.IO/JM2AK). This repository also contains additional tables with analysis results and proteome data, including those from the pilot study. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank our colleagues in the Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, for valuable discussions. We thank Katharina Zettl and Verena Grummel for technical support and for preparing the StageTips for peptide clean-up and sample preparation. We thank Andreas Brunner for input in setting up the workflow and Till Andlauer for support in designing the pilot study and analyzing its data. We thank Vadim Demichev for guidance on how to use DIA-NN for mass spectrometry data processing and Philipp Remes (Thermo) for support with the operation of the Stellar mass spectrometer and analysis of the resulting data.

Funding was provided by the Max-Planck Society for Advancement of Science and the German Federal Ministry of Research, Technology, and Space through project CLINSPECT-M. S.R. was supported by the Helmholtz Association under the joint research school "Munich School for Data Science"-MUDS. M.T. received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 874839 (ISLET). A.B. received funding from the German Federal Ministry of Research, Technology, and Space (grant 01ZZ2102B). C.G. received funding from the German Federal Ministry of Research, Technology, and Space, the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), the Hertie Foundation, and the Hans and Klementia Langmatz Stiftung. B.H. received funding for the study from the European Union's Horizon 2020 Research and Innovation Program (grant MultipleMS, EU RIA 733161 and grant WISDOM) and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy within the framework of the Munich Cluster for Systems Neurology (EXC 2145 SyNergy—ID 390857198).

AUTHOR CONTRIBUTIONS

J.M.B., P.S., and M.T. established the proteomic workflow and performed the measurements. J.M.B., W.-F.Z., and C.A. processed the proteomic raw data. J.M.B., M.W., and M.B.L. established the targeted mass spectrometry assay and processed its raw data. C.G., F.H., L.P., A.B., C.M., and B.H. collected clinical samples and designed the cohorts. J.M.B., C.M., S.R., M.T.S., and I.B. performed the biological analysis of proteomic data. B.S., F.J.T., C.G., B.H., and M.M. supervised and guided the project. J.M.B., C.M., C.G., B.H., and M.M. wrote the manuscript. C.G., B.H., and M.M. conceptualized the study.

DECLARATION OF INTERESTS

M.M. is an indirect investor in Evosep Biosciences. F.J.T. consults for Immunai Inc., Singularity Bio B.V., CytoReason Ltd, and Cellarity and has ownership interest in Dermagnostix GmbH and Cellarity. A.B. has received consulting and/or speaker fees from Alexion, Argenx, Biogen, Horizon, Novartis, Roche, and Sandoz/Hexal, and his institution has received compensation for clinical trials from Alexion, Biogen, Merck, Novartis, Roche, and Sanofi Genzyme, all of which are outside the present work. B.H. has served on scientific advisory boards for Novartis; he has served as a DMSB member for AllergyCare, Sandoz, Polpharma, Biocon, and TG therapeutics; and his institution received research grants from Regeneron and Roche for multiple sclerosis research. He has received honoraria from GLG consulting. He holds part of two patents: one for the detection of antibodies against KIR4.1 in a subpopulation of patients with multiple sclerosis and one for genetic determinants of neutralizing antibodies to interferon. None of these conflicts are relevant to the topic of the study.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used ChatGPT by OpenAI and Claude by Anthropic in order to improve the readability and conciseness of the manuscript. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Study collective – Discovery cohort
  - Study collective – Replication cohort
- METHOD DETAILS
  - Clinical sampling, measurements, and quality control
  - Dataset overview
  - Sample organization and pooled samples for quality control
  - Sample preparation
  - Loading of Evotips with sample peptides for measurement
  - Preparation of samples for library generation
  - Mass spectrometry for discovery proteomics
  - Library generation for discovery proteomics
  - Processing of discovery proteomics raw data
  - Heavy Isotope-labeled peptides for targeted mass spectrometry
  - Spectral library for targeted mass spectrometry
  - Targeted mass spectrometry assay and cohort measurement
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Bioinformatic analysis
  - Differential abundance analysis

- Protein annotation and enrichment analysis
- Machine learning-based sample classification and model explanation
- Analysis of biomarkers separating MS from other CNS inflammation
- Sample pseudo-times and protein trajectories

- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2026.01.017.

Received: December 11, 2023
Revised: July 28, 2025
Accepted: January 23, 2026
Published: February 25, 2026

REFERENCES

1. Thompson, A.J., Banwell, B.L., Barkhof, F., Carroll, W.M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., Freedman, M.S., et al. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol. *17*, 162–173. https://doi.org/10.1016/S1474-4422(17)30470-2.
2. Zürrer, W.E., Cannon, A.E., Ilchenko, D., Gaitán, M.I., Granberg, T., Piehl, F., Solomon, A.J., and Ineichen, B.V. (2024). Misdiagnosis and underdiagnosis of multiple sclerosis: A systematic review and meta-analysis. Mult. Scler. *30*, 1409–1422. https://doi.org/10.1177/13524585241274527.
3. Wang, Y., Bou Rjeily, N., Koshorek, J., Grkovski, R., Aulakh, M., Lin, D., Solomon, A.J., and Mowry, E.M. (2023). Clinical and radiologic characteristics associated with multiple sclerosis misdiagnosis at a tertiary referral center in the United States. Mult. Scler. *29*, 1428–1436. https://doi.org/10.1177/13524585231196795.
4. Tieppo, E.M. de S., Silva, G.D., da Silva, T.F.F., de Araujo, R.S., de Oliveira, M.B., Spricigo, M.G.P., Pimentel, G.A., Campana, I.G., Castrillo, B.B., Mendes, N.T., et al. (2023). Misdiagnosis in multiple sclerosis in a Brazilian reference center: Clinical, radiological, laboratory profile and failures in the diagnostic process-Cohort study. Mult. Scler. *29*, 1755–1764. https://doi.org/10.1177/13524585231199323.
5. Katsarogiannis, E., Landtblom, A.-M., Kristoffersson, A., Wikström, J., Semnic, R., and Berntsson, S.G. (2023). Absence of Oligoclonal Bands in Multiple Sclerosis: A Call for Differential Diagnosis. J. Clin. Med. *12*, 4656. https://doi.org/10.3390/jcm12144656.
6. Deisenhammer, F., Zetterberg, H., Fitzner, B., and Zettl, U.K. (2019). The Cerebrospinal Fluid in Multiple Sclerosis. Front. Immunol. *10*, 726. https://doi.org/10.3389/fimmu.2019.00726.
7. Tumani, H., Hartung, H.-P., Hemmer, B., Teunissen, C., Deisenhammer, F., Giovannoni, G., and Zettl, U.K.; BioMS Study Group (2009). Cerebrospinal fluid biomarkers in multiple sclerosis. Neurobiol. Dis. *35*, 117–127. https://doi.org/10.1016/j.nbd.2009.04.010.
8. Di Filippo, M.D., Gaetani, L., Centonze, D., Hegen, H., Kuhle, J., Teunissen, C.E., Tintoré, M., Villar, L.M., Willemse, E.A.J., Zetterberg, H., et al. (2024). Fluid biomarkers in multiple sclerosis: from current to future applications. Lancet Reg. Health Eur. *44*, 101009. https://doi.org/10.1016/j.lanepe.2024.101009.
9. Levraut, M., Landes-Chateau, C., Mondot, L., Cohen, M., and Lebrun-Frenay, C. (2025). The Kappa Free Light Chains Index and Central Vein Sign: Two New Biomarkers for Multiple Sclerosis Diagnosis. Neurol. Ther. *14*, 711–731. https://doi.org/10.1007/s40120-025-00737-7.
10. Hegen, H., Walde, J., Berek, K., Arrambide, G., Gnanapavan, S., Kaplan, B., Khalil, M., Saadeh, R., Teunissen, C., Tumani, H., et al. (2023). Cerebrospinal fluid kappa free light chains for the diagnosis of multiple sclerosis: A systematic review and meta-analysis. Mult. Scler. *29*, 169–181. https://doi.org/10.1177/13524585221134213.
11. Toscano, S., and Patti, F. (2021). CSF biomarkers in multiple sclerosis: beyond neuroinflammation. Neurosciences *8*, 14–41. https://doi.org/10.20517/2347-8659.2020.12.
12. Ferraro, D., Trovati, A., Bedin, R., Natali, P., Franciotta, D., Santangelo, M., Camera, V., Vitetta, F., Varani, M., Trenti, T., et al. (2020). Cerebrospinal fluid kappa and lambda free light chains in oligoclonal band-negative patients with suspected multiple sclerosis. Eur. J. Neurol. *27*, 461–467. https://doi.org/10.1111/ene.14121.
13. Świderek-Matysiak, M., Oset, M., Domowicz, M., Galazka, G., Namiecińska, M., and Stasiołek, M. (2023). Cerebrospinal Fluid Biomarkers in Differential Diagnosis of Multiple Sclerosis and Systemic Inflammatory Diseases with Central Nervous System Involvement. Biomedicines *11*, 425. https://doi.org/10.3390/biomedicines11020425.
14. Teunissen, C.E., and Khalil, M. (2012). Neurofilaments as biomarkers in multiple sclerosis. Mult. Scler. *18*, 552–556. https://doi.org/10.1177/1352458512443092.
15. Kuhle, J., Plattner, K., Bestwick, J.P., Lindberg, R.L., Ramagopalan, S.V., Norgren, N., Nissim, A., Malaspina, A., Leppert, D., Giovannoni, G., et al. (2013). A comparative study of CSF neurofilament light and heavy chain protein in MS. Mult. Scler. *19*, 1597–1603. https://doi.org/10.1177/1352458513482374.
16. Åkesson, J., Hojjati, S., Hellberg, S., Raffetseder, J., Khademi, M., Rynkowski, R., Kockum, I., Altafini, C., Lubovac-Pilav, Z., Mellergård, J., et al. (2023). Proteomics reveal biomarkers for diagnosis, disease activity and long-term disability outcomes in multiple sclerosis. Nat. Commun. *14*, 6903. https://doi.org/10.1038/s41467-023-42682-9.
17. Elkjaer, M.L., Nawrocki, A., Kacprowski, T., Lassen, P., Simonsen, A.H., Marignier, R., Sejbaek, T., Nielsen, H.H., Wermuth, L., Rashid, A.Y., et al. (2021). CSF proteome in multiple sclerosis subtypes related to brain lesion transcriptomes. Sci. Rep. *11*, 4132. https://doi.org/10.1038/s41598-021-83591-5.
18. Huang, J., Khademi, M., Fugger, L., Lindhe, Ö., Novakova, L., Axelsson, M., Malmeström, C., Constantinescu, C., Lycke, J., Piehl, F., et al. (2020). Inflammation-related plasma and CSF biomarkers for multiple sclerosis. Proc. Natl. Acad. Sci. USA *117*, 12952–12960. https://doi.org/10.1073/pnas.1912839117.
19. Komori, M., Matsuyama, Y., Nirasawa, T., Thiele, H., Becker, M., Alexandrov, T., Saida, T., Tanaka, M., Matsuo, H., Tomimoto, H., et al. (2012). Proteomic pattern analysis discriminates among multiple sclerosis-related disorders. Ann. Neurol. *71*, 614–623. https://doi.org/10.1002/ana.22633.
20. Mosleth, E.F., Vedeler, C.A., Liland, K.H., McLeod, A., Bringeland, G.H., Kroondijk, L., Berven, F.S., Lysenko, A., Rawlings, C.J., Eid, K.E.-H., et al. (2021). Cerebrospinal fluid proteome shows disrupted neuronal development in multiple sclerosis. Sci. Rep. *11*, 4087. https://doi.org/10.1038/s41598-021-82388-w.
21. Opsahl, J.A., Vaudel, M., Guldbrandsen, A., Aasebø, E., Van Pesch, V., Franciotta, D., Myhr, K.-M., Barsnes, H., Berle, M., Torkildsen, Ø., et al. (2016). Label-free analysis of human cerebrospinal fluid addressing various normalization strategies and revealing protein groups affected by multiple sclerosis. Proteomics *16*, 1154–1165. https://doi.org/10.1002/pmic.201500284.
22. Probert, F., Yeo, T., Zhou, Y., Sealey, M., Arora, S., Palace, J., Claridge, T.D.W., Hillenbrand, R., Oechtering, J., Kuhle, J., et al. (2022). Determination of CSF GFAP, CCN5, and vWF Levels Enhances the Diagnostic Accuracy of Clinically Defined MS From Non-MS Patients With CSF Oligoclonal Bands. Front. Immunol. *12*, 811351. https://doi.org/10.3389/fimmu.2021.811351.
23. Ottervald, J., Franzén, B., Nilsson, K., Andersson, L.I., Khademi, M., Eriksson, B., Kjellström, S., Marko-Varga, G., Végvári, A., Harris, R.A., et al. (2010). Multiple sclerosis: Identification and clinical evaluation of novel CSF biomarkers. J. Proteom. *73*, 1117–1132. https://doi.org/10.1016/j.jprot.2010.01.004.

24. Held, F., Makarov, C., Gasperi, C., Flaskamp, M., Grummel, V., Berthele, A., and Hemmer, B. (2025). Proteomics Reveals Age as Major Modifier of Inflammatory CSF Signatures in Multiple Sclerosis. Neurol. Neuroimmunol. Neuroinflamm. *12*, e200322. https://doi.org/10.1212/NXI.0000000000200322.

25. Fissolo, N., Matute-Blanch, C., Osman, M., Costa, C., Pinteac, R., Miró, B., Sanchez, A., Brito, V., Dujmovic, I., Voortman, M., et al. (2021). CSF SERPINA3 Levels Are Elevated in Patients With Progressive MS. Neurol. Neuroimmunol. Neuroinflamm. *8*, e941. https://doi.org/10.1212/NXI.0000000000000941.

26. Kosa, P., Barbour, C., Varosanec, M., Wichman, A., Sandford, M., Greenwood, M., and Bielekova, B. (2022). Molecular models of multiple sclerosis severity identify heterogeneity of pathogenic mechanisms. Nat. Commun. *13*, 7670. https://doi.org/10.1038/s41467-022-35357-4.

27. Zelek, W.M., Fathalla, D., Morgan, A., Touchard, S., Loveless, S., Tallantyre, E., Robertson, N.P., and Morgan, B.P. (2020). Cerebrospinal fluid complement system biomarkers in demyelinating disease. Mult. Scler. *26*, 1929–1937. https://doi.org/10.1177/1352458519887905.

28. Martynova, E., Goyal, M., Johri, S., Kumar, V., Khaibullin, T., Rizvanov, A.A., Verma, S., Khaiboullina, S.F., and Baranwal, M. (2020). Serum and Cerebrospinal Fluid Cytokine Biomarkers for Diagnosis of Multiple Sclerosis. Mediat. Inflam. *2020*, 2727042. https://doi.org/10.1155/2020/2727042.

29. Gaetani, L., Bellomo, G., Di Sabatino, E., Sperandei, S., Mancini, A., Blennow, K., Zetterberg, H., Parnetti, L., and Di Filippo, M. (2023). The Immune Signature of CSF in Multiple Sclerosis with and without Oligoclonal Bands: A Machine Learning Approach to Proximity Extension Assay Analysis. Int. J. Mol. Sci. *25*, 139. https://doi.org/10.3390/ijms25010139.

30. Arrambide, G., Comabella, M., and Tur, C. (2024). Big data and artificial intelligence applied to blood and CSF fluid biomarkers in multiple sclerosis. Front. Immunol. *15*, 1459502. https://doi.org/10.3389/fimmu.2024.1459502.

31. Kodosaki, E., Watkins, W.J., Loveless, S., Kreft, K.L., Richards, A., Anderson, V., Hurler, L., Robertson, N.P., Zelek, W.M., and Tallantyre, E.C. (2024). Combination protein biomarkers predict multiple sclerosis diagnosis and outcomes. J. Neuroinflamm. *21*, 52. https://doi.org/10.1186/s12974-024-03036-4.

32. Cencioni, M.T., Magliozzi, R., Palmisano, I., Suwan, K., Mensi, A., Fuentes-Font, L., Villar, L.M., Fernández-Velasco, J.I., Migallón, N.V., Costa-Frossard, L., et al. (2024). Soluble CD27 is an intrathecal biomarker of T-cell-mediated lesion activity in multiple sclerosis. J. Neuroinflamm. *21*, 91. https://doi.org/10.1186/s12974-024-03077-9.

33. Sakiz, E., Amanzadeh Jajin, E., Cubeddu, L., Gamsjaeger, R., and Avsar, T. (2025). Comprehensive Meta-Analysis of Differentially Expressed Proteins in Cerebrospinal Fluid Associated with Multiple Sclerosis. Int. J. Mol. Sci. *26*, 6171. https://doi.org/10.3390/ijms26136171.

34. Qureshi, F., Hu, W., Loh, L., Patel, H., DeGuzman, M., Becich, M., Rubio da Costa, F., Gehman, V., Zhang, F., Foley, J., et al. (2023). Analytical validation of a multi-protein, serum-based assay for disease activity assessments in multiple sclerosis. Proteomics Clin. Appl. *17*, e2200018. https://doi.org/10.1002/prca.202200018.

35. Chitnis, T., Foley, J., Ionete, C., El Ayoubi, N.K., Saxena, S., Gaitan-Walsh, P., Lokhande, H., Paul, A., Saleh, F., Weiner, H., et al. (2023). Clinical validation of a multi-protein, serum-based assay for disease activity assessments in multiple sclerosis. Clin. Immunol. *253*, 109688. https://doi.org/10.1016/j.clim.2023.109688.

36. Bader, J.M., Geyer, P.E., Müller, J.B., Strauss, M.T., Koch, M., Leypoldt, F., Koertvelyessy, P., Bittner, D., Schipke, C.G., Incesoy, E.I., et al. (2020). Proteome profiling in cerebrospinal fluid reveals novel biomarkers of Alzheimer's disease. Mol. Syst. Biol. *16*, e9356. https://doi.org/10.15252/msb.20199356.

37. Zeng, W.-F., Zhou, X.-X., Willems, S., Ammar, C., Wahle, M., Bludau, I., Voytik, E., Strauss, M.T., and Mann, M. (2022). AlphaPeptDeep: a modular deep learning framework to predict peptide properties for proteomics. Nat. Commun. *13*, 7238. https://doi.org/10.1038/s41467-022-34904-3.

38. Karayel, O., Virreira Winter, S., Padmanabhan, S., Kuras, Y.I., Vu, D.T., Tuncali, I., Merchant, K., Wills, A.-M., Scherzer, C.R., and Mann, M. (2022). Proteome profiling of cerebrospinal fluid reveals biomarker candidates for Parkinson's disease. Cell Rep. Med. *3*, 100661. https://doi.org/10.1016/j.xcrm.2022.100661.

39. Mun, D.-G., Budhraja, R., Bhat, F.A., Zenka, R.M., Johnson, K.L., Moghekar, A., and Pandey, A. (2023). Four-dimensional proteomics analysis of human cerebrospinal fluid with trapped ion mobility spectrometry using PASEF. Proteomics *23*, e2200507. https://doi.org/10.1002/pmic.202200507.

40. Barkovits, K., Linden, A., Galozzi, S., Schilde, L., Pacharra, S., Mollenhauer, B., Stoepel, N., Steinbach, S., May, C., Uszkoreit, J., et al. (2018). Characterization of Cerebrospinal Fluid via Data-Independent Acquisition Mass Spectrometry. J. Proteome Res. *17*, 3418–3430. https://doi.org/10.1021/acs.jproteome.8b00308.

41. Niu, L., Geyer, P.E., Wewer Albrechtsen, N.J., Gluud, L.L., Santos, A., Doll, S., Treit, P.V., Holst, J.J., Knop, F.K., Vilsbøll, T., et al. (2019). Plasma proteome profiling discovers novel proteins associated with non-alcoholic fatty liver disease. Mol. Syst. Biol. *15*, e8793. https://doi.org/10.15252/msb.20188793.

42. Niu, L., Thiele, M., Geyer, P.E., Rasmussen, D.N., Webel, H.E., Santos, A., Gupta, R., Meier, F., Strauss, M., Kjaergaard, M., et al. (2022). Noninvasive proteomic biomarkers for alcohol-related liver disease. Nat. Med. *28*, 1277–1287. https://doi.org/10.1038/s41591-022-01850-y.

43. Geyer, P.E., Kulak, N.A., Pichler, G., Holdt, L.M., Teupser, D., and Mann, M. (2016). Plasma Proteome Profiling to Assess Human Health and Disease. Cell Syst. *2*, 185–195. https://doi.org/10.1016/j.cels.2016.02.015.

44. Geyer, P.E., Wewer Albrechtsen, N.J., Tyanova, S., Grassl, N., Iepsen, E.W., Lundgren, J., Madsbad, S., Holst, J.J., Torekov, S.S., and Mann, M. (2016). Proteomics reveals the effects of sustained weight loss on the human plasma proteome. Mol. Syst. Biol. *12*, 901. https://doi.org/10.15252/msb.20167357.

45. Geyer, P.E., Mann, S.P., Treit, P.V., and Mann, M. (2021). Plasma Proteomes Can Be Reidentifiable and Potentially Contain Personally Sensitive and Incidental Findings. Mol. Cell. Proteomics *20*, 100035. https://doi.org/10.1074/mcp.RA120.002359.

46. Messner, C.B., Demichev, V., Bloomfield, N., Yu, J.S.L., White, M., Kreidl, M., Egger, A.-S., Freiwald, A., Ivosev, G., Wasim, F., et al. (2021). Ultra-fast proteomics with Scanning SWATH. Nat. Biotechnol. *39*, 846–854. https://doi.org/10.1038/s41587-021-00860-4.

47. Szyrwiel, L., Gille, C., Mülleder, M., Demichev, V., and Ralser, M. (2024). Fast proteomics with dia-PASEF and analytical flow-rate chromatography. Proteomics *24*, e2300100. https://doi.org/10.1002/pmic.202300100.

48. Petrera, A., von Toerne, C., Behler, J., Huth, C., Thorand, B., Hilgendorff, A., and Hauck, S.M. (2021). Multiplatform Approach for Plasma Proteomics: Complementarity of Olink Proximity Extension Assay Technology to Mass Spectrometry-Based Protein Profiling. J. Proteome Res. *20*, 751–762. https://doi.org/10.1021/acs.jproteome.0c00641.

49. Ammar, C., Schessner, J.P., Willems, S., Michaelis, A.C., and Mann, M. (2023). Accurate Label-Free Quantification by directLFQ to Compare Unlimited Numbers of Proteomes. Mol. Cell. Proteomics *22*, 100581. https://doi.org/10.1016/j.mcpro.2023.100581.

50. Niu, L., Stinson, S.E., Holm, L.A., Lund, M.A.V., Fonvig, C.E., Cobuccio, L., Meisner, J., Juel, H.B., Thiele, M., Krag, A., et al. (2023). Plasma Proteome Variation and its Genetic Determinants in Children and Adolescents. Nat. Genet. *57*, 635–646. https://doi.org/10.1038/s41588-025-02089-2.

51. Aasebø, E., Opsahl, J.A., Bjørlykke, Y., Myhr, K.-M., Kroksveen, A.C., and Berven, F.S. (2014). Effects of blood contamination and the rostro-caudal gradient on the human cerebrospinal fluid proteome. PLoS One *9*, e90429. https://doi.org/10.1371/journal.pone.0090429.

52. Knox, E.G., Aburto, M.R., Clarke, G., Cryan, J.F., and O'Driscoll, C.M. (2022). The blood-brain barrier in aging and neurodegeneration. Mol. Psychiatry *27*, 2659–2673. https://doi.org/10.1038/s41380-022-01511-z.

53. Rieckmann, J.C., Geiger, R., Hornburg, D., Wolf, T., Kveler, K., Jarrossay, D., Sallusto, F., Shen-Orr, S.S., Lanzavecchia, A., Mann, M., et al. (2017). Social network architecture of human immune cells unveiled by quantitative proteomics. Nat. Immunol. *18*, 583–593. https://doi.org/10.1038/ni.3693.

54. Sharma, K., Schmitt, S., Bergner, C.G., Tyanova, S., Kannaiyan, N., Manrique-Hoyos, N., Kongi, K., Cantuti, L., Hanisch, U.-K., Philips, M.-A., et al. (2015). Cell type– and brain region–resolved mouse brain proteome. Nat. Neurosci. *18*, 1819–1831. https://doi.org/10.1038/nn.4160.

55. Reboul, J., Lambin, P., Gervais, A., Gaillard, O., Delattre, J., Bricaire, F., Lyon-Caen, O., and Schuller, E. (1995). Frequency and relevance of IgM intrathecal synthesis in multiple sclerosis. Eur. J. Neurol. *2*, 429–434. https://doi.org/10.1111/j.1468-1331.1995.tb00152.x.

56. Shaimardanova, A.A., Chulpanova, D.S., Solovyeva, V.V., Mullagulova, A.I., Kitaeva, K.V., Allegrucci, C., and Rizvanov, A.A. (2020). Metachromatic Leukodystrophy: Diagnosis, Modeling, and Treatment Approaches. Front. Med. (Lausanne) *7*, 576221. https://doi.org/10.3389/fmed.2020.576221.

57. Oh, J., Takahashi, R., Kondo, S., Mizoguchi, A., Adachi, E., Sasahara, R.M., Nishimura, S., Imamura, Y., Kitayama, H., Alexander, D.B., et al. (2001). The Membrane-Anchored MMP Inhibitor RECK Is a Key Regulator of Extracellular Matrix Integrity and Angiogenesis. Cell *107*, 789–800. https://doi.org/10.1016/S0092-8674(01)00597-9.

58. Oechtering, J., Schaedelin, S., Benkert, P., Müller, S., Achtnichts, L., Vehoff, J., Disanto, G., Findling, O., Fischer-Barnicol, B., Orleth, A., et al. (2021). Intrathecal Immunoglobulin M Synthesis is an Independent Biomarker for Higher Disease Activity and Severity in Multiple Sclerosis. Ann. Neurol. *90*, 477–489. https://doi.org/10.1002/ana.26137.

59. Kuhle, J., Barro, C., Disanto, G., Mathias, A., Soneson, C., Bonnier, G., Yaldizli, Ö., Regeniter, A., Derfuss, T., Canales, M., et al. (2016). Serum neurofilament light chain in early relapsing remitting MS is increased and correlates with CSF levels and with MRI measures of disease severity. Mult. Scler. *22*, 1550–1559. https://doi.org/10.1177/1352458515623365.

60. Gasperi, C., Salmen, A., Antony, G., Bayas, A., Heesen, C., Kümpfel, T., Linker, R.A., Paul, F., Stangel, M., Tackenberg, B., et al. (2019). Association of Intrathecal Immunoglobulin G Synthesis With Disability Worsening in Multiple Sclerosis. JAMA Neurol. *76*, 841–849. https://doi.org/10.1001/jamaneurol.2019.0905.

61. Fainardi, E., Castellazzi, M., Bellini, T., Manfrinato, M.C., Baldi, E., Casetta, I., Paolino, E., Granieri, E., and Dallocchio, F. (2006). Cerebrospinal fluid and serum levels and intrathecal production of active matrix metalloproteinase-9 (MMP-9) as markers of disease activity in patients with multiple sclerosis. Mult. Scler. *12*, 294–301. https://doi.org/10.1191/135248506ms1274oa.

62. Ambegaonkar, A.A., Holla, P., Dizon, B.L., Sohn, H., and Pierce, S.K. (2022). Atypical B cells in chronic infectious diseases and systemic autoimmunity: puzzles with many missing pieces. Curr. Opin. Immunol. *77*, 102227. https://doi.org/10.1016/j.coi.2022.102227.

63. Chorazy, M., Wawrusiewicz-Kurylonek, N., Adamska-Patruno, E., Czarnowska, A., Zajkowska, O., Kapica-Topczewska, K., Posmyk, R., Kretowski, A.J., Kochanowicz, J., and Kułakowska, A. (2021). Variants of Novel Immunomodulatory Fc Receptor Like 5 Gene Are Associated With Multiple Sclerosis Susceptibility in the Polish Population. Front. Neurol. *12*, 631134. https://doi.org/10.3389/fneur.2021.631134.

64. Franco, A., Kraus, Z., Li, H., Seibert, N., Dement-Brown, J., and Tolnay, M. (2018). CD21 and FCRL5 form a receptor complex with robust B-cell activating capacity. Int. Immunol. *30*, 569–578. https://doi.org/10.1093/intimm/dxy052.

65. Li, H., Dement-Brown, J., Liao, P.-J., Mazo, I., Mills, F., Kraus, Z., Fitzsimmons, S., and Tolnay, M. (2020). Fc receptor-like 4 and 5 define human atypical memory B cells. Int. Immunol. *32*, 755–770. https://doi.org/10.1093/intimm/dxaa053.

66. Sullivan, R.T., Kim, C.C., Fontana, M.F., Feeney, M.E., Jagannathan, P., Boyle, M.J., Drakeley, C.J., Ssewanyana, I., Nankya, F., Mayanja-Kizza, H., et al. (2015). FCRL5 Delineates Functionally Impaired Memory B Cells Associated with Plasmodium falciparum Exposure. PLoS Pathog. *11*, e1004894. https://doi.org/10.1371/journal.ppat.1004894.

67. Zurbuchen, Y., Michler, J., Taeschler, P., Adamo, S., Cervia, C., Raeber, M.E., Acar, I.E., Nilsson, J., Warnatz, K., Soyka, M.B., et al. (2023). Human memory B cells show plasticity and adopt multiple fates upon recall response to SARS-CoV-2. Nat. Immunol. *24*, 955–965. https://doi.org/10.1038/s41590-023-01497-y.

68. Lees, J.R., and Cross, A.H. (2007). A little stress is good: IFN-γ, demyelination, and multiple sclerosis. J. Clin. Investig. *117*, 297–299. https://doi.org/10.1172/JCI31254.

69. Lighvani, A.A., Frucht, D.M., Jankovic, D., Yamane, H., Aliberti, J., Hissong, B.D., Nguyen, B.V., Gadina, M., Sher, A., Paul, W.E., et al. (2001). T-bet is rapidly induced by interferon-γ in lymphoid and myeloid cells. Proc. Natl. Acad. Sci. USA *98*, 15137–15142. https://doi.org/10.1073/pnas.261570598.

70. Hoff, S.T., Salman, A.M., Ruhwald, M., Ravn, P., Brock, I., Elsheikh, N., Andersen, P., and Agger, E.M. (2015). Human B cells produce chemokine CXCL10 in the presence of Mycobacterium tuberculosis specific T cells. Tuberculosis (Edinb) *95*, 40–47. https://doi.org/10.1016/j.tube.2014.10.005.

71. Magliozzi, R., Howell, O.W., Calabrese, M., and Reynolds, R. (2023). Meningeal inflammation as a driver of cortical grey matter pathology and clinical progression in multiple sclerosis. Nat. Rev. Neurol. *19*, 461–476. https://doi.org/10.1038/s41582-023-00838-7.

72. Reali, C., Magliozzi, R., Roncaroli, F., Nicholas, R., Howell, O.W., and Reynolds, R. (2020). B cell rich meningeal inflammation associates with increased spinal cord pathology in multiple sclerosis. Brain Pathol. *30*, 779–793. https://doi.org/10.1111/bpa.12841.

73. Wiendl, H., Gold, R., Berger, T., Derfuss, T., Linker, R., Mäurer, M., Aktas, O., Baum, K., Berghoff, M., Bittner, S., et al. (2021). Multiple Sclerosis Therapy Consensus Group (MSTCG): position statement on disease-modifying therapies for multiple sclerosis (white paper). Ther. Adv. Neurol. Disord. *14*, 17562864211039648. https://doi.org/10.1177/17562864211039648.

74. Thompson, A.J., Baranzini, S.E., Geurts, J., Hemmer, B., and Ciccarelli, O. (2018). Multiple sclerosis. Lancet *391*, 1622–1636. https://doi.org/10.1016/S0140-6736(18)30481-1.

75. Lünemann, J.D., Jelcić, I., Roberts, S., Lutterotti, A., Tackenberg, B., Martin, R., and Münz, C. (2008). EBNA1-specific T cells from patients with multiple sclerosis cross react with myelin antigens and co-produce IFN-gamma and IL-2. J. Exp. Med. *205*, 1763–1773. https://doi.org/10.1084/jem.20072397.

76. Lanz, T.V., Brewer, R.C., Ho, P.P., Moon, J.-S., Jude, K.M., Fernandez, D., Fernandes, R.A., Gomez, A.M., Nadj, G.-S., Bartley, C.M., et al. (2022). Clonally expanded B cells in multiple sclerosis bind EBV EBNA1 and GlialCAM. Nature *603*, 321–327. https://doi.org/10.1038/s41586-022-04432-7.

77. Comabella, M., Hegen, H., Villar, L.M., Rejdak, K., Sao-Avilés, A., Behrens, M., Sastre-Garriga, J., Mongay, N., Berek, K., Martínez-Yelamos, S., et al. (2024). Increased EBNA1-specific antibody response in primary-progressive multiple sclerosis. J. Neurol. *272*, 26. https://doi.org/10.1007/s00415-024-12763-w.

78. Santiago, O., Gutierrez, J., Sorlozano, A., de Dios Luna, J., Villegas, E., and Fernandez, O. (2010). Relation between Epstein-Barr virus and multiple sclerosis: analytic study of scientific production. Eur. J. Clin. Microbiol. Infect. Dis. *29*, 857–866. https://doi.org/10.1007/s10096-010-0940-0.

79. Bjornevik, K., Cortese, M., Healy, B.C., Kuhle, J., Mina, M.J., Leng, Y., Elledge, S.J., Niebuhr, D.W., Scher, A.I., Munger, K.L., et al. (2022). Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated

with multiple sclerosis. Science *375*, 296–301. https://doi.org/10.1126/science.abj8222.

80. Bjornevik, K., Münz, C., Cohen, J.I., and Ascherio, A. (2023). Epstein–Barr virus as a leading cause of multiple sclerosis: mechanisms and implications. Nat. Rev. Neurol. *19*, 160–171. https://doi.org/10.1038/s41582-023-00775-5.

81. Misra, M.K., Damotte, V., and Hollenbach, J.A. (2018). The immunogenetics of neurological disease. Immunology *153*, 399–414. https://doi.org/10.1111/imm.12869.

82. Gasperi, C., Andlauer, T.F.M., Keating, A., Knier, B., Klein, A., Pernpeintner, V., Lichtner, P., Gold, R., Zipp, F., Then Bergh, F., et al. (2020). Genetic determinants of the humoral immune response in MS. Neurol. Neuroimmunol. Neuroinflamm. *7*, e827. https://doi.org/10.1212/NXI.0000000000000827.

83. Balnyte, R., Rastenyte, D., Vaitkus, A., Mickeviciene, D., Skrodeniene, E., Vitkauskiene, A., and Uloziene, I. (2013). The importance of HLA DRB1 gene allele to clinical features and disability in patients with multiple sclerosis in Lithuania. BMC Neurol. *13*, 77. https://doi.org/10.1186/1471-2377-13-77.

84. Dammer, E.B., Ping, L., Duong, D.M., Modeste, E.S., Seyfried, N.T., Lah, J.J., Levey, A.I., and Johnson, E.C.B. (2022). Multi-platform proteomic analysis of Alzheimer's disease cerebrospinal fluid and plasma reveals network biomarkers associated with proteostasis and the matrisome. Alzheimers Res. Ther. *14*, 174. https://doi.org/10.1186/s13195-022-01113-5.

85. Katz, D.H., Robbins, J.M., Deng, S., Tahir, U.A., Bick, A.G., Pampana, A., Yu, Z., Ngo, D., Benson, M.D., Chen, Z.-Z., et al. (2022). Proteomic profiling platforms head to head: Leveraging genetics and clinical traits to compare aptamer- and antibody-based methods. Sci. Adv. *8*, eabm5164. https://doi.org/10.1126/sciadv.abm5164.

86. Wesenhagen, K.E.J., Gobom, J., Bos, I., Vos, S.J.B., Martinez-Lage, P., Popp, J., Tsolaki, M., Vandenberghe, R., Freund-Levi, Y., Verhey, F., et al. (2022). Effects of age, amyloid, sex, and APOE ε4 on the CSF proteome in normal cognition. Alzheimers Dement. (Amst) *14*, e12286. https://doi.org/10.1002/dad2.12286.

87. Shi, Y., Ding, Y., Li, G., Wang, L., Osman, R.A., Sun, J., Qian, L., Zheng, G., and Zhang, G. (2021). Discovery of Novel Biomarkers for Diagnosing and Predicting the Progression of Multiple Sclerosis Using TMT-Based Quantitative Proteomics. Front. Immunol. *12*, 700031. https://doi.org/10.3389/fimmu.2021.700031.

88. Jankovska, E., Lipcseyova, D., Svrdlikova, M., Pavelcova, M., Kubala Havrdova, E., Holada, K., and Petrak, J. (2022). Quantitative proteomic analysis of cerebrospinal fluid of women newly diagnosed with multiple sclerosis. Int. J. Neurosci. *132*, 724–734. https://doi.org/10.1080/00207454.2020.1837801.

89. Kroksveen, A.C., Guldbrandsen, A., Vaudel, M., Lereim, R.R., Barsnes, H., Myhr, K.-M., Torkildsen, Ø., and Berven, F.S. (2017). In-Depth Cerebrospinal Fluid Quantitative Proteome and Deglycoproteome Analysis: Presenting a Comprehensive Picture of Pathways and Processes Affected by Multiple Sclerosis. J. Proteome Res. *16*, 179–194. https://doi.org/10.1021/acs.jproteome.6b00659.

90. Liu, H., Wang, Z., Li, H., Li, M., Han, B., Qi, Y., Wang, H., and Gao, J. (2022). Label-free Quantitative Proteomic Analysis of Cerebrospinal Fluid and Serum in Patients With Relapse-Remitting Multiple Sclerosis. Front. Genet. *13*, 635–646. https://doi.org/10.3389/fgene.2022.892491.

91. Zhu, M., Lan, Z., Park, J., Gong, S., Wang, Y., and Guo, F. (2024). Regulation of CNS pathology by Serpina3n/SERPINA3: The knowns and the puzzles. Neuropathol. Appl. Neurobiol. *50*, e12980. https://doi.org/10.1111/nan.12980.

92. Hagan, N., Kane, J.L., Grover, D., Woodworth, L., Madore, C., Saleh, J., Sancho, J., Liu, J., Li, Y., Proto, J., et al. (2020). CSF1R signaling is a regulator of pathogenesis in progressive MS. Cell Death Dis. *11*, 904. https://doi.org/10.1038/s41419-020-03084-7.

93. Hosoi, E., Nishizaki, C., Gallagher, K.L., Wyre, H.W., Matsuo, Y., and Sei, Y. (2001). Expression of the ryanodine receptor isoforms in immune cells. J. Immunol. *167*, 4887–4894. https://doi.org/10.4049/jimmunol.167.9.4887.

94. Osipchuk, N.C., Soulika, A.M., and Fomina, A.F. (2021). Modulation of Ryanodine Receptors Activity Alters the Course of Experimental Autoimmune Encephalomyelitis in Mice. Front. Physiol. *12*, 770820. https://doi.org/10.3389/fphys.2021.770820.

95. Tahmasebi, F., Barati, S., and Kashani, I.R. (2021). Effect of CSF1R inhibitor on glial cells population and remyelination in the cuprizone model. Neuropeptides *89*, 102179. https://doi.org/10.1016/j.npep.2021.102179.

96. Batiste, S.M., Blackwell, D.J., Kim, K., Kryshtal, D.O., Gomez-Hurtado, N., Rebbeck, R.T., Cornea, R.L., Johnston, J.N., and Knollmann, B.C. (2019). Unnatural verticilide enantiomer inhibits type 2 ryanodine receptor-mediated calcium leak and is antiarrhythmic. Proc. Natl. Acad. Sci. USA *116*, 4810–4815. https://doi.org/10.1073/pnas.1816685116.

97. Kryshtal, D.O., Blackwell, D.J., Egly, C.L., Smith, A.N., Batiste, S.M., Johnston, J.N., Laver, D.R., and Knollmann, B.C. (2021). RYR2 Channel Inhibition Is the Principal Mechanism of Flecainide Action in CPVT. Circ. Res. *128*, 321–331. https://doi.org/10.1161/CIRCRESAHA.120.316819.

98. Kumar, V., Donthireddy, L., Marvel, D., Condamine, T., Wang, F., Lavilla-Alonso, S., Hashimoto, A., Vonteddu, P., Behera, R., Goins, M.A., et al. (2017). Cancer-Associated Fibroblasts Neutralize the Anti-tumor Effect of CSF1 Receptor Blockade by Inducing PMN-MDSC Infiltration of Tumors. Cancer Cell *32*, 654–668.e5. https://doi.org/10.1016/j.ccell.2017.10.005.

99. Li, J., Chen, Z., Kim, G., Luo, J., Hori, S., and Wu, C. (2023). Cathepsin W restrains peripheral regulatory T cells for mucosal immune quiescence. Sci. Adv. *9*, eadf3924. https://doi.org/10.1126/sciadv.adf3924.

100. Watanabe, M., Nakamura, Y., Michalak, Z., Isobe, N., Barro, C., Leppert, D., Matsushita, T., Hayashi, F., Yamasaki, R., Kuhle, J., et al. (2019). Serum GFAP and neurofilament light as biomarkers of disease activity and disability in NMOSD. Neurology *93*, e1299–e1311. https://doi.org/10.1212/WNL.0000000000008160.

101. Paul, A., Comabella, M., and Gandhi, R. (2019). Biomarkers in Multiple Sclerosis. Cold Spring Harb. Perspect. Med. *9*, a029058. https://doi.org/10.1101/cshperspect.a029058.

102. Rajasekharan, S., and Bar-Or, A.; CIHR/MSSC NET in Clinical Autoimmunity (2012). From bench to MS bedside: Challenges translating biomarker discovery to clinical practice. J. Neuroimmunol. *248*, 66–72. https://doi.org/10.1016/j.jneuroim.2012.01.006.

103. Sastre-Garriga, J., Pareto, D., Battaglini, M., Rocca, M.A., Ciccarelli, O., Enzinger, C., Wuerfel, J., Sormani, M.P., Barkhof, F., Yousry, T.A., et al. (2020). MAGNIMS consensus recommendations on the use of brain and spinal cord atrophy measures in clinical practice. Nat. Rev. Neurol. *16*, 171–182. https://doi.org/10.1038/s41582-020-0314-x.

104. Scalfari, A., Neuhaus, A., Degenhardt, A., Rice, G.P., Muraro, P.A., Daumer, M., and Ebers, G.C. (2010). The natural history of multiple sclerosis: a geographically based study 10: relapses and long-term disability. Brain *133*, 1914–1929. https://doi.org/10.1093/brain/awq118.

105. Abdelhak, A., Hottenrott, T., Morenas-Rodríguez, E., Suárez-Calvet, M., Zettl, U.K., Haass, C., Meuth, S.G., Rauer, S., Otto, M., Tumani, H., et al. (2019). Glial Activation Markers in CSF and Serum From Patients With Primary Progressive Multiple Sclerosis: Potential of Serum GFAP as Disease Severity Marker? Front. Neurol. *10*, 280. https://doi.org/10.3389/fneur.2019.00280.

106. Bar-Or, A., Gelfand, J., Agbim, C., von Büdingen, H.-C., Cameron, B., Herman, A., Jia, X., Ramesh, A., Winger, R., Jiang, J., et al. (2025). Evaluation of CHIT1 and SERPINA3 as Candidate Cerebrospinal Fluid Biomarkers of Progressive Biology in Multiple Sclerosis (S33.003). Neurology *104*, 2542. https://doi.org/10.1212/WNL.0000000000210408.

107. Kong, A.T., Leprevost, F.V., Avtonomov, D.M., Mellacheruvu, D., and Nesvizhskii, A.I. (2017). MSFragger: ultrafast and comprehensive peptide

identification in mass spectrometry–based proteomics. Nat. Methods *14*, 513–520. https://doi.org/10.1038/nmeth.4256.

108. Demichev, V., Messner, C.B., Vernardis, S.I., Lilley, K.S., and Ralser, M. (2020). DIA-NN: Neural networks and interference correction enable deep proteome coverage in high throughput. Nat. Methods *17*, 41–44. https://doi.org/10.1038/s41592-019-0638-x.

109. Chen, T., and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD '16 (Association for Computing Machinery), pp. 785–794. https://doi.org/10.1145/2939672.2939785.

110. Lundberg, S.M., and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems NIPS'17 (Curran Associates Inc.), pp. 4768–4777.

111. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat. Methods *13*, 731–740. https://doi.org/10.1038/nmeth.3901.

112. Cox, J., and Mann, M. (2012). 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. BMC Bioinform. *13*, S12. https://doi.org/10.1186/1471-2105-13-S16-S12.

113. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47. https://doi.org/10.1093/nar/gkv007.

114. Polman, C.H., Reingold, S.C., Edan, G., Filippi, M., Hartung, H.-P., Kappos, L., Lublin, F.D., Metz, L.M., McFarland, H.F., O'Connor, P.W., et al. (2005). Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria.". Ann. Neurol. *58*, 840–846. https://doi.org/10.1002/ana.20703.

115. Polman, C.H., Reingold, S.C., Banwell, B., Clanet, M., Cohen, J.A., Filippi, M., Fujihara, K., Havrdova, E., Hutchinson, M., Kappos, L., et al. (2011). Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann. Neurol. *69*, 292–302. https://doi.org/10.1002/ana.22366.

116. Teunissen, C.E., Petzold, A., Bennett, J.L., Berven, F.S., Brundin, L., Comabella, M., Franciotta, D., Frederiksen, J.L., Fleming, J.O., Furlan, R., et al. (2009). A consensus protocol for the standardization of cerebrospinal fluid collection and biobanking. Neurology *73*, 1914–1922. https://doi.org/10.1212/WNL.0b013e3181c47cc2.

117. Reiber, H., Otto, M., Trendelenburg, C., and Wormek, A. (2001). Reporting Cerebrospinal Fluid Data: Knowledge Base and Interpretation Software. Clin. Chem. Lab. Med. *39*, 324–332. https://doi.org/10.1515/CCLM.2001.051.

118. Manouchehrinia, A., Westerlind, H., Kingwell, E., Zhu, F., Carruthers, R., Ramanujam, R., Ban, M., Glaser, A., Sawcer, S., Tremlett, H., et al. (2017). Age Related Multiple Sclerosis Severity Score: Disability ranked by age. Mult. Scler. *23*, 1938–1946. https://doi.org/10.1177/1352458517690618.

119. Kulak, N.A., Geyer, P.E., and Mann, M. (2017). Loss-less Nano-fractionator for High Sensitivity, High Coverage Proteomics. Mol. Cell. Proteomics *16*, 694–705. https://doi.org/10.1074/mcp.O116.065136.

120. Meier, F., Brunner, A.-D., Frank, M., Ha, A., Bludau, I., Voytik, E., Kaspar-Schoenefeld, S., Lubeck, M., Raether, O., Bache, N., et al. (2020). diaPASEF: parallel accumulation–serial fragmentation combined with data-independent acquisition. Nat. Methods *17*, 1229–1236. https://doi.org/10.1038/s41592-020-00998-0.

121. Skowronek, P., Thielert, M., Voytik, E., Tanzer, M.C., Hansen, F.M., Willems, S., Karayel, O., Brunner, A.-D., Meier, F., and Mann, M. (2022). Rapid and In-Depth Coverage of the (Phospho-)Proteome With Deep Libraries and Optimal Window Design for dia-PASEF. Mol. Cell. Proteomics *21*, 100279. https://doi.org/10.1016/j.mcpro.2022.100279.

122. Meier, F., Brunner, A.-D., Koch, S., Koch, H., Lubeck, M., Krause, M., Goedecke, N., Decker, J., Kosinski, T., Park, M.A., et al. (2018). Online Parallel Accumulation–Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. Mol. Cell. Proteomics *17*, 2534–2545. https://doi.org/10.1074/mcp.TIR118.000900.

123. Wallmann, G., Skowronek, P., Brennsteiner, V., Lebedev, M., Thielert, M., Steigerwald, S., Kotb, M., Heymann, T., Zhou, X.-X., Schwörer, M., et al. (2025). AlphaDIA enables End-to-End Transfer Learning for Feature-Free Proteomics. Nat. Biotechnol. https://doi.org/10.1038/s41587-025-02791-w.

124. Pernemalm, M., Sandberg, A., Zhu, Y., Boekel, J., Tamburro, D., Schwenk, J.M., Björk, A., Wahren-Herlenius, M., Åmark, H., Östenson, C.G., et al. (2019). In-depth human plasma proteome analysis captures tissue proteins and transfer of protein variants across the placenta. Elife *8*, e41608. https://doi.org/10.7554/eLife.41608.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human CSF	This study	N/A
Chemicals, peptides, and recombinant proteins		
lysyl Endopeptidase®, mass spectrometry grade (Lys-C)	FUJIFILM Wako	Cat# WAKO129-02541
trypsin, porcine, mass spectrometry grade	Sigma Aldrich /Merck	Cat# T6567
styroldivinylbenzol-reversed phase sulfonate (SDB-RPS)	Empore 3M	Cat# 13427427
Isotope labeled heavy spike-in peptides, custom synthesized in manufacturer grade 'SpikeMix L' and 'MaxiSpikeTides L' for a few peptides longer than 20 amino acids. Overview table on peptide sequences included in PRIDE repository for targeted data, see below	JPT Technologies, Berlin, Germany	N/A
N Albumin-Standard	Siemens	N/A
N IgG-Standard	Siemens	N/A
Trichloroacetic acid for total protein measurement	Siemens	N/A
Deposited data		
Raw proteomic data measured on the timsTOF Pro 2 or Orbitrap platforms (Q Exactive HF-X, Exploris 480).	This paper, access via PRIDE repository	PRIDE: PXD045058
Raw proteomic data measured on the Astral platform	This paper, access via PRIDE repository	PRIDE: PXD064570
Raw targeted proteomic data measured on the Stellar platform	This paper, access via PRIDE repository	PRIDE: PXD064513
Processed proteomic data, key R and Python code to analyze the data, and key analysis results generated in this study	this paper, access via OSF platform	OSF: https://doi.org/10.17605/OSF.IO/JM2AK
Brain cell type proteomes, specifically protein expression data of cultured CNS cell types for individual replicates and developmental stage	Sharma et al.[54]	Table S5; https://static-content.springer.com/esm/art%3A10.1038%2Fnn.4160/MediaObjects/41593_2015_BFnn4160_MOESM42_ESM.xlsx
Proteomes of human immune cells in blood, specifically the total proteomes files	Rieckmann et al.[53]	Table S3; https://static-content.springer.com/esm/art%3A10.1038%2Fni.3693/MediaObjects/41590_2017_BFni3693_MOESM10_ESM.xlsx
OLINK CSF proteome dataset, comparing MS to healthy controls	Åkesson et al.[16]	Table S1; https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-023-42682-9/MediaObjects/41467_2023_42682_MOESM4_ESM.xlsx
OLINK CSF proteome dataset on variability explained by QAlb	Held et al.[24], access via OSF repository	osf.io/n6hx5/?view_only=05edc0dc8bec405686cca8d1d7669811
Software and algorithms		
Fragpipe suite including MSFragger, mass spectrometry search engine for library generation	Kong et al.[107]	https://fragpipe.nesvilab.org/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
DIA-NN version 1.8.1, mass spectrometry search engine for processing of CSF samples	Demichev et al.[108]	https://github.com/vdemichev/DiaNN
directLFQ, normalization algorithm for MS data	Ammar et al.[49]	https://github.com/MannLabs/directlfq
AlphaPeptDeep, spectral library prediction	Zeng et al.[37]	https://github.com/MannLabs/alphapeptdeep
Skyline daily suite to analyze targeted MS data, version 24.1.1.398	MacCoss Lab	https://skyline.ms/project/home/software/Skyline/begin.view
PRM conductor	Philipp Remes (Thermo Fisher Scientific)	https://skyline.ms/skyts/home/software/Skyline/tools/details.view?name=PRM%20Conductor
XGBoost, machine learning classifier	Chen and Guestrin[109]	https://xgboost.readthedocs.io/en/stable/
SHAP, package for Shapley value extraction	Lundberg and Lee[110]	https://shap.readthedocs.io/en/latest/
Perseus computational platform	Tyanova et al.[111]	https://maxquant.net/perseus/
1D annotation enrichment	Cox and Mann[112]	NA
Python version 3.10.09	Python Software Foundation	https://www.python.org/
Protis, protein testing	Siemens	https://www.siemens-healthineers.com/diagnostics-it/decision-support/protis-system
R version 4.3.0	R Core Team	https://www.R-project.org/
Limma for differential expression analysis, version 3.50.0	Ritchie et al.[113]	https://bioconductor.org/packages/release/bioc/html/limma.html

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Study collective – Discovery cohort

We first measured CSF proteomes in a pilot study comprising cross-sectional samples from 482 individuals across two cohorts. More information about this pilot study can be found on the PRIDE archive. The main study included cross-sectional samples from 5,059 individuals in one cohort and used a different experimental setup but still included more than 95% of the samples of the pilot study. From here, we will focus on describing the main study. It comprised cross-sectional CSF samples from 5059 donors. For data analysis, we excluded 14 samples with less than 500 identified proteins, reducing this to 5045 samples. In that set, there were 1343 samples from donors with multiple sclerosis (MS), 641 with other neurological autoimmune diseases, 391 with neurodegenerative diseases, 300 with infectious diseases, 241 with neoplasia, 178 with stroke, 599 neurological controls, and 1352 with other CNS diseases. The predominant subtype of MS was RRMS, with 1190 samples, followed by PPMS and SPMS, with 93 and 54 samples, respectively. The RRMS group was predominantly treatment-naïve and included RRMS and CIS samples due to changes in the diagnostic criteria during the time of sample collection. Stroke samples were predominantly of ischemic type, with 164 samples; 14 hemorrhagic stroke samples were excluded from biological analyses. The group of other autoimmune diseases was further categorized into three groups, namely demyelinating events not fulfilling MS diagnostic criteria, such as myelitides (referred to as "DM AIM"), other autoimmune diseases of the central nervous system not related to MS, including Neuromyelitis optica spectrum disorders ("CNS AIM"), and other autoimmune diseases of the peripheral nervous system, for example polyneuropathies ("PNS AIM"). The DM AIM group included 290 samples, CNS AIM 155 samples, and PNS AIM 196 samples.

Informed consent was obtained from all individuals, and the local ethics committee approved the study (S324/2019). The diagnoses were extracted from the hospital records and were made according to standard guidelines and diagnostic criteria. The control group consists of individuals who received a diagnostic workup after presenting with headache with 1) no evidence of a structural neurological disease or 2) a diagnosis of idiopathic intracranial hypertension. Individuals with MS were diagnosed according to the McDonald criteria valid at the time.[74,114,115]

A subset of 1595 individuals were selected as stratified subset for re-measurement by the Astral with the intention of finding biomarkers separating MS from other inflammatory diseases. Stratification aimed to reduce covariate differences and retain only samples with the clearest and most relevant diagnoses to the above aim. This selection of the most relevant autoimmune and inflammatory diseases in the context of an MS diagnosis was performed by clinical specialists. Additionally, samples were only included if they had none or little blood contamination (0 or (+)). For MS samples, this was stricter and required no evidence of blood contamination.

Further, MS samples were matched to the group of individuals with demyelinating events not fulfilling MS diagnostic criteria in a 4:1 ratio based on age, sex, difference to age adjusted Qalb, IgG ratio, OCB presence and leukocyte count. The resulting stratified subset comprised 195 individuals with infectious diseases, 755 individuals with MS, 373 with other autoimmune diseases and 272 neurological controls. Amongst autoimmune diseases, the DM AIM subgroup included 182 individuals, the CNS AIM group 96, and the PNS AIM group 95. More information about samples is in Table S1.

Study collective – Replication cohort

The replication cohort for the follow-up measurements consisted of 616 samples, including samples from 220 individuals with MS, 37 individuals with infectious diseases, 155 neurological controls, and 204 individuals with other autoimmune diseases. The group of individuals with other autoimmune diseases was again separated into demyelinating events not fulfilling MS diagnostic criteria with 109 samples, other autoimmune diseases of the central nervous system not related to MS with 36 samples, and other autoimmune diseases of the peripheral nervous system with 59 samples. Inclusion required available information on Qalb, blood contamination and OCB status. This cohort includes 351 samples that were also included in full discovery cohort measured on the timsTOF platform, but not in the stratified subset of the discovery cohort, meaning that they were first measured on the Astral platform in the replication cohort. As we collected samples for the replication cohort, few OCB-negative MS samples were available, reflecting their low natural prevalence. To reach sufficient numbers, we reduced the stratified discovery subset from 139 to 100 OCB-neg. MS samples, transferring 39 randomly chosen samples to the replication cohort. For these 39 samples, Astral measurements in both replication and stratified subset of the discovery cohort exist. We ensured that these samples were excluded from the discovery dataset in our analyses.

METHOD DETAILS

Clinical sampling, measurements, and quality control

All samples were obtained and processed in the department of Neurology at TUM. After lumbar puncture, samples were immediately transferred to the laboratory, stored on ice, and processed within one hour maximum in the standardized process described in this section, which was guided by community guidelines.[116] Clinical and proteomic measurements were conducted in a blinded fashion.

For cell count determination, 3µl CSF stored in polypropylene tubes were pipetted into a Rosenthal counting chamber, and leukocytes and erythrocytes were counted manually under the microscope (Zeiss, AXIO Lab.A1) and were visually assessed. Erythrocyte count was categorized into groups 0 to +++, depending on the number of erythrocytes observed, marking it as ''bloody'' when there are few erythrocytes in the sample (group 0 or +), but the color was nevertheless visibly changed. For measuring protein in CSF and blood serum, whole blood samples were centrifuged at 3500 rpm for 10 minutes and the serum was decanted. Samples were then analyzed by nephelometry using a ProSpec instrument from Siemens with specific reagents: N Albumin-Standard and N IgG-Standard for albumin and IgG, respectively, and trichloroacetic acid for total protein analysis. Data are processed with Protis (Siemens), and ratios between serum and CSF were automatically computed. Total protein measurement in CSF was discontinued in 2019 because certified control sera were not available any longer, resulting in missing information for ca. 20% or the samples. The quality of the measurements was controlled by daily test measurements of Liquicheck Spinal Fluid Control 1 + 2 (neolab). These measurements were later replaced by an Optilite system (Binding Site), in use for approximately 45 samples of this cohort. The amount of oligoclonal bands in CSF and serum were indicated by 0, (+) or +, and an optional free text field. These indicators were not present for all samples. CSF-specific oligoclonal bands were classified as ''positive'' if the indicator for serum was 0 and + for CSF, or serum was (+) and CSF +, or both + with at least 2 CSF-specific bands specified in the free text field. They were classified as ''borderline'', if serum was 0 and CSF (+), both (+), or serum was (+) and CSF +, or both + with less than 2 CSF-specific bands specified in the free text field. In all other cases with available information, the CSF-specific OCBs were classified as negative, or NA for unavailable data. For all samples included in this study at least 200µl remained.

After the above quality controls, which took 15 minutes, CSF cells were removed by dual centrifugation and CSF was frozen for long-term storage. The first centrifugation was performed at 400g for 10 minutes, and cellular material was stored separately. The CSF was again centrifuged at 1200g for 10 minutes and the supernatant was frozen at -80°C. The ''difference to age-adjusted QAlb'' is computed according to the Reiber formula by subtracting the measured QAlb value from the maximum QAlb value defined as healthy depending on age (QAlb_max = 4+age/15).[117] EDSS scores were converted to global ARMSS scores as described by Manouchehrinia et al.[118]

Dataset overview

We first measured a pilot cohort of about 500 samples on quadrupole-Orbitrap mass spectrometers (Q Exactive HF-X, Exploris 480) in a workflow analogous to our previous Alzheimer's Disease (AD) study.[36] These samples were again fully contained in our main discovery cohort comprising more than 5000 samples described above, which were measured on the timsTOF Pro 2 mass spectrometer. We did not analyze pilot cohort proteomes in this manuscript, but the pilot cohort spectra refined the library used to process the timsTOF data. Later in our study, we built an improved proteomic workflow around the Orbitrap Astral. With this Astral workflow we re-measured about 2000 samples of the discovery cohort which we had prepared for timsTOF measurement before, and 600 new samples of an additional replication cohort. These discovery samples comprised the about 1600-strong stratified subset to identify

biomarkers to separate MS from other inflammatory CNS conditions, and additional samples to study MS disease course differences. Lastly, we built a targeted assay for 29 of the Astral data-derived MS biomarkers on the Stellar mass spectrometer which we applied to again measure the stratified discovery subset and the replication cohort.

Sample organization and pooled samples for quality control

Regarding the main discovery cohort, blocking randomization was used to assign cross-sectional samples to 56 96-well plates. Sample plates were stratified for age, sex, and MS status (MS cases or controls). Additionally, we included samples from a pool of MS CSF as well as samples from a pool of non-MS CSF for quality control purposes and to calculate technical variability. Two pool samples of each type were included on each of the 56 plates, reflecting inter-plate preparation. Positions A1 and D6 contained pooled MS CSF, and positions A2 and D7 contained pooled non-MS CSF. Additionally, 16 pool samples of each type were placed on plate 57, reflecting intra-plate sample preparation. Both of these two preparation types were combined with measurement along cohort samples of the study or in blocks of pool samples only, to integrate or exclude, respectively, the variability of the analytical system across the entire cohort. For that purpose, aliquots of pool samples after sample preparation were transferred from the 56 cohort plates to three separate pool plates ('POL') for joint Evotip loading and measurement in block. The replication cohort was prepared later and contained freshly prepared but similar pool samples for MS and non-MS. Within the replication cohort, blocking randomization across was again used as before.

Sample preparation

Regarding the pilot cohort, sample preparation followed our AD study protocol. Regarding the main discovery and replication cohort, we optimized the sample preparation for high-throughput processing and improved protein depth at short gradients. CSF was gently thawed on ice and 40µl aliquoted to 96-well plates on ice. For each sample plate, denaturing buffer was freshly prepared by combining 4ml of pure water, 0.5ml of 1M Tris/HCl pH8, 0.4ml of 0.55M chloroacetamide, and 0.1ml of 0.5M TCEP solution. To each sample, 40µl of denaturing buffer was added. The chloroacetamide solution was prepared at the beginning of the study, split into 0.4ml aliquots, and stored at -20°C. The plate was sealed, shaken at 2000rpm on a shaking incubator, and spun at 500g for 1 minute to settle the liquid. The samples were denatured at 87°C at 300rpm on a shaking incubator. The temperature was not raised to ensure that no lid popped open due to pressurization. After cooling for 5 minutes, the samples were again spun to settle the condensate and then frozen at -20°C until further processing.

Further liquid handling was automated on the Agilent Bravo platform with custom protocols for digestion, peptide clean-up by StageTips, and Evotip loading. Processing was done in batches of eight sample plates. Samples were thawed at room temperature, and a 20µl solution containing 0.16µg of both trypsin and LysC each was added. An enzyme solution for all sample plates was prepared at the beginning of the study and aliquoted, this solution was freshly diluted and dispensed to a 96well plate for robot handling. Sample plates with enzyme were mixed at 2000rpm for 1min on a shaking incubator and then digested at 37°C for 18 hours over night. The next day, the samples were acidified by the addition of 20µl 50% (v/v) trifluoroacetic acid and mixed by pipetting up and down by the robot. The samples were transferred to StageTips for clean-up, removing lipids, salts, and potential particles. StageTips were filled with three 14-gauge plugs of SDB-RPS resin. Packing was done by one person to minimize variability. StageTips were centrifuged for 15min at 1700g and visually checked for complete flow through of the samples. In the rare case of StageTip clogging, a thin gel loader pipette tip was used to remove potential particles off the resin disc. The StageTip discs were then washed with 100µl of 1% (v/v) trifluoroacetic acid in isopropanol. Complete flow-through and dryness of the disc were visually checked before the addition of 50µl elution solution containing 50% (v/v) acetonitrile and 1% ammonium hydroxide. Peptides were eluted directly into 96well plates by centrifugation at 300g for 3 minutes and 1000g for 2 minutes. Eluates were dried in a vacuum concentrator for 90 minutes at 60°C and resuspended in 70µl of 2% (v/v) acetonitrile, 2% acetic acid, and 0.5% trifluoroacetic acid. The samples were then frozen at -20°C and stored until loading onto Evotips shortly before measurement.

Loading of Evotips with sample peptides for measurement

For Evotip loading, samples were thawed and mixed in a shaking incubator at 2000rpm for one minute. Evotips were soaked in 1-propanol poured into an Evotip box for about 3 minutes until the liquid was 1-2mm above the C18 disc. Evotips were then transferred to 3D-printed custom trays that fit into the Agilent robot spaces for robot handling. The soaked 1-propanol was removed by the addition of 50µl 0.1% (v/v) formic acid (EvoA) and centrifugation for 1 minute at 700g. To prevent drying of the EvoTip, 15µl of EvoA was added without delay, and the Evotips spun quickly up to 300g to settle the liquid on top of the discs. Then 50µl of EvoA was added, which stayed higher up in the tip absent centrifugation. Into that liquid, the sample peptides were pipetted and mixed by the robot. For measurement on the timsTOF Pro 2, 4µl corresponding to about 400ng of peptides were loaded. For Astral measurements, this was reduced to 2µl. For targeted measurements on the Stellar, first 20µl containing 20fm of each isotope-labeled heavy peptide in 1% (v/v) acetonitrile were added, and then 10µl of CSF peptides. For a separate CSF loading experiment, a reduced amount of CSF peptides of only 7 and 4µl were additionally loaded for two sample plates of the replication cohort. After peptide loading, Evotips were centrifuged at 700g for 1min to bind the peptides to the C18 disc, washed with 50µl EvoA as before, filled with 100µl of EvoA, and transferred back to an Evotip box, which was also filled with EvoA. The loaded Evotip boxes were stored

at 4°C until LC-MS/MS analysis, typically no longer than one week for discovery proteomics. For targeted proteomics on the Stellar, this was reduced to a maximum of three days through more frequent sample loading and 1mM of dithiothreitol was added to EvoA to counteract potential peptide oxidation.

Preparation of samples for library generation

For processing of pilot cohort data, we generated 24 high pH fractions from digested MS CSF using an automated high-pH fractionator, similar to our AD study.[36,119]

For the main study, we extended the experimental library to a total of four fraction sets, collecting initially 48 and subsequently 96 fractions of tryptic digests of each depleted and undepleted CSF. To add to total identified peptides, we still included the 48 fraction sets. For depletion, we used an equal volume of CSF and a 50% slurry of High-Select Top14 plasma depletion resin (Pierce/ThermoFisher). (Note that there is no commercial CSF-specific depletion kit to remove highly abundant plasma proteins in CSF.) Depleted and undepleted CSF were used to obtain a deep library while keeping depleted proteins in the overall library. Fractionation required high starting amounts of peptides. To avoid overloading the SDB-RPS resin during clean-up, the digested CSF was distributed onto multiple StageTips, loading no more than 10μg per tip. The eluted and vacuum concentrator-dried desalted peptides were resuspended, pooled, and their concentration was measured by absorbance at 280nm using a spectrophotometer (Nanodrop 2000, Thermo Fisher). About 50-100μg of peptide in 45μl were used for high pH fractionation. While less comprehensive than 96 fractions, we included the sets of 48 fractions in the measurement as we expected it to add identifications to the overall experimental library due to the stochasticity of data-dependent acquisition during measurement. After fractionation, we measured the peptide concentration in each fraction by spectrophotometry as above and used about 50-80ng for LC-MS/MS analysis. The Astral library was created computationally and did not require CSF sample measurements.

Mass spectrometry for discovery proteomics

Orbitrap setup measurements were done on two related instruments: A first partition of the pilot cohort was measured on a Q Exactive HF-X, equivalent to our prior AD study.[36] The second partition was measured on the successor instrument, the Exploris 480, with equivalent settings. For chromatography, we used a home-packed 50 cm column (inner diameter of 75μm, packed with 1.9 A Reprosil AQ Beads, Dr. Maisch), and the EASY-nLC 1200 system (Thermo Fisher). The data-independent acquisition (DIA) scheme to measure cohort samples included one MS1 survey scan followed by 33 MS2 scans. The high pH CSF fractions for the library were measured in data-dependent acquisition (DDA) mode, selecting the top 12 most abundant precursors for fragmentation. We used a 100-minute gradient, corresponding to a throughput of 12 samples per day (SPD) when accounting for sample loading overheads.

The timsTOF Pro 2 (Bruker Daltonics) was coupled to an Evosep One chromatography system (Evosep). Peptides were separated by a 21-minute gradient with a flow rate of 1 μl/min during peptide elution and a 2 μl/min flow at the start and end of the gradient. This gradient corresponds to a throughput of 60 samples per day. A commercial 8 cm x 150 um ID column packed with 1.5 um C18 functionalized beads (PepSep) was used at 40°C, connected to a 10 μm ID fused silica emitter (ZDV Captive Spray Emitter), which was housed in the captive spray ion source (all Bruker Daltonics) interfacing liquid chromatography and mass spectrometry. The mass spectrometer was operated in dia-PASEF mode[120] with variable window width for measurement of the CSF study samples. The windows were arranged with the py_diAID tool to cover the precursor cloud efficiently in the m/z – ion mobility (IM) plane.[121] For that purpose, the precursor density distribution in m/z and IM was modeled based on a tryptic high pH fractionation library of pooled CSF acquired by the same gradient. From previous experience with the 'high-speed' dia-PASEF method, we used the constraints of m/z 400 to 1200.[120] Additionally, we found that constraining the ion mobility to 0.8 - 1.3 Vs cm-2 $1/K_0$ was optimal, leaving out some very short peptides. The optimized dia-PASEF method comprised one MS1 scan followed by ten dia-PASEF scans with two IM ramps, corresponding to 20 MS2 windows (Table S1). The accumulation and ramp time were 100 ms, resulting in an overall cycle time of 1.16 seconds. The collision energy was a linear ramp from 20 eV at 1/K0 = 0.6 Vs cm-2 to 59 eV at 1/K0 = 1.6 Vs cm-2. The instrument was controlled by HyStar (6.0.30.0) and timsControl (3.0.20.0) data acquisition software (all Bruker Daltonics). To measure high pH fractions for the library, the MS was operated in dda-PASEF mode with one MS1 survey TIMS-MS and 4 PASEF MS/MS scans per acquisition cycle.[122] All other settings were the same as in our dia-PASEF method, including the ranges of IM and m/z. Generally, the measurement order on a sample plate was from left to right and subsequently from top to bottom rows. Plates were measured mainly according to plate order. When analytical performance fell below a threshold or an issue occurred, measurement of the entire plate was repeated later on, as this required loading the samples on Evotips again. The measurement order can be found in the supplement.

The Astral mass spectrometer (Thermo) was also coupled to the EvoSep One. However, we used a 11.5-minute gradient, corresponding to a throughput of 100 samples per day, and an 8cm (ion opticks) column. A FAIMS device was mounted to the mass spectrometer and operated at a compensation voltage of -40V, limiting entry of contaminants and reducing the need for cleaning. The ion source was operated at 1900V, the ion transfer tube at 280°C. Survey (MS1) scans were acquired by the Orbitrap at a resolution of 120000, covering a range of 380-980 m/z. This range was split into 150 isolation windows of 4 m/z for data-independent acquisition by the Astral mass analyzer. The normalized AGC target was set to 500%. The maximum injection time per isolation window was limited to 7ms. Isolated ions were fragmented in the HCD cell at a normalized collision energy of 25. Fragment spectra (MS2) were recorded for the 150-2000 m/z range.

Library generation for discovery proteomics

Separate spectral libraries were generated for each analytical platform. Specifically, this included the Orbitrap (Q Exactive HF-X/Exploris 480) setup with a 100-minute gradient, the timsTOF Pro 2 setup with a 21-minute gradient, and the Orbitrap Astral setup with a 11.5-minute gradient. For the first two setups, we generated libraries through a combination of experimentally acquisition and computationally prediction. The Astral library was entirely computationally predicted.

Regarding experimental libraries, data-dependent acquisition (DDA) raw files of high pH-fractionated CSF peptides were searched by MSFragger (software version 3.5) in the Fragpipe suite (18.0) using Philospher (4.2.2), EasyPQP (0.1.29) and Python (3.9).[107] The proteome fasta file was downloaded from UniProt in March 2022 and generally comprised SwissProt entries as well as reviewed isoforms. For analysis of DDA raw files of high pH fractions, UniProt TrEmbl entries were also considered. Additionally, common contaminants and enzymes for digestion preloaded in Fragpipe were included in the search space but later removed before bioinformatic analysis. Fragpipe generated an equally sized sequence-reversed decoy database for FDR scoring. Default search parameters were used unless stated otherwise. In brief, the peptide length was set to 7 to 45 amino acids, and the peptide mass was constrained to 500 to 5,000. The protein digest was set to trypsin, and one miscleavage, as well as the variable modifications N-terminal acetylation and methionine oxidation, were allowed. Carbamidomethlyation was set as a fixed modification. For the Orbitrap setup, the 24 high pH fraction raw files of multiple sclerosis CSF were complemented with two sets of 24 raw files each of AD and control CSF from our prior AD study which had an equivalent analytical setup. Regarding the timsTOF setup, the four sets of fractions (48 or 96) combined with optional depletion totaled 288 raw files for library generation.

For computational library generation, spectra were predicted by the AlphaPeptDeep software for the entire human UniProt proteome and the top 16 most abundant fragments were retained.[37] Within the same analytical setup, experimental libraries were used as training data to fine tune the prediction of spectra, retention time, and ion mobilities. To limit the already large size of the predicted library, we constrained the prediction of peptides to the dominant precursor properties in our experimental libraries. Specifically, peptide length was limited to seven to 35 amino acids, peptide charge to two to four charges. Likewise, a maximum of one missed cleavage, and a maximum of one methionine oxidation per peptide was permitted. Ions were removed from the predicted library if they were already contained in the experimental library.

The quality of the prediction was assessed by comparing predicted and experimental values for the ions with experimental observations in a train-test-split setting. Regarding fragmentation spectra, more than 90% and 63% of precursors in the test set of the Orbitrap and timsTOF setup, respectively, had a Pearson correlation coefficient of at least 90% for the correlation of predicted and experimental fragment intensities. Likewise, predictions of retention times and collisional cross-sections were close to true values, with regressions between experimental and predicted yielding R^2 values between 0.98 and 0.99.

Both the experimental and predicted libraries contained precursors and fragment ions not observable in donor samples which were unfractionated and undepleted. The large library size, in particular due to the prediction, slowed the search, making the processing of large cohorts time-consuming. Thus, a subset of the timsTOF discovery cohort comprising the first five 96-well plates was searched against the combined experimental and predicted library using the DIA-NN software version 1.8.1, thereby generating a pruned library of observable ions. The same approach was applied to the Orbitrap setup, however, using all pilot cohort samples, which likewise numbered around 500. The libraries of observed ions for these two analytical setups, Orbitrap and timsTOF, overlapped but also had unique ions. This may reflect a mixed effect of distinct separations of peptides between the analytical setups as well as stochasticity in DDA sampling, which can only be partially compensated by prediction due to FDR constraints associated with large predicted libraries. Thus, we computationally merged the two observed libraries to yield a final combined library. For that purpose, we used AlphaPeptDeep's transfer learning functionalities to predict ion mobilities for Orbitrap setup-specific ions. The ion mobilities of the timsTOF setup library were used as training data. The retention times (RT) were aligned into a uniform normalized RT space by 'Locally Estimated Scatterplot Smoothing' (LOESS) regression. DIA-NN can handle this abstract RT space in the library by aligning it to the gradient encountered in raw files based on a temporary first search.

The Astral library was generated by AlphaPeptDeep.[37] Fragment spectra were predicted by end-to-end transfer learning, as recently described for the combination with AlphaDIA.[123] The training data consisted of six replicates of a HeLa cell lysate tryptic digest measured on the Astral for routine quality control of instrument performance with a 21min gradient. These data were split into 70%, 20%, and 10% for training, validation, and test set because early stopping and dynamic learning rate scheduling was used. Predictions correlated strongly ($R^2 \sim 0.98$) with experimental values. This yielded a generic Astral-tuned fragment prediction model for multi-purpose use in the laboratory, e.g. even across species. For this study, the model was used to predict the top 16 most abundant fragments of the human UniProt proteome comprising canonical and isoform SwissProt entries.

To provide refined library retention times that matched those in the CSF cohort samples with a 11.5-minute gradient, transfer learning was again employed. Training data could be derived from the CSF cohort sample raw files themselves, removing the need to measure any samples specifically for this library. To wit, the cohort samples were searched by DIA-NN using our above merged CSF library with normalized RT space. The results from this search contained about 20000 precursors with RTs in the desired 11.5-minute gradient. The RT was normalized to range zero to one while preserving the peptide order, and was split into 80% training and 20% testing data. AlphaPeptDeep's retrained RT model yielded high correlation ($R^2 > 0.99$) in the test set, and was applied to the human proteome.

Processing of discovery proteomics raw data

DIA raw files of CSF cohort samples were processed with DIA-NN version 1.8.1.[108] Orbitrap and timsTOF raw files were separately searched against the merged library using the GUI version of DIA-NN. Astral raw files were searched against the Astral-specific predicted library using the command-line version of DIA-NN on a linux cluster to speed up the search. Uniform mass accuracy and scan window cutoffs to all raw files were applied within each search. To derive these values, one sample plate per experimental setup was searched with 'Unrelated runs' enabled and we calculated the cross-sample median of the values that DIA-NN recommends. For the timsTOF setup, these corresponded to 13 ppm MS1 accuracy ('mass' in DIA-NN), 14 ppm MS2 accuracy, and 8 as scan window parameter. For the Astral setup, values were 4ppm, 8ppm, and 6, respectively. To process entire cohorts, 'Unrelated Runs' was disabled, but the options 'match between runs', 'heuristic protein inference', and the 'no shared spectra' function were enabled. Protein inference was set to isoform IDs, neural network classifier to single-pass mode, quantification strategy to robust LC (high precision), and cross-run normalization to RT-dependent. We used DIA-NN's main report.tsv file of the second pass, which contains ion-level quantities. For Orbitrap and timsTOF setup, ions were filtered for a library protein group q value (Lib.PG.Q.Value) below 1% because DIA-NN's 'match between runs' had been applied in the library generation procedure of the 'merged library', see above. For the Astral setup where this was not the case, the ions were filtered according to the global protein group q value (Global.PG.Q.Value) below 1%, DIA-NN's standard recommendation. To infer protein quantities, we used the directLFQ package v. 0.2.9 with default settings and the 'input_type_to_use' switch set to 'diann_precursor_ms1_and_ms2'. directLFQ performs protein quantity estimation to aggregate the filtered precursor-level quantities into protein quantities and additionally normalizes proteomes between samples.[49] Following the standard procedure in mass spectrometry proteomics, normalization aims to report as little protein regulation for as many proteins as possible, a parsimonious explanation of apparent differences. To achieve this, the intensities of all proteins in a given sample are together shifted up or down, preserving intra-sample protein order, to overall best match other samples. This effectively addresses technical variability that affects all proteins of a sample, for instance loading differential sample amounts or drifts in analytical sensitivity. The Astral discovery and replication cohort raw files were separately processed by DIA-NN and directLFQ. This was the main approach and served maximum independence for identification and quantification of the MS biomarkers across both cohorts. For two of the 29 markers, MBP and ACTG1, protein grouping was slightly different in the replication cohort. Protein-specific peptides were identified, but not the identical set of isoform-specific ones - in the discovery cohort, these also originated only from on a subset of samples. In these two cases, protein groups were manually mapped. However, prediction of MS disease course types was based on the entire proteome. Application of its discovery cohort-trained model to the replication cohort required uniform protein grouping across thousands of proteins. Thus, for this specific analysis both cohorts were again processed together. All three datasets were deposited to the PRIDE repository.

Heavy Isotope-labeled peptides for targeted mass spectrometry

Based on Astral discovery cohort proteomic data 29 biomarker candidate proteins to target were selected. Thus, peptide-level Astral data were queried to select peptide targets. To wit, up to 15 peptides were selected per protein, prioritizing peptides with lower workflow coefficients of variation and less than 20 amino acids for easier synthesis. For several proteins, only one to five peptides were identified in the Astral data, including CD7, CD27, SDC1, MBP, CTSW, HAMP, and SPP1. For these, additional undetected proteotypic tryptic peptides were determined based on the human UniProt proteome and included, typically one to three per protein. A total of about 300 heavy isotope-labeled peptides were custom synthesized by JPT Peptide Technologies (Berlin, Germany). Peptides were supplied as lyophilizate in pools of 31 peptides, with ten aliquots per pool and 0.7nmol per peptide and aliquot, a format JPT Technologies terms 'SpikeMix labeled (L)'. Four peptides were 21-25 amino acids long, synthesized in the 'MaxiSpikeTides labeled (L)' format, and supplied as individual lyophilizates. Predominantly, the C-terminal lysine or arginine were ^{13}C and ^{15}N isotope-labeled, corresponding to mass shifts of 8 and 10 Daltons, respectively. In four peptides with a single miscleavage, an internal lysine or arginine were labeled. Similarly, six peptides derived from a protein C-terminus and were labeled at an internal leucine, conferring a mass shift of 7 Daltons. Cysteines were carbamidomethylated, equivalent to CSF peptides after sample preparation. The full table of isotope-labeled spike-in peptides is provided on the PRIDE proteomic data repository. Peptides were resuspended in 50% acetonitrile and 0.1% formic acid to create 5μM stock solutions. For measurement of cohort samples, 20fm of each heavy spike-in peptide were loaded in 20μl of 1.5% acetonitrile and 0.1% formic acid solution containing 1nM per peptide.

Spectral library for targeted mass spectrometry

To guide targeted assay development and peak identification later in the cohort samples, a spectral library specific to the peptide targets was recorded on the Astral. For that purpose, 40fm of heavy spike-in in 200ng of CSF peptides were measured in data-independent acquisition mode. To provide deep identification of heavy (spike-in) and light (endogenous) peptides, we measured four 200 m/z slices from 300 to 1100 m/z in a gas phase fractionation approach. This was combined with a FAIMS compensation voltage sweep in steps of 5V from -25V to -65V. Raw files were searched by DIA-NN, with isotope labels set as variable or fixed modifications in separate searches to avoid inflating the search space, the resulting speclib files and ion reports were loaded into Skyline daily (24.1.1.398) for library creation. Data and more details are provided on the PRIDE repository.

Targeted mass spectrometry assay and cohort measurement

A targeted assay was built on the Stellar mass spectrometer (Thermo). For chromatography, we used the EvoSep One, an 8cm ion opticks column, and a 21-minute gradient, corresponding to a throughput of 60SPD. The Stellar was operated in parallel reaction monitoring (PRM) mode, selecting light (endogenous) and heavy (spike-in) precursor peptides according to a pre-defined precursor table for fragmentation and MS2 or MS3 spectrum measurement. Abundance of endogenous CSF peptides was quantified as ratio to the 20fm of heavy spike-in peptide based on peak area integration in the Skyline software. RT windows for targeting were typically 1.5-3 minutes long and adjusted by the Stellar's adaptive RT scheduling function based on a survey scan series of covering the 400-1000m/z range in twelve 50 m/z windows in each targeting cycle. The cycle time was set to 1 second, and the maximum injection time mode to dynamic. The MS2 scan rate was set to 125, the MS3 scan rate to 67 kDa/s. The FAIMS compensation voltages were optimized by two sweeps, first with 5V steps and subsequently with 1V steps around the initial optimum, repeatedly measuring CSF pooled from an entire sample plate and heavy spike-ins added. The PRM conductor plug-in in the Skyline software suite was used to distribute the initially about 300 targets across multiple assay variants optimally according to their retention times. Data were loaded into Skyline for manual evaluation, selecting peptide targets for a given protein, suitable fragment ions, and optimal compensations voltages. Targets with low or noisy signals of endogenous peptides were iteratively excluded. Likewise, targets lacking a clear spike-in signal were removed in rare cases. From about one hundred pre-filtered targets, an assay covering all 29 proteins by 1-3 peptides to a total of 42 peptides in a single measurement was constructed. Selection was based on the signals of endogenous peptides, giving priority to the best peptide(s) of low abundant proteins. Peptides for more abundant proteins were chosen dependently to reduce RT overlap across the gradient, typically having multiple peptides per protein as options. To boost the signal of low abundant targets, the collision type, i.e. CID or HCD, and the associated collision energy were optimized. For 20 of the 42 peptides, in particular low abundant ones, a MS3 assay was developed to remove interference or improve the signal to noise ratio. Typically, 2-4 MS2 fragments were selected for MS3 fragmentation, guided by MS2 fragment abundance, charge, and length, as well as empirical testing of different MS2 fragments. The Stellar targeting method, the full precursor table, the table of fragments used for quantification in Skyline, and the cohort raw files were deposited to the PRIDE repository. After cohort measurement, peptides were filtered based on CVs in the discovery cohort. Specifically, peptides were removed if their workflow CV for MS pool samples was higher than 50% or if the biological effect across CSF donor samples, calculated as CV, was not at least twice the workflow CV. Additionally, only the best peptide per protein was kept, pruning the marker list to 22 peptides and proteins. The Astral marker set contained the SPP1 isoforms three and five in separate protein groups, namely P10451-3 and P10451-5. The final 22 peptides included one peptide specific to isoform five, QNLLAPQTLPSK, as well as a protein-specific peptide shared by all SPP1 isoforms, ISHELDSASSEVN.

QUANTIFICATION AND STATISTICAL ANALYSIS

Bioinformatic analysis

Data were generally analyzed in Python (version 3.10.09) using numpy (version 1.23.5), pandas (1.5.2), sci-kit-learn (1.2.1), and scanpy (1.9.3) unless stated otherwise. Protein level quantities as calculated by directLFQ were loaded and log2-transformed for further analysis except for coefficient of variation analysis. In the timsTOF dataset, the fourteen samples with less than 500 quantified proteins were excluded. There was no clear disease or other pattern across these samples. Additionally, another fourteen samples with hemorrhagic stroke were removed from further biological analyses, leaving 164 ischemic stroke samples in the dataset. In the Astral dataset, one sample with less than 1000 quantified proteins was excluded. Six MS samples with unknown disease courses were excluded from MS disease course-related analyses. For calculation of Pearson correlation for clinical measures of QAlb and total protein content in linear space, three outliers with a very high QAlb greater than 150 were excluded.

For most analyses, missing protein quantities were not imputed, and calculations were based on present quantities. For calculation of covariate effect on the CSF proteome by principal component analysis, which requires complete data, proteins were filtered for present quantifications across at least 75% of samples, corrected for sample plate batches, and missing values were imputed according to the five K-nearest neighbors (scikit-learn). Explained variabilities by covariates were calculated by integrating the R^2 of the linear regressions between covariates and projections of samples in the top 20 principal components. The R^2 of each component was weighed according to variability accounted for by the principal component.

Batch correction was done by a custom function shifting the means of batches onto the global mean and likewise scaling the standard deviation of a batch onto the global standard deviation. For biological analyses, the 96well plate was taken as batch. Hierarchical clustering was based on Pearson correlation as a distance metric and the average linkage mode.

Relevant statistical data, including for individual proteins, can be found in the supplemental tables associated with the respective figures. Descriptions of statistical tests and analyses are provided in individual sections below. Samples relate to CSF from different donors. Generally, all available samples were used. The above STAR Methods sections on study collectives give the numbers of samples across diagnosis groups and cohorts, after exclusion of samples, thus reflecting the number of samples used for analysis. Table S1 summarizes the sample numbers across all primary analyses and further breaks them down by oligoclonal band status which is relevant to the biomarker application case. Few proteins were observed across all samples in the discovery proteomics datasets acquired by Astral and timsTOF platforms. Thus, the number of observations across the cohort is provided for each protein in Table S1, more specifically in the spreadsheets that also contain protein intensity values across samples. In bar plots and similar visualizations, center values represent the mean and error bars the standard deviation. Except for differential abundance analysis,

which is further discussed below, all other analyses were descriptive (e.g., PCA, variance decomposition), correlation-based, or machine-learning-based, and therefore did not require formal testing of statistical assumptions related to normality or variance homogeneity.

Differential abundance analysis

For all differential abundance analyses in this paper, proteins with observations in less than 20% of samples were removed. For completeness, we provide both the results for this filtered protein set and an additional table with results for all proteins in the Table S2. Samples with missing QAlb values and strong erythrocyte contamination (labeled ++, +++, or ''bloody'') were removed from the analyses. The analyses were performed on log2-transformed data using the R package limma (limma version 3.50.0,[113] R version 4.3.0). Empirical Bayes moderation of the standard errors was applied as per the *limma* framework to improve variance estimation. Adjustment for multiple testing was applied using Benjamini-Hochberg correction, with a significance level defined at FDR < 0.05.

No formal tests were performed to assess whether the data met the assumptions of limma's linear modeling framework. However, the use of binary diagnosis group indicators and $\log_2$-transformed continuous variables was considered appropriate for the linear model. This transformation mitigated the right-skewed distributions of proteins and covariates and was expected to support approximate normality of model residuals. Limma does not require normality of raw data and is robust to moderate deviations from normality. Heterogeneity of variance across proteins was addressed by empirical Bayes moderation, stabilizing variance estimates and improving the reliability of statistical inference in high-dimensional settings. This moderation benefitted from the large number of both samples and proteins in our dataset.

The primary differential abundance analyses presented in the paper's main figures include age, sex, status of erythrocyte presence (0 for ''no presence'' or +, ++, +++, 'bloody' for ''presence''), difference to age-adjusted QAlb, and 96-well plate as covariates, alongside the diagnosis groups. This equally applies to ancillary analyses in the supplement unless stated otherwise. One such supportive analysis in Figures S2N–S2Q additionally included CSF leukocyte count as covariate. Leukocyte count was excluded from the primary analyses because elevated values are inherently tied to certain diseases, resulting in redundancy for the regression model. Except for age, numeric clinical covariates were $\log_2$-transformed to reduce right-skewness and to align with the $\log_2$-transformed protein intensities. Because some samples had a CSF leukocyte count of zero, one was added to all leukocyte count values prior to $\log_2$ transformation.

Generally, we used patient and clinical CSF parameters as covariates separate to the diagnosis group status, correcting for covariate influence through the linear regression, in a single global regression model that covered all diagnoses. In supportive analyses, we studied potential limits of this approach. To explore potential disease-specific covariate effects, covariates effects were fitted separately in each diagnosis group and subsequently correlated to one another or to the results from the standard model across diagnosis groups, see Figures S2P–S2R. Accordingly, these individual regression models included all of the above covariates but did not contain a diagnosis group variable. In a related analysis that focused on age, we reverted to the global regression model covering all diagnosis groups but this time enabled interaction terms between the diagnosis group and age and sex, see Figure S3A. In an additional supportive analysis, we explored determining healthy aging and sex effects and correcting for these effects in disease samples to permit a refined analysis of disease effects. This followed a two-step approach: Initially, a first regression model was fitted within the neurological control group to estimate the effects of age and sex, which were then subtracted from the global protein measurements to adjust for these covariates. Subsequently, a second regression model was applied to the adjusted data across all diseases, which included the same set of covariates as in our standard regression model above, except for age and sex which had been corrected for. To contextualize this two-step approach, we compared the resulting disease effects to those disease effects from the standard regression analysis, see Figure S3B. To investigate pitfalls of this correction strategy itself, we analyzed the consistency of aging effects across age segments. This is because correction based on linear regression can lead to artefacts, if effects are far from linear, in particular when extrapolating from the age range observed in samples for fitting, i.e. controls, to a different age range observed in the target samples such as disease samples. For that purpose, we divided the controls into age four quantiles, fitted separate models within each quantile, and analyzed the correlation of age effect estimates across these quartiles, see Figure S3C.

Beyond the above control analyses on how to best integrate covariates, we focused on disease effects obtained from regression analyses. In an initial survey analysis, broader diagnosis groups like multiple sclerosis or other autoimmunity were used as conditions. Analyses involving specific subgroups thereof, such as progressive or relapsing MS, or other forms of other CNS and PNS autoimmunity, were performed in separate regression models, keeping the same model parameters, including the above-described covariates. Additionally, for MS, we determined the set of proteins that are significantly regulated (p. adj. < 5%) in all comparisons with other individual diagnosis groups, see Figures 3C–3E. For application of that significance filter, the broad group of other autoimmune conditions was used. For the retained proteins, the fold changes between MS and specific other autoimmune conditions are additionally shown in the heatmap.

Protein annotation and enrichment analysis

Proteins were annotated with gene ontology (GO) terms for biological process, cellular compartment, molecular function, as well as UniProt annotations for Keywords, protein family, glycosylation, and presence of transmembrane domains, all downloaded from

UniProt in February 2023. Annotation of cell type origin of CSF proteins was derived from cell type proteomes from two literature datasets on brain cell types and immune cells.[53,54] Regarding the brain cell type dataset, only the most adult/longest cultured condition per cell type was included. Similarly, erythrocytes and thrombocytes from the immune cell dataset were excluded as they had substantially fewer identified proteins. Proteins were then filtered for at least 70% of quantifications in at least one cell type within datasets, and remaining missing values were imputed by a downshifted distribution using default parameters in the Perseus software.[111] Proteins with ANOVA-significant abundance across cell type ($q < 5\%$, $s0 = 0.1$) were kept and normalized by subtracting the median intensity across samples. Proteins were clustered according to Pearson correlation, and protein clusters were annotated for cell types if there were clear trends. These cell type associations were matched to CSF proteins based on gene names for annotation enrichment.

Analysis of annotation enrichment on continuous scales was performed using the 1D annotation enrichment method built into Perseus.[111,112] This was applied to the continuous scales of regression estimators for diagnoses and covariates, as well as the aggregate SHAP values of proteins. Annotation enrichment based on dichotomic association, e.g., with the set of top 250 drivers of the diagnosis classification model, was done with Fisher's Exact test.

Machine learning-based sample classification and model explanation

Machine learning classification of CSF samples was performed in Python (version 3.9.13) with the classifier XGBoost (version 1.7.1), which was compatible with missing values in the protein matrix and thus did not require imputation or filtering to proteins quantified in all samples.[109] Classification based on the entire proteome was performed for multiple applications. This included a multi-class model surveying all CNS conditions of the full discovery cohort and binary prediction of MS vs. neurological controls to determine pseudo-times, both based on timsTOF data. For the classification of progressive and relapsing MS disease courses, Astral data were used. Protein quantities, the sample plate batch, and the clinical covariates age, sex, QAlb, QIgG, and leukocyte counts in the CSF were used as features. Omitting the batch information or using batch-corrected data did not materially alter the performance, and the batch itself never presented as top feature when included. Following a cross-validation design, test-train-splitting with 5 splits was repeated 50 times. Dataset splitting was the first step of machine learning. The splits were stratified for class frequency using scikit learn's RepeatedStratifiedKFold function. The random state was set to 42. To address class imbalance when classifying two classes with discrepant sample numbers, the minority class was weighed more strongly by setting the balanced_weights parameter to True and tuning the scale_pos_weight parameter. That parameter was set to 0.2 and 0.02 to classify MS vs neurological controls and RRMS vs PMS, respectively. Multi-class prediction was not compatible with class balancing.

Regarding MS disease course types, 91 PPMS and 51 SPMS samples in the discovery cohort were pooled as PMS due to the low sample number reflecting their lower natural frequency compared to RRMS, of which there were 836 in the Astral dataset. Since the PMS and RRMS inherently differ in age and leukocyte count, we performed a control analysis on a stratified subset to exclude artefactual protein associations that reflected these covariates rather than disease course-specific effects. For that purpose, we downsampled the RRMS type to 300 samples to match the age and leukocyte count distributions of the PMS samples. The downsampling was implemented by modeling a multivariate normal distribution for PMS samples and drawing 300 random datapoints from it. For each PMS data point, the closest existing RRMS sample in the multivariate space was identified, selected for the stratified subset, and excluded from repeated selection. To enable harmonic integration of both age and log2-transformed leukocyte counts, these dimensions were Z-scored, and the Euclidian L1 distance was taken as a metric. For application to the replication cohort, two models were trained on the discovery cohort samples, once on the entire set and once on the stratified subset.

To explain the machine learning models, SHAP values were calculated with the shap package (0.42.1) for each combination of feature (protein), donor CSF sample, cross-validation repeat, and in case of multi-class prediction, also a class representing disease to predict for or against it.[110] Thus, positive SHAP values indicate that a protein predicts for a given disease in a given sample, for instance if the protein's abundance is high in that sample. Conversely, a negative SHAP values indicates that the protein predicts against that disease, for instance if the protein's abundance is low in another sample. Beeswarm plots showing the relationship between protein intensity and SHAP value of the top 20 features across samples were directly generated by the shap package.

To aggregate the SHAP information within proteins, the mean absolute value across cross-validation repeats and then samples was calculated. To add directionality information, we mean-averaged values across cross-validation repeats first and then calculated the two mean SHAP values across samples with a high and low value of the original protein intensity, using the top and bottom third quantile of samples by intensity, respectively. Samples without protein quantity were excluded for the directionality determination. If a SHAP mean was higher in samples with high protein abundance, the directionality sign was positive, meaning that high protein abundance predicts for the disease. Conversely, the directionality sign was negative if the SHAP mean was higher in samples with low protein abundance. For principal component analysis, the SHAP matrix was flattened by combining class and feature, and missing SHAP values reflective of no importance were imputed with 0.

Analysis of biomarkers separating MS from other CNS inflammation

Classification followed the survey multi-class machine learning analysis with adaptations. To generate a biomarker panel that does not require measurements from existing markers, only age and sex were included as covariates alongside the proteome. Similarly, the sample plate was not included as batch information, because application in clinical routine would not permit this, instead

necessitating biomarkers with strong biological effects that supersede potential batch effects. Likewise, this was reflected in our replication cohort, having a different batch status not available to classifier training in our discovery cohort.

Feature selection in the Astral discovery cohort followed a nested cross-validation design. Both outer and inner set used 5-fold test-train-splitting, the outer one was repeated 10 times. Feature selection was performed separately in each inner set to avoid test-train leakage between inner and outer sets. An initial model was trained and features ranked according to the mean absolute SHAP value, pruning the data to 100 features in a computationally efficient manner. Further selection was based on recursive elimination in steps of five, using the scikit-learn package. At each iteration, ROC AUCs were calculated for the identification of MS vs all other diseases based on the probabilities of the multi-class model. Based on saturation of classification performance in the inner set and compatibility with a downstream targeted mass spectrometry assay, 30 features were chosen for further analysis. Model performance was further characterized with regards to which diseases were hard to separate from each other. For that purpose, the ROC AUC for each disease 'A' was determined in the outer test set, once in the entire test set and additionally in all pair-wise subsets of the disease of interest 'A' and a second other disease 'B'. In the multi-class classification, wrong predictions also include other diseases (C etc.). Thus, unlike binary classification, the ROC AUCs for disease A and B are independent and a given ROC AUC for disease A does not determine the ROC AUC for disease B in the cohort subset containing only the true cases of A and true cases of B.

The performance of the proteomic model was compared to a control model of clinical markers in a side-by-side fashion on identical outer test sets. OCB status was the only accepted CSF biomarker for MS, meaning included in the McDonald criteria. Other routine CSF parameters that guide clinical decisions like QAlb, QIgG, and leukocyte count, as well as the generally available age and sex, were additionally included for a high comparative strictness. Kappa free light chain information was not available, but literature data indicated performance equal to the included OCB status, making omission not a problem. The discovery cohort, but not the replication cohort, contained samples lacking OCB status information. As these were selectively non-MS samples, a classifier would have learnt and applied the artefactual association that absence of OCB testing excluded MS. To avoid that, we excluded these samples from the control model of established markers. Consequentially, these samples were always part of the inner training data for the proteomic model, ensuring that the outer testing set remained identical for proteomic and control model, yielding paired ROC AUCs for both models. Each cross-validation repeat yielded such a pair. Thus, a paired two-sided t-test was used for statistical analysis in the discovery cohort. As metrics of the individual models, the mean and standard deviation of ROC AUC values across cross-validation repeats were calculated and plotted in the bar plots.

For application to the replication cohort, both a final proteomic and a final control model were trained on the discovery cohort. For each model, this yielded a single prediction in the replication cohort. To estimate the ROC AUC spread from such a single prediction, 1000 bootstraps were drawn using the scikit-learn resample function. Each bootstrap yielded ROC AUC values for both models, again in a paired setting. Significance between models was estimated by a two-sided empirical p-value approach, counting the fraction of bootstraps with ROC AUC differences below or above zero and doubling the smaller of the two values. Similar to before, the mean and standard deviation of ROC AUCs across bootstraps were calculated as metrics of individual model performance. These ROC data are provided in Table S7.

For the biomarker panel application here, the calculation of SHAP values was simplified. Instead of aggregating SHAP values across cross-validation splits, the SHAP values were obtained from the above final model trained on the entire discovery cohort and applied to the replication cohort. As this happened after feature selection, it provided a homogenous SHAP dataset that included all biomarker panel proteins. This circumvented challenges with integrating SHAP values across cross-validation splits that could have different protein sets after feature selection.

Targeted assay data was analyzed equivalently, both with regards to SHAP value calculation and model performance comparison. However, it was further simplified as we did not use machine-learning based feature selection.

Sample pseudo-times and protein trajectories

Protein trajectories were derived by placing CSF samples on pseudo-times between two states. These pseudo-times resulted from the supervised classification of MS vs neurological controls as well as PMS vs RRMS, once for the entire dataset and once for the subset stratified for age and leukocyte count. Pseudo-times were calculated as the SHAP value sum of all features (proteins) within one sample. Regarding directionality, positive values predicted for MS and PMS, respectively. Trajectories along the pseudo-times were calculated both for the intensity of quantified proteins and the local detection frequency of proteins using the dataset corrected for sample plate batches. For intensity trajectories, the proteins were filtered for at least 20% quantifications in the dataset to remove noise. Similarly, for detection trajectories, proteins had to have quantifications between 5% and 95% of samples. The local detection rate was calculated as fraction of quantifications in a 30-sample window sliding along the pseudo-time. Thus, the window size was much smaller than the datasets of 1935 samples (MS vs. neurological controls, timsTOF data), 978 or 442 samples (PMS vs RRMS, Astral data, full set or RMS-down sampled). LOESS regression from the statsmodels package (v0.13.5) was used to fit the trajectories with a fraction parameter of 0.3 to the data point cloud of protein intensities or local detection rates along the pseudo-time. The trajectories were prioritized based on the explained variance and linearity. The explained variance (r^2) of the fit was calculated according to the summed square method using the real datapoints and prediction by the curve fit. The linearity was calculated as the Pearson coefficient between the trajectory and the pseudo-time to exclude non-monotonous trajectories. For clarity, PMS-associated trajectories had a positive Pearson coefficient, RRMS-associated ones had a negative Pearson coefficient because the RRMS cases

predominated at negative pseudo-times and PMS cases at positive pseudo-times. To cluster the trajectories, Z-scoring was applied within trajectories and trajectories with an r^2 of 0.05 or less were excluded. Hierarchical clustering was based on Pearson coefficients and average linkage. To analyze concordance between the full dataset and the stratified subset regarding RRMS vs PMS trajectories, proteins were additionally filtered for an absolute Pearson coefficient of 0.8 or higher in both datasets. To identify immune cell type sources of secreted proteins, proteins with PMS-associated trajectories in both dataset variants were mapped to a literature dataset on immune cell types.[53]

ADDITIONAL RESOURCES

This study has not generated or contributed to a new website/forum and it is not part of a clinical trial.

CellPress
OPEN ACCESS

Supplemental figures

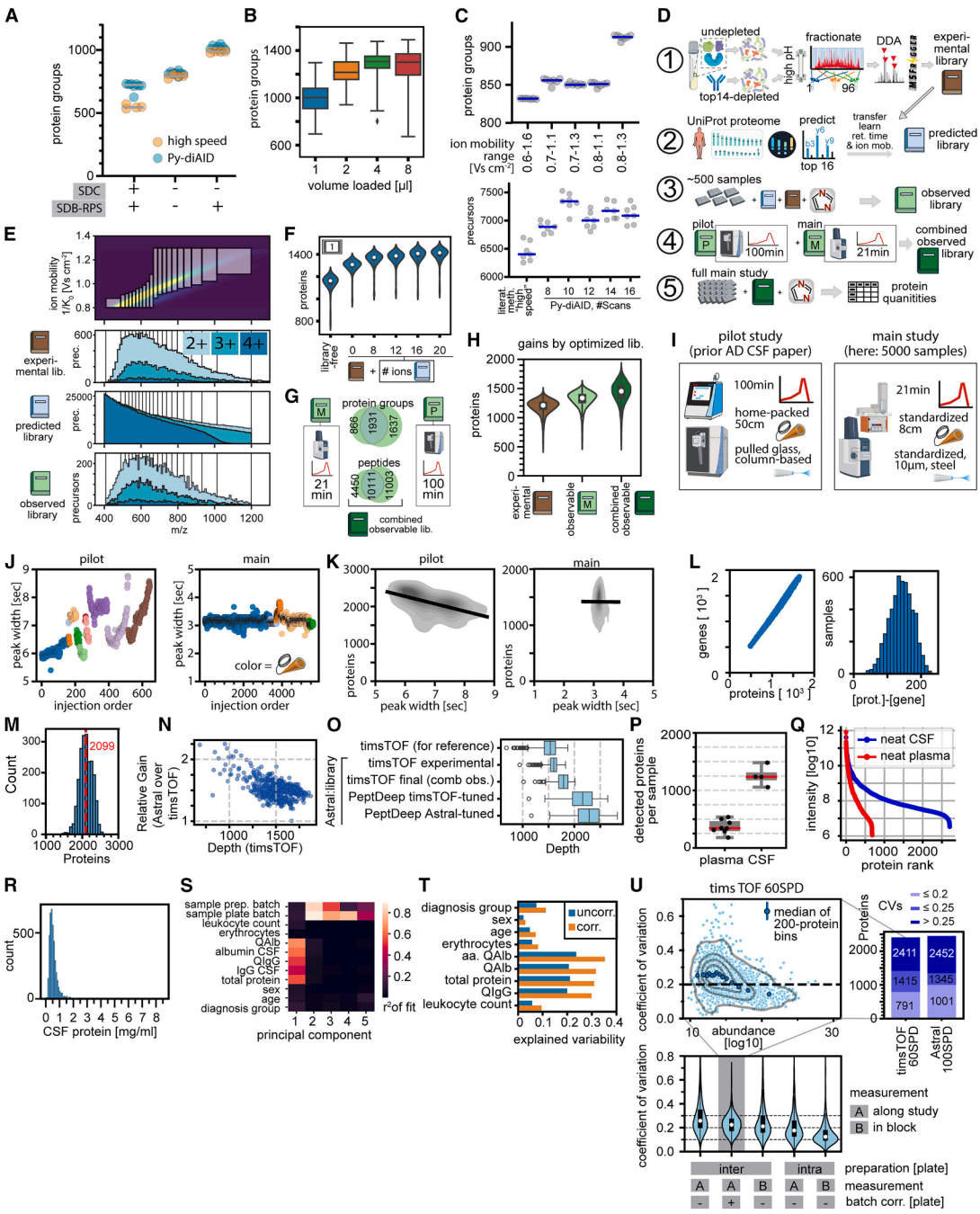


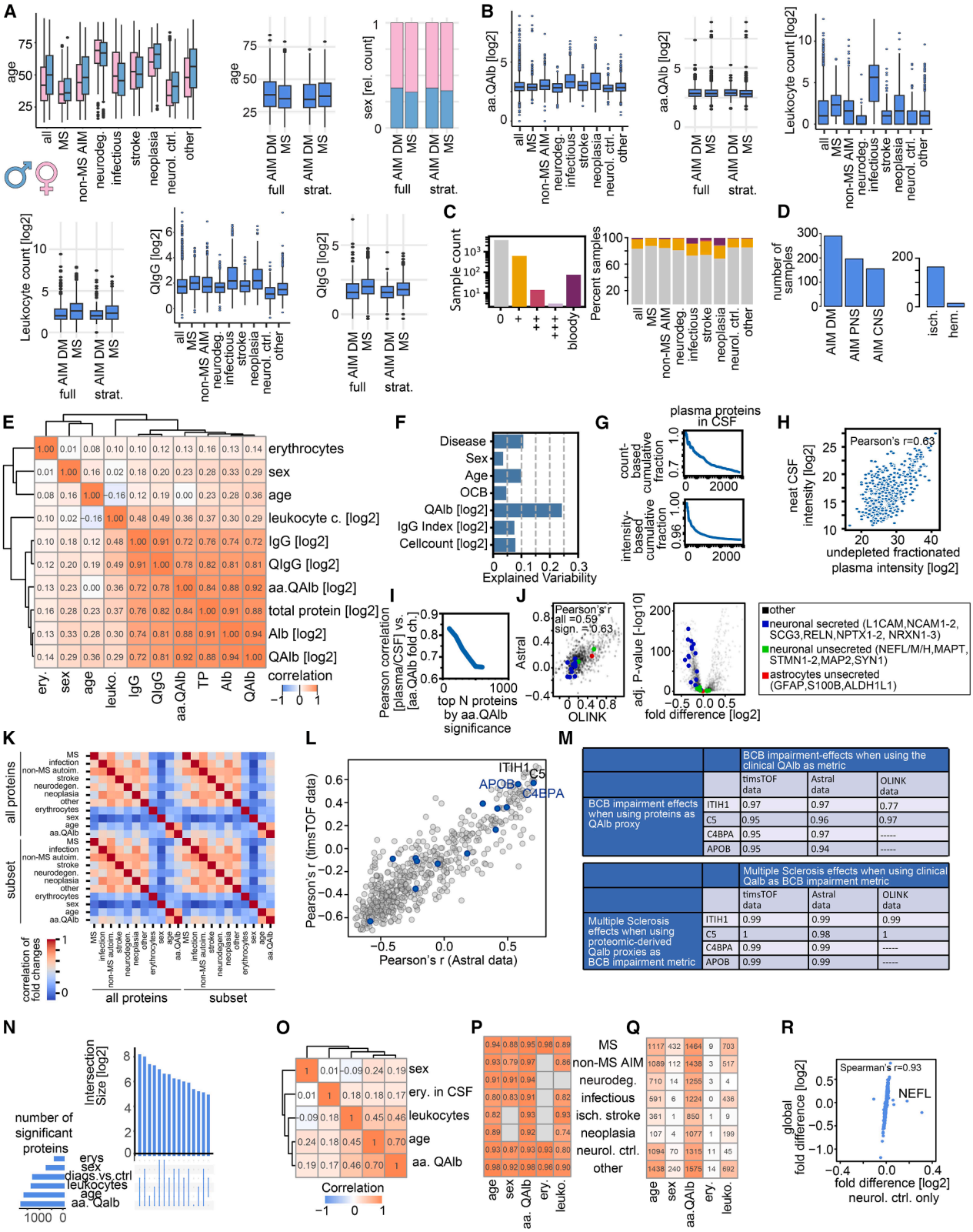
Figure S1. Workflow development and robustness, related to Figure 1

(A) Optimization of sample denaturation and peptide clean-up. The number of identified protein groups depends on the inclusion of sodium deoxycholate (SDC) in the solution added to CSF and on the inclusion of a styroldivinylbenzol-reversed-phase sulfonate (SDB-RPS) resin-based peptide clean-up in the StageTip format. Data acquisition by "high-speed" dia-PASEF method from the original paper or method with windows optimized by the Py-diAID package. Measurement on the timsTOF Pro 2 instrument with the Evosep One 21-min gradient and an 8-cm PepSep column. Ion mobility covered 0.6–1.6 V·s·cm^{-2}. Peptide amount loaded was 500 ng. Protein search in Spectronaut 15 using an experimental library from high-pH CSF fractions generated by Pulsar.

(legend continued on next page)

(B) Optimization of the volume loaded of peptide solution from cohort CSF samples. Number of identified protein groups depending on the volume of peptide solution after SDB-RPS clean-up. 1 μL corresponds to about 100 ng of peptide. The 96-well plate 1 of the cross-sectional study was measured twice for each loaded volume. Data were acquired on the timsTOF Pro 2, with a final dia-PASEF method that was also used for cohort measurement.
(C) Optimization of the dia-PASEF method. Top: optimization of the ion mobility range. Number of identified protein groups depending on ion mobility covered by a variant of the high-speed dia-PASEF method described in the original paper. The DIA windows of the acquisition method were adjusted for their ion mobility boundaries (x) but kept equidistant in the *m/z* dimension. Peptide amount loaded was 400 ng. Search in library-free mode. Bottom: optimization of the number of scans per cycle. Each scan comprises two windows that together cover the entire ion mobility range from upper and lower end, see (E) for illustration. The number of scans was varied from 8 to 16 and resulting windows designed by the Py-diAID package to cover equal numbers of precursors based on data from the experimental high-pH fraction library. This library was also used for the spectral searches. All methods used an ion mobility range of 0.8–1.3 $V{\cdot}s{\cdot}cm^{-2}$. The method with 10 scans corresponding to 20 windows was chosen as the final.
(D) Overall library generation scheme for processing the cohort, which was measured on the timsTOF Pro 2 platform. Step 1: generation of the final experimental library combining high-pH fractions, top14 plasma depletion kit, 24–96 high-pH fractions, DDA measurement, and MSFragger processing. Step 2: prediction of top 16 fragments for charge two to four precursors of tryptic peptides of the entire human genome. Transfer learning using the experimental library to predict retention time and ion mobility. Removal of entries contained in the experimental library. Software AlphaPeptDeep. Step 3: DIA-NN search of around 500 samples using both the experimental and predicted library, producing a library with ions actually observable in unfractionated CSF samples. Step 4: computational merging of observable libraries of the pilot and main study, both generated by steps one to three. Retention time alignment and ion mobility prediction by transfer learning in AlphaPeptDeep. Step 5: DIA-NN search to obtain protein quantities across the full set of samples, separately for the pilot and main study.
(E) Precursor distributions across libraries and the final dia-PASEF acquisition method. Top: final acquisition method package on top of the precursor distribution according to the experimental library, which guided design of the method. Bottom: precursor distribution in *m/z* dimension across libraries generated as part of the library generation workflow, as in (D). The distributions of precursors are color coded according to precursor charge. Vertical lines are the window boundaries of the acquisition method.
(F) Benchmarking of library complementation by prediction. Number of identified protein groups (y) depending on the number of fragments per precursor (x) in the predicted libraries (blue book icon), which was used in addition to the experimental libraries (brown icon). Results derived from DIA-NN searching the first 96-well plate of CSF samples from the timsTOF Pro 2 cohort measurement.
(G) Overlap and merging of observed libraries, corresponding to step 4 in the library generation scheme, see (D). Overlap of peptides and proteins between libraries of the main ("M") study on the timsTOF Pro 2 and the pilot study ("P") on Orbitrap platforms (Q Exactive HF-X, Exploris 480). Liquid chromatography gradients were 21 min (M) and 100 min (P). The combined observable library was constructed as the union of both above libraries. The retention time was harmonized by LOESS regression to a normalized range and ion mobility predicted for entries from the pilot observed library (STAR Methods).
(H) Overall benefit of library optimization in the final dataset of the main timsTOF Pro 2 study. Number of proteins per sample across the entire main study cohort of 5,059 samples, depending on the library used.
(I) Mass spectrometry setup for the pilot and main study.
(J) Stability of chromatographic peak separation across the pilot and main study. Chromatographic performance as peak width in retention time averaged across peptides along the injection order, for the pilot study (left) and the main study (center). Datapoints are samples in order of mass spectrometric measurement, and liquid chromatography columns are color coded. Distribution of peak width values (right).
(K) Effect of chromatographic peak separation on protein identification across the pilot and main study.
(L) Detection of isoforms across the main study cohort (timsTOF Pro2) processed with the combined observed library, the final library, in DIA-NN. Correlation of protein group number (x) in thousands to gene number (y) across samples (left). Difference in protein and gene numbers reflects resolved isoforms that derive from the same gene. Distribution of resolved isoforms across samples (right).
(M) Distribution of protein identifications by the Orbitrap Astral across samples. Re-measurement with this mass spectrometer covered about 2,000 samples of the discovery cohort. The median is indicated by the vertical line.
(N) Identification boost by the Orbitrap Astral compared with the timsTOF Pro 2. Relative gain in number of proteins per sample (y) vs. number of proteins identified by the timsTOF platform (x). Datapoints are samples. Samples with few identifications benefit particularly from the Astral.
(O) Benchmarking of predicted libraries for analysis of the Orbitrap Astral data. Tukey's boxplots show the distribution of identified proteins (x) depending on library (y). Libraries were predicted by AlphaPeptDeep. Samples of the first ten 96-well plates of the discovery cohort were included. The very top row shows timsTOF Pro 2-acquired data searched against the final library for this platform as a reference for comparison with the Astral data. The below rows show corresponding Astral data from the same samples, searched against various libraries: from top to bottom, this includes the experimental high-pH fraction library acquired on the timsTOF platform, the above final timsTOF library, an AlphaPeptDeep-predicted library in which fragment spectra prediction was based on HeLa training data acquired on the timsTOF platform, and a likewise predicted library in which training data were acquired on the Astral platform itself.
(P) Typical number of detected proteins in neat, non-fractionated and non-depleted, plasma and CSF across comparable literature studies. Datapoints are averages across samples within the individual studies, which we reference in our paper.
(Q) Abundance distribution of proteins in neat CSF and plasma. The data show that CSF has a more moderate protein intensity distribution than plasma, leading to less of a dynamic range challenge. Plasma data from Geyer et al.[43] Intensities normalized for visualization so that the most highly abundant proteins have equal intensity.
(R) Total protein in CSF across samples of our discovery cohort measured by nephelometry in the clinic
(S) Structure of dataset before batch correction. Contribution of technical batches, diagnosis group status and diagnosis-unspecific covariates to the first five principal components (top). Variance ratios for fraction of total variability explained by a given principal component (box on top) or variable (box on right side). Among technical batches, 96-well sample plate explains most variability, sample preparation accounts for substantial fraction of that variability between plates. Uniform manifold approximation and projection (UMAP) embedding for sample preparation batch and plate (bottom).
(T) Explained variability of biological variables before and after batch correction.
(U) Coefficients of variation (CVs) for proteomic measurements of non-MS quality control pool samples. Distributions of CVs depending on the sample preparation and measurement design (left). Preparation of samples on the same 96-well plate (intra-plate) or on different plates on different days over the course of sample preparation for the study (inter-plate). Measurement of CV samples along the study (A) or in block with one CV sample measured directly after another (B). Batch correction for CV samples with inter-plate preparation and measurement along the study. Correlation of CVs with protein abundance (right). Batch-corrected measurements along study of inter-plate prepared samples used. Data points as proteins, density maps modeled by gray lines. Blue dots represent median CVs and 95% confidence intervals thereof for bins of 200 proteins with binning according to protein abundance starting with highest abundance. Blue dot placed at the abundance center of the bin.

CellPress
OPEN ACCESS



		BCB impairment-effects when using the clinical QAlb as metric		
		timsTOF data	Astral data	OLINK data
BCB impairment effects when using proteins as QAlb proxy	ITIH1	0.97	0.97	0.77
	C5	0.95	0.96	0.97
	C4BPA	0.95	0.97	-----
	APOB	0.95	0.94	-----

		Multiple Sclerosis effects when using clinical Qalb as BCB impairment metric		
		timsTOF data	Astral data	OLINK data
Multiple Sclerosis effects when using proteomic-derived Qalb proxies as BCB impairment metric	ITIH1	0.99	0.99	0.99
	C5	1	0.98	1
	C4BPA	0.99	0.99	-----
	APOB	0.99	0.99	-----

(legend on next page)

Figure S2. Clinical CSF variables and their relation to differential abundance results, related to Figures 1 and 2

(A) Distribution of age for the diagnosis groups and sex in the full discovery cohort (left). Additionally, age (center) and sex (right) distributions are shown for MS and other DM autoimmunity not fulfilling the diagnostic criteria for MS before and after stratification. Before stratification means in the full discovery cohort, after means in the stratified subset of the discovery cohort. Note that stratification was primarily for inflammatory parameters.

(B) Distribution of the age-adjusted albumin ratio, leukocyte count in CSF, and immunoglobulin G ratio. Distributions for each covariate are shown across all diagnosis groups in the full discovery cohort (each left) and specifically for MS and non-MS DM AIM before and after stratification (each right), analogously to (A).

(C) Distribution of semi-quantitative erythrocyte abundance in CSF across samples. Number of samples (y) depending on erythrocyte abundance (x) category (left). Relative frequency of categories across diagnosis groups (right).

(D) Sample sizes for subgroups of non-MS autoimmune diseases (left) and stroke (right). Hemorrhagic stroke samples were excluded from analysis. AIM stands for autoimmunity, DM for demyelinating AIM not fulfilling MS diagnostic criteria, PNS for AIM of the peripheral nervous system with systemic or CNS involvement, and CNS for other AIM of the central nervous system.

(E) Pearson correlation of covariates across discovery cohort samples. Erythrocyte content was dichotomized into groups of none (0) and any content (+, ++, +++, bloody) for correlation. For numeric covariates, $\log_2$-transformed quantities were used.

(F) Variability explained by covariates and disease in the Held et al.[24] OLINK dataset.

(G) Fraction of CSF proteins that are also found in plasma. Fractions (y) are based on the number of proteins (i.e., counts, top) or based on the proteins' intensities in CSF (bottom). Fractions are cumulated along protein abundance ranks from highest- to lowest-abundance CSF protein (x). CSF proteomes are from our timsTOF data. Two plasma proteome datasets were combined for deep coverage: Pernemalm et al.[124] used high-abundance protein depletion, whereas Geyer et al.,[43] did not.

(H) Correlation of protein abundances between CSF (our data) and plasma. Plasma data from deep dataset without high-abundance protein depletion, see (G).

(I) Similarity between QAlb effects across donor CSF and the corresponding effects between neat plasma and CSF pooled from a few donors without brain-CSF-barrier impairment. QAlb fold changes were from our multivariate regression model across CSF samples. Pearson correlation between these fold changes (y), depending on the maximum significance rank of proteins regarding the QAlb effect in the regression (x). QAlb effects across donor CSF correlate with differences between CSF and actual plasma. When only the most significant proteins in the regression are considered, this correlation is particularly high. Plasma-free CSF was CSF from donors with an intact blood-CSF barrier, as characterized by QAlb < 5. The plasma dataset without depletion from (G).

(J) Correlation of age-adjusted QAlb effects between OLINK and Astral data (left). Overall, effects are correlated. However, Astral fold changes are down-shifted below the dashed equality line (y = x), likely due to normalization to the entire proteome in mass spectrometry combined with a large influx of plasma proteins seen in QAlb. Volcano plot for amino acids (aa). QAlb associations in Astral data (right). Independent of technology, secreted neuronal proteins surprisingly have lower QAlb fold changes than unsecreted neuronal or astrocytic proteins. QAlb may correlate with inflammation, leading to the release of intracellular proteins that are normally not secreted, which causes this apparent pattern. Several unsecreted proteins were not contained in OLINK data, thus there are more labels in the Astral volcano plot.

(K) Correlation of all disease and covariate fold changes between standard and subset normalization. Regression fold changes were calculated on differentially normalized timsTOF data. Data were normalized by the DirectLFQ algorithm using the entire proteome in the standard approach. For the alternative subset normalization, proteins found to be not strongly affected by BCB impairment were used. This subset is derived from Aasebø et al.[51]

(L) Correlation of protein abundance with QAlb across samples. Pearson coefficients were calculated using protein quantities from the Astral (x) or timsTOF (y) platform. The top two QAlb-correlating proteins, ITIH1 and C5, were selected as QAlb substitute metrics. Proteins included in our biomarker panel separating MS from other inflammatory conditions are shown in blue. Their location suggests that machine learning unbiasedly selected APOB and C4BPA as an informative feature capturing the individuals' QAlb status.

(M) Benchmarking of pseudo-QAlb. One benchmarking metric is similarity of BCB-impairment effects to those obtained when using actual QAlb (top). Pearson correlation values between BCB-impairment effects fitted by original QAlb and corresponding effects when fitted by an individual protein-based pseudo-QAlb substitute (rows) across proteomic technologies (columns). BCB-impairment effects are regression estimators across all proteins in the proteome. A second benchmarking metric is the similarity of MS effects that are obtained in the above multivariate regressions when BCB-impairment effects are fitted using either the actual QAlb or a given pseudo-QAlb (bottom).

(N) Number of regulated proteins by variable in a multivariate regression model when including leukocyte count in addition to the other covariates. The label diags vs. ctrl. represents all proteins that are significantly differentially abundant in the comparison of any neurological condition to controls.

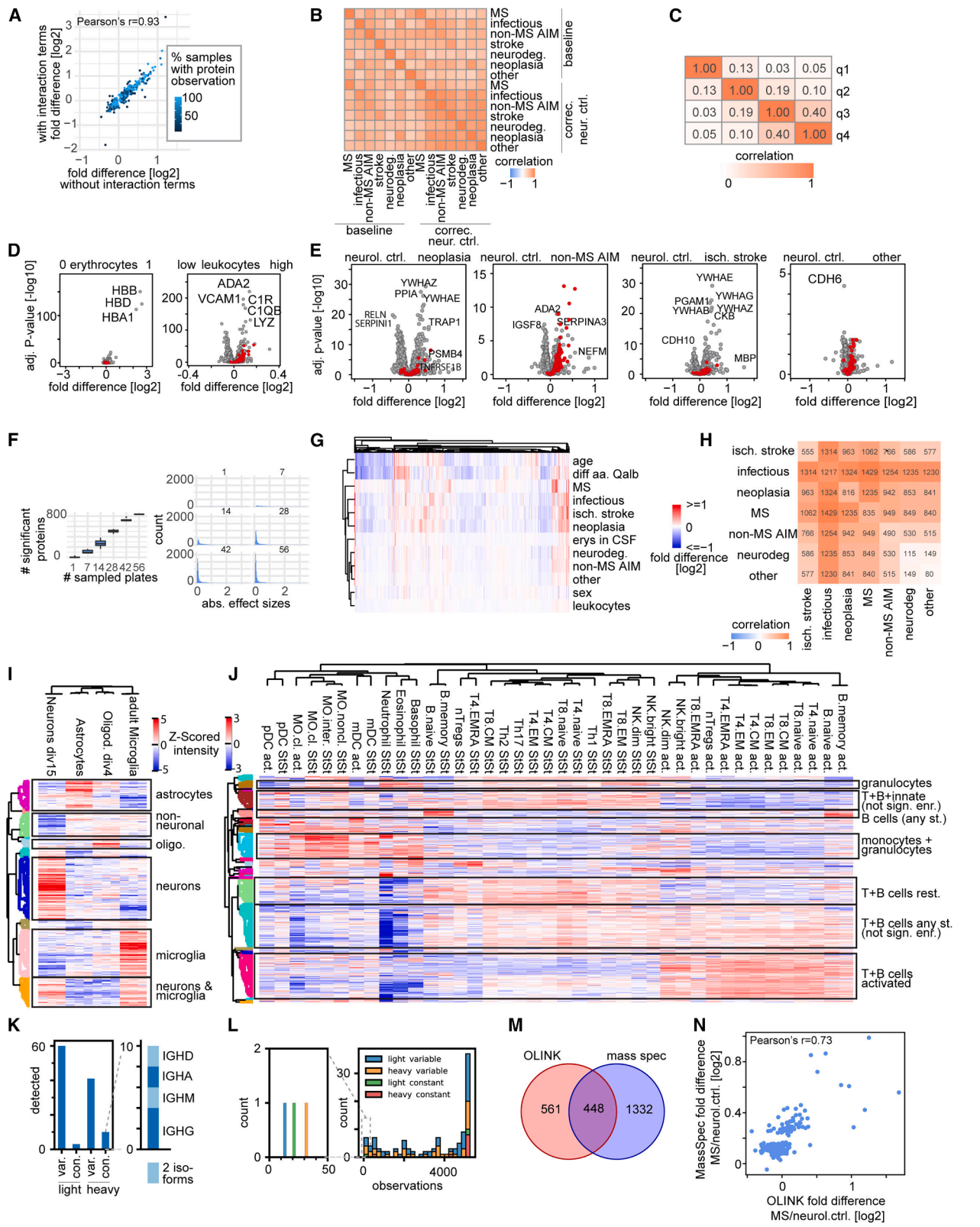
(O) Correlation of protein regulation across clinical covariates when including leukocyte count in addition to the other covariates.

(P) Correlation of the effect sizes in each covariate, for proteins that are significantly differentially abundant in each of the diagnosis groups separately, compared with the complete cohort when including leukocyte count in addition to the other covariates. Gray corresponds to NA when 4 or fewer proteins were significantly differentially abundant in the diagnosis group and no correlation could be computed. One outlier in the stroke group for the QAlb effect sizes was excluded.

(Q) Number of proteins that are used to compute correlations in (P).

(R) Correlation of age fold changes between the regression model in neurological controls only and the global regression model across all disease entities on proteins significant in at least one of the models.

CellPress
OPEN ACCESS



(legend on next page)

Figure S3. Relation of differential expression results across diagnosis groups, related to Figures 2 and 3

(A) Correlation of MS estimators between regression models with and without interaction terms. Datapoints are proteins. The color represents the fraction of samples with an observation of that given protein.

(B) Correlation of disease fold change estimators between global regression model variants that differentially integrate covariates. The standard global regression model ("baseline") is computed on unprocessed data and integrates clinical parameters as model covariates. The alternative model does not include clinical parameters but is calculated on data corrected for covariate effects. This correction was done by fitting a helper regression model for these covariates in neurological controls and subtracting covariate effects from all disease entities accordingly. Both global regression models are calculated for samples from all disease entities.

(C) Correlation of estimators of age fold changes resulting from regression models in individual age quantiles of neurological controls. Proteins were included if they were significantly regulated by age in a regression model across all neurological controls. Age quantiles were 16–26, 27–35, 36–48, and 49–80 years.

(D) Protein regulation by clinical covariates for erythrocytes and leukocytes. Immunoglobulins are shown in red.

(E) Protein regulation by disease groups, each in comparison with neurological controls. Immunoglobulins are shown in red.

(F) Number of significant proteins for different numbers of sampled plates (left). Histogram of absolute effect sizes for these proteins (right).

(G) Effect sizes for all proteins in the diagnosis groups and covariates, including the leukocyte count in addition to the other covariates.

(H) Intersection of differentially abundant proteins in the diagnosis groups.

(I) Clusters of proteins specific to brain cell types. These clusters were used to analyze enrichment of these cell types in the CSF data.

(J) Clusters of proteins specific to blood immune cell types. These clusters were used to analyze enrichment of these cell types in the CSF data.

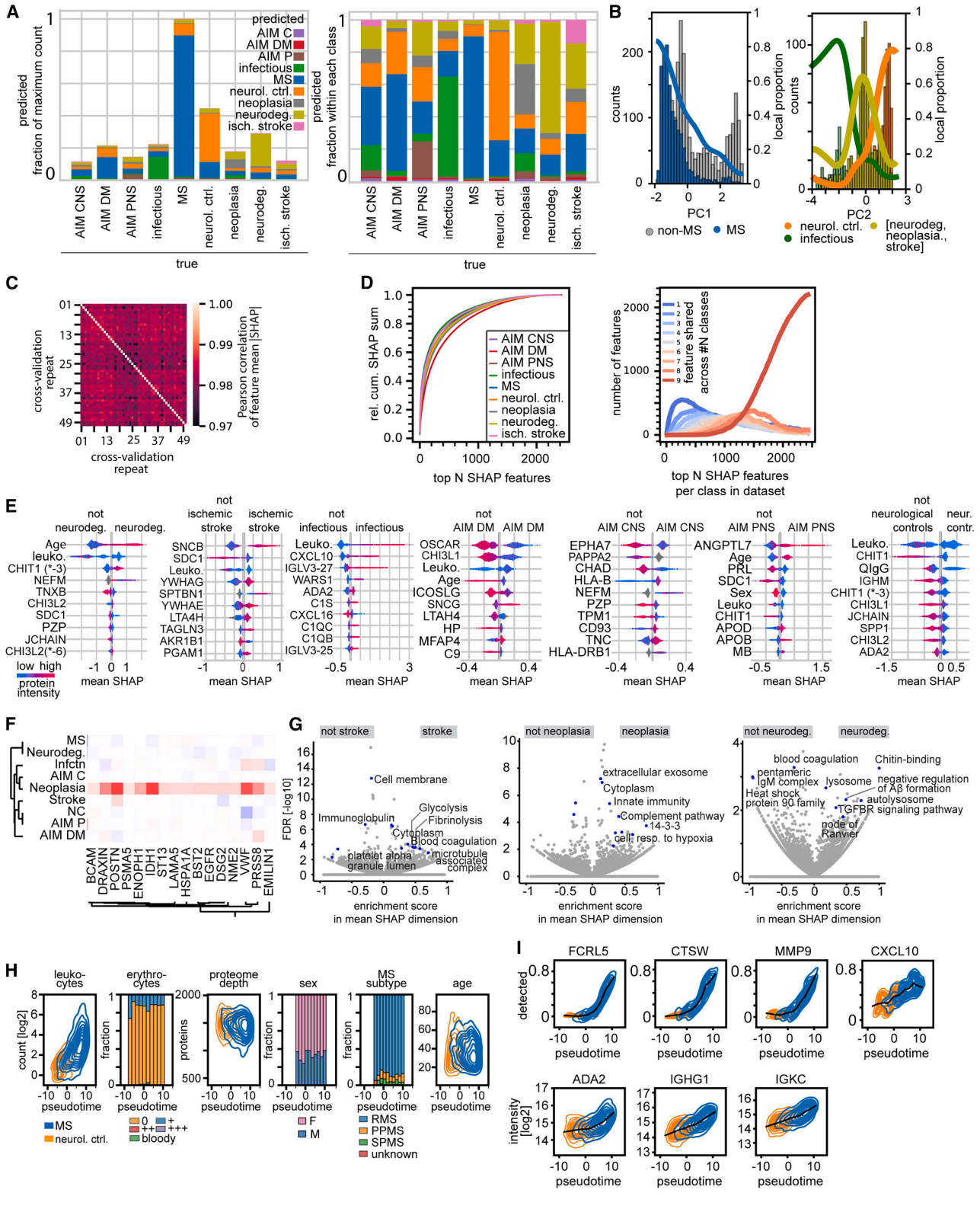
(K) Number of detected immunoglobulin proteins in the CSF.

(L) Number of observations across CSF samples of immunoglobulins.

(M) Overlap in measured proteins between the OLINK dataset and our timsTOF dataset.

(N) Correlation of MS estimators for shared proteins in the OLINK dataset and our timsTOF dataset. Proteins were included if they were significantly regulated in at least one of the datasets. OLINK data from Åkesson et al.[16]

CellPress
OPEN ACCESS



(legend on next page)

Figure S4. Structure and function of features driving supervised classification of diagnosis groups, related to Figure 4

(A) Predicted (y) and true disease classes (x) in the analysis surveying diverse neurological diseases by a multi-class model. The number of predictions (y) is shown for each disease (color coded), either normalized to the maximum count across diseases (left) or within individual diseases (right). Prediction counts are from a confusion matrix but are shown in a bar plot for illustration.

(B) Partial separation of MS samples by principal component one of the SHAP values matrix (left). Similar partial separation by principal component two for disease groups broadly according to the level of inflammation and BCB impairment (right). Infection and neurological controls are well separated, with other less inflammatory diseases lying between them.

(C) Correlation of protein features across cross-validation repeats in the multi-class model with 50 repeats. Absolute Shapley values were mean-aggregated across samples in the cross-validation repeat for each protein feature and protein vectors were correlated.

(D) Information for classification provided by additional features based on their rank. Normalized cumulative distribution function of absolute SHAP values for the top N SHAP features across diagnosis groups (left). Extent of features being unique to or shared across classes, i.e., diagnosis groups, depending on the set size of features included. Top N features according to their SHAP value rank within a class were included as the set for that class. Curve for class overlap of 1 relates to features unique to one class and overlap of 2 to a feature present in the sets of two classes, etc. (right).

(E) Top ten predictive features across diseases in a multi-class model. For each feature, the information contributed to the model is shown by a distribution across samples. Samples are datapoints, the intensity of the protein is color coded, and its contribution is measured by SHAP values to the model, shown on the *x* axis. Positive SHAP values indicate prediction for the given disease class, and negative values indicate the prediction against it, all in the given sample.

(F) Protein cluster predicting for neoplasia. This cluster is a zoom-in from the global clustering of SHAP values in Figure 4F.

(G) Annotation enrichment in the SHAP value dimension for stroke (left), neoplasia (center), and neurodegenerative disease (right). Positive scores (x) reflect enrichment of terms among proteins predicting for the given disease, whereas negative scores conversely reflect enrichment among proteins predicting against that disease. Significance of enrichment is shown on the *y* axis.

(H) Distribution of covariates and proteome depth along the pseudo-time between neurological controls and MS. Proteome depth reflects the number of detected proteins. For numeric variables (leukocytes, age, and proteome depth), distributions are shown separately for MS (blue) and neurological controls (yellow).

(I) Profiles of selected outlier proteins along the pseudo-time between neurological controls and MS. Profiles reflect protein intensities (bottom) or local detection rates (top). Underlying distributions are shown separately for MS (blue) and neurological controls (yellow). The averaged profile is shown by a black line.

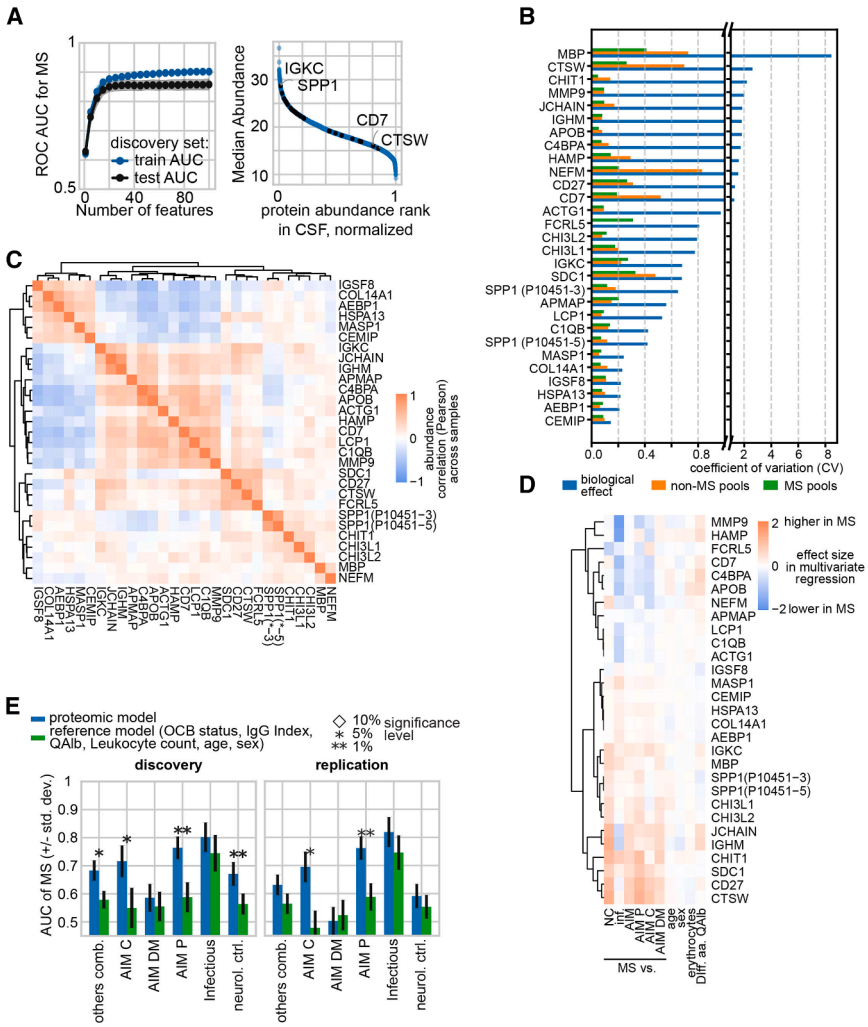


Figure S5. Determination and characterization of an MS biomarker panel using Astral data, related to Figure 5

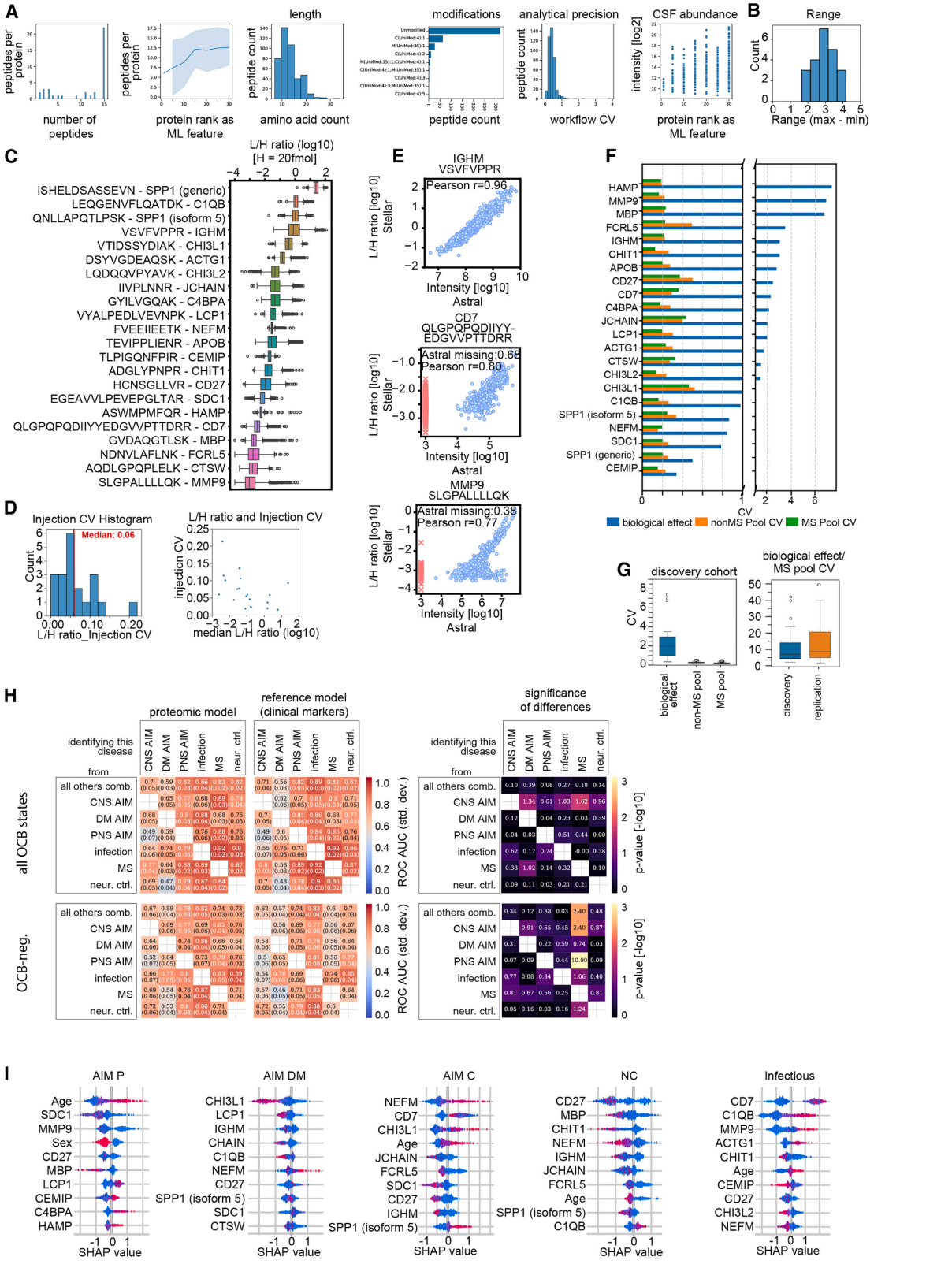
(A) Determination of the biomarker panel in the discovery cohort. ROC AUC for MS during feature selection in the discovery cohort (left). Location of the 29 proteins in the biomarker panel (black dots) across the CSF protein abundance spectrum (blue dots).

(B) Coefficients of variation (CVs) for the 29 biomarker panel proteins. CVs covering the variability of the entire workflow are shown for MS and non-MS pools that were measured on each 96-well plate. These CVs are contrasted with the biological effect across donor CSF samples, which was calculated equivalently. To reflect clinical reality, proteomic data without batch correction were used for the calculation.

(C) Abundance correlation across samples for proteins in the biomarker panel.

(D) Effect sizes in the multivariate regression model for proteins in the biomarker panel. Effect sizes are fold differences between MS and other diseases and estimators for covariate effects.

(E) Performance results for a control analysis, where training data are limited to OCB-negative samples. A biomarker panel of 30 features was determined equivalently to our standard analysis in which training data cover samples of all OCB states. Performance shown in this control pertains to ROC AUCs in OCB-negative samples. Note that the overall pattern in these samples is equivalent to our standard analysis, meaning that there is no particular benefit to limiting training to OCB-negative samples only. AUC for classifying MS when trained on OCB-negative samples only.



(legend on next page)

Figure S6. Characterization of targeted assay for biomarkers separating MS from inflammatory conditions, related to Figure 6

(A) Properties of peptides of the biomarker panel proteins that were synthesized as heavy isotope-labeled spike-in standards. Properties are based on Astral DIA measurements. Per protein, up to 15 protein-specific peptides were selected, which was possible for most proteins (first from the left), yet especially proteins contributing strongly to machine learning had fewer peptides (second), which were also typically low abundance (sixth). Length (third) and modification status (fourth) of selected peptides were compatible with synthesis. For the selection of peptides, those with the lowest workflow CV in Astral data were prioritized (fifth).

(B) Dynamic range of light-to-heavy ratios across donor CSF samples. Distribution for peptides in the final panel of the targeted assay. Summary statistic for (C).

(C) Distribution of light (L) to heavy (H) ratios across donor CSF for peptides in the final panel of the targeted assay. Light refers to the endogenous peptide and heavy to the isotope-labeled spike-in. For heavy peptides, 20 fmol was loaded.

(D) Coefficient of variation (CV) for repeated injections of the same sample loaded on different Evotips. Distribution (left) and dependence on the light-to-heavy ratio (right).

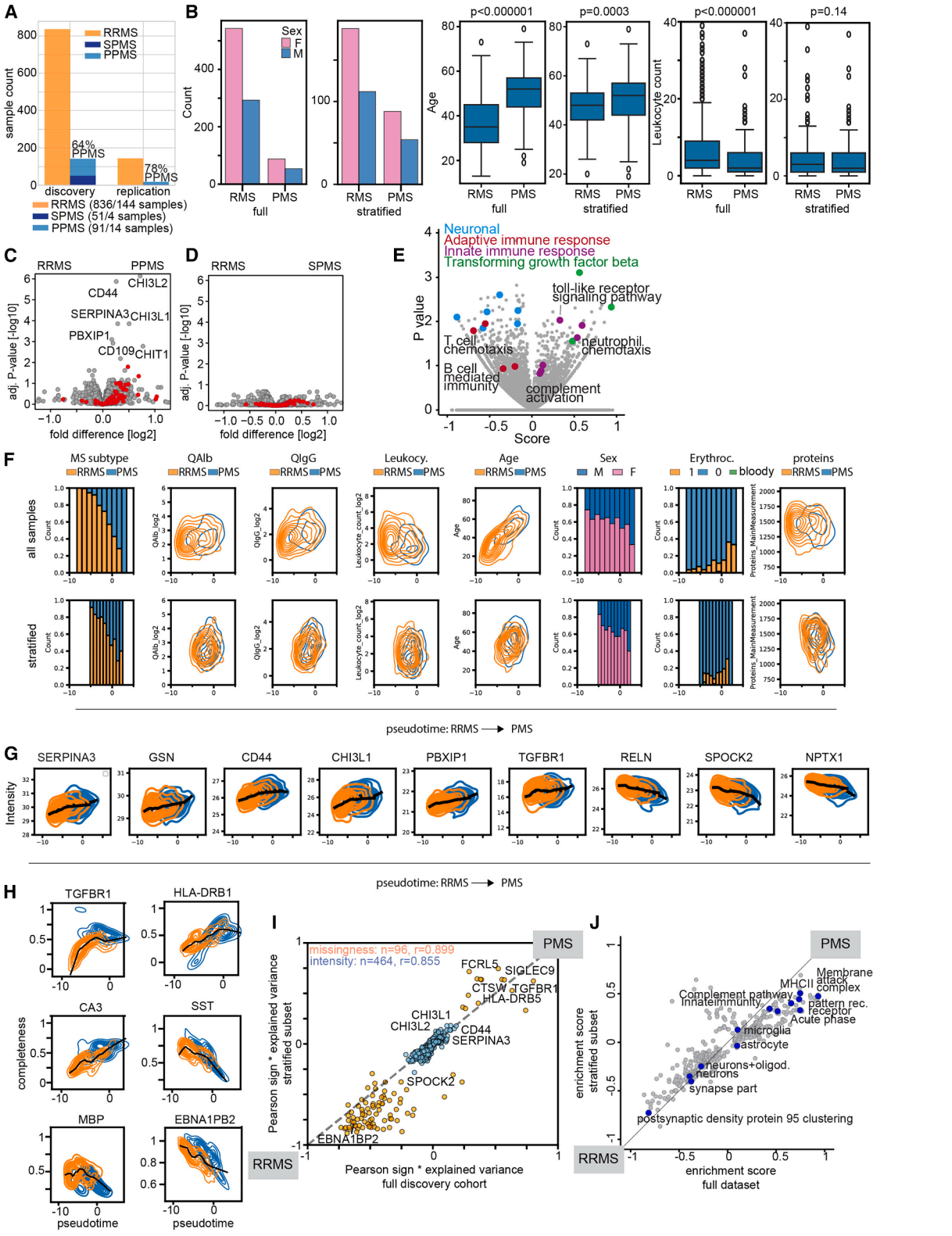
(E) Comparison of quantification between Astral DIA data (x) and the targeted assay measured on the Stellar (y) for high-abundance proteins (IGHM) and low-abundance proteins (CD7 and MMP9). In the Astral DIA data, there are missing values for proteins not observed in some samples. For visualization in this plot, these Astral quantities are imputed at a value of three and shown by a red cross. Calculation of the Pearson coefficient was based on the observed quantities only.

(F) CVs for the 22 final proteins included in the biomarker panel. Peptides are the same as in (C). The workflow CVs for both pooled MS and pooled non-MS samples are shown. Samples of these pools were measured across 96-well plates, along with donor CSF samples. To align with clinical requirements, quantities for CV calculation were not batch corrected. For comparison, the biological effect across donor CSF samples was calculated analogously to a CV.

(G) Distribution of CVs across the 22 final proteins in the discovery cohort (left). Signal-to-noise ratios calculated as the biological effect divided by the MS pool workflow CV (right).

(H) Performance comparison between the proteomic and the reference model, which uses current clinical markers. The ROC AUC (left) in the replication cohort is shown, within OCB-negative samples only (bottom) and across samples of all OCB states (top). One proteomic and one reference multi-class model were trained on the stratified subset of the discovery cohort and applied once to the replication cohort (STAR Methods). Subsets of the predictions were bootstrapped to calculate the standard deviations and the significance of performance differences between the proteomic and the reference models (right). Values are not identical across the diagonal because there are more than two outcomes in the multi-class model (STAR Methods). The vertical slices for MS correspond to the data shown in Figure 6A.

(I) Ten most influential features for the classification of the disease groups other than MS. Samples are datapoints; the intensity of the protein in the samples is color coded. The SHAP value is a metric for the contribution to decision-making by the model (x). Positive SHAP values mean that the protein predicts for that disease in the given sample, whereas negative values mean that the protein predicts against that disease.



(legend on next page)

Figure S7. MS subtype stratification and comparison of classification and differential abundance results, related to Figure 7

(A) Sample counts for MS disease course types in the discovery and replication cohorts.

(B) Covariate distributions for MS subtypes in the complete dataset and the subset stratified for age and leukocyte count. Stratification was implemented by down-sampling RRMS to 300 samples.

(C) Protein regulation between RRMS and PPMS in our discovery cohort. Immunoglobulins are shown in red.

(D) Protein regulation between RRMS and SPMS in our discovery cohort. Immunoglobulins are shown in red.

(E) Enrichment of annotation terms in SHAP values predicting for PMS (positive score) or for RRMS (negative score). Selected terms of interest are grouped and color coded by common processes.

(F) Distribution of PMS and RRMS samples, covariates, and the number of identified proteins along the pseudo-time between RRMS (low time value) and PMS (high time value). Data for the analysis are based on the full discovery cohort (top) or the subset stratified for age and leukocyte count (bottom). RRMS samples are enriched at negative pseudo-times and PMS samples at positive pseudo-times. The distribution of numerical variables is shown by density maps separately for RRMS samples (orange) and PMS samples (blue).

(G) Intensity profiles of selected proteins along the pseudo-time. The intensity distribution is shown by density maps separately for RRMS samples (orange) and PMS samples (blue). The protein profile averaged across all samples is shown by a black line.

(H) Detection rate profiles of selected proteins, analogous to (G).

(I) Concordance of protein profiles between the full dataset (x) and the stratified subset (y). Protein profiles are based on intensity (blue) or local detection rate (yellow). Profiles were pre-filtered for linearity by excluding those that had an absolute Pearson correlation of less than 0.8 with the pseudo-time in one of the datasets. The sign of the Pearson correlation was kept to reflect the directionality of association. Positive values reflect profiles that increase toward PMS, whereas negative values reflect profiles that increase toward RRMS. Profiles were further ranked by the variability that they explained for their given protein, which was quantified by the LOESS fit R^2.

(J) Annotation term enrichment in protein profiles. Data points are enrichment terms. Enrichment is shown for the full discovery cohort (x) and the stratified subset (y). Enrichment of terms was calculated in the protein profile linearity dimension, which itself was quantified as the Pearson coefficient between the protein profile and pseudo-time.