Proteomic Signatures of Conversion Risk and Disease Severity in Multiple Sclerosis

Laure Bastide[1,2], Virginie Imbault[3], Gaetano Perrotta[4], Serena Borrelli[1,5], Sophie Elands[1], Vincent van Pesch[6,7], Eva Borràs[8,9], Eduard Sabido[8,9], Nicolas Gaspard[1,10], David Communi[3] and Xavier Bisteau[3]

Abstract

Despite important advances in understanding the etiopathology of multiple sclerosis, factors determining disease progression remain partially understood and often difficult to predict. Specific diagnostic and prognostic biomarkers are needed to optimize the risk-benefit ratio of treatment for each patient. The aim of our study was to identify a cerebrospinal fluid proteomic signature associated with diagnosis and short- to mid-term prognosis across the multiple sclerosis continuum.

Our multicentric cohort study analyzed CSF samples from 120 patients using a proteomics data-independent acquisition strategy. Differentially expressed proteins were identified across diagnostic groups: 62 patients with multiple sclerosis, 15 patients with clinically isolated syndrome, and 43 healthy controls. We also compared the CSF of patients with no evidence of disease activity with those with disease activity at 2 and 5 years of follow-up. A diagnostic and prognostic classification model was built using iterative cross-validated logistic regression models on shared differentially expressed proteins across these two comparisons.

A total of 1,257 proteins were quantified, and 162 differentially expressed proteins were identified across comparisons. We identified a set of ten proteins associated with the diagnosis and prognosis of multiple sclerosis, including previously identified potential biomarkers (CH3L2, IGHG1, IGKC, LAMP2, ADA2), proteins known to be involved in the pathophysiology of multiple sclerosis (A0A8J8YUT9, AT2A2, CO3A1) and two yet unreported proteins (DSC2 and MMRN2). Multivariate models based on these proteins achieved good accuracy for the diagnosis of MS compared with CIS (area under the receiver operating characteristics curve [AUROC] up to 80% using 3 proteins) and prognosis (NEDA vs. EDA; AUROC up to 96% at 2 and 5 years; using 5 proteins).

These results, which will require further investigation to validate the new biomarkers, open new perspectives on multiple sclerosis pathophysiology and therapeutic targets.

Author affiliations:

1 Department of Neurologie, Hôpital Universitaire de Bruxelles, Université Libre de Bruxelles, Brussels, Belgium

2 Department of Neurology, CHU Nîmes, Université Montpellier, Nîmes, France

3 IRIBHM J.E. Dumont, Université Libre de Bruxelles ULB, Campus Erasme, Brussels, Belgium

4 Neuroinflammation Imaging Lab (NIL), Institute of NeuroScience, Université catholique de Louvain, Brussels, Belgium.

5 Laboratory of Neurochemistry, Institute of Neuroscience, Université Catholique de Louvain (UCLouvain), 1200 Brussels, Belgium

6 Department of Neurology, Cliniques Universitaires Saint-Luc, Université catholique de Louvain (UCLouvain), 1200 Brussels, Belgium

7 Department of Neurology, Hôpitaux Iris Sud, Brussels, Belgium

8 Proteomics Unit, Centre de Regulació Genòmica (CRG), Dr. Aiguader 88, 08003 Barcelona, Spain

9 Proteomics Unit, Universitat Pompeu Fabra (UPF), Dr. Aiguader 88, 08003 Barcelona, Spain

10 Neurology Department, Yale School of Medecine, New Haven, CT USA

Correspondance to: Laure Bastide,

Erasme Hôpital Universitaire de Bruxelles, Université Libre de Bruxelles,

Route de Lennik 808,

1070 Brussels,

Belgium;

E-mail: laure.bastide@chu-nimes.fr

and Xavier Bisteau,

IRIBHM J.E. Dumont, Université Libre de Bruxelles ULB, Campus Erasme,

Route de Lennik 808,

Brussels,

Belgium

E-mail: xavier.bisteau@ulb.be

Running title: New biomarkers for the diagnostic and prognostic of multiple sclerosis.

Keywords: Proteomics; multiple sclerosis; biomarkers signature; cerebrospinal fluid; SWATH-MS

Introduction

Multiple sclerosis is a complex and heterogeneous disease driven by two parallel pathogenic events, a progressive degenerative process superimposed by acute activities of inflammation leading to clinical and MRI relapses and accumulated disability.[1]

Recently, the classification of multiple sclerosis based on three discrete categories characterized by different clinical courses—relapsing-remitting (RRMS), primary progressive (PPMS), and secondary progressive (SPMS)—has evolved toward a spectrum of manifestations, starting with prodromal radiologically isolated syndrome (RIS), a first clinically isolated syndrome (CIS) and evolution to relapsing and progressive multiple sclerosis. This spectrum is thought to be underpinned by the variable expression since the beginning of these two parallel mechanisms, neurodegeneration and inflammation, which are also modulated by patient-specific factors, such as sex, age, social and environmental exposures, genetics, and disease duration.[2] The newly published 2024 revisions of the McDonald criteria allow the diagnosis of multiple sclerosis in patients presenting with RIS and have unified the framework for the diagnosis of relapsing and progressive forms.[3] Despite this important progress in understanding multiple sclerosis etiopathology, the course of the disease remains partially understood and poorly predictable. A RIS cohort study showed a cumulative probability of a first clinical event is only 51.2% at 10 years of follow-up.[4] Furthermore, according to several CIS cohort studies—prior to the revision of the McDonald criteria in 2017—17% to 25% of patients diagnosed with CIS with a normal brain MRI will develop multiple sclerosis within 5 to 20 years of follow-up and with highly variable severity.[5–9] At the same time, many disease-modifying therapies (DMT) have emerged over the past two decades and different lines of treatment are available depending on the aggressiveness of the disease.[10,11] Their effectiveness has been reported to be greater on reducing the course of the disease if they are initiated early during the disease progression as

soon as CIS or even RIS state, delaying the appearance of disability by years.[12–14] Specific biomarkers are still needed to better diagnose but also refine the prognosis and predict the disease progression, helping clinicians to optimize the risk-benefit of treatment to each patient.

Several studies have sought to identify biomarkers in blood or CSF for a clearer and earlier diagnosis of multiple sclerosis as well as to more accurately establish the prognosis. Multiple proteins have emerged as biomarkers, such as kappa free light chain (KFLC),[15–17] neurofilament light chain[18] or chitinase-3-like 1 protein (CH3L1).[19] Some of these are currently recommended for use in clinical routine but none are multiple sclerosis-specific.

Clinical proteomics has emerged as an important approach for biomarker discovery in MS, aiding in diagnosis, prognosis, elucidating etiopathogenesis, and monitoring treatment response. Although several technologies are used for proteomic analyses, liquid chromatography coupled with tandem mass spectrometry has established itself over the years as the method of choice for the identification and quantification of proteins.[20,21] Progress over the last decade in instrument sensitivity, speed of acquisition, and data analysis has enabled the accurate and sensitive identification and quantification of proteins with increasingly deep proteome coverage across large-scale sample sets. Among mass spectrometry strategies, data-independent acquisition (DIA) enables systematic fragmentation of all precursor ions across the full mass range, thereby reducing the stochasticity inherent to data-dependent acquisition (DDA) methods and improving quantitative reproducibility.[22] Sequential windowed acquisition of all theoretical fragment ion spectra (SWATH-MS) is one of the most widely implemented DIA approaches and has become an essential unbiased and consistently reproducible strategy predominantly used in clinical proteomics focused on neurological diseases; various protocols or tutorials have been developed.[23–35] This unbiased but highly specific protein identification proteomics approach makes SWATH-MS the ideal tool for

diseases as complex as multiple sclerosis, when searching for a unique biomarker, as it allows the identification and quantification of a protein signature.

In this study, we identified proteomic signatures for the diagnosis of multiple sclerosis and the prognosis stratification at the onset, short-term and mid-term using SWATH-MS data acquisition.

Materials and methods

Study design and population

The first cohort comprised 69 retrospectively identified patients (31 with multiple sclerosis, 10 with CIS, and 28 controls [CTL]) with leftover CSF samples collected between July 2014 and June 2017 and archived in the Hôpital Universitaire de Bruxelles (HUB) Neurologie/Neurochirurgie biobank (Reference B2014/001). The second cohort comprised 51 prospectively identified patients (31 with multiple sclerosis, 5 with CIS and 15 CTL) with additional CSF collection for research purposes. Samples were collected between March 2017 and December 2022 from patients enrolled in the study when evaluated for suspected multiple sclerosis in one of the three participating departments of neurology in Brussels, Belgium: Departments of Neurology of HUB, Centre Hospitalier Universitaire Brugmann (CHUB), and Cliniques Universitaires de Saint-Luc (CUSL).

The inclusion criteria were as follows: (i) age > 18 years, (ii) availability of lumbar puncture and MRI data, and (iii) ability to provide informed consent.

Patients fulfilling the 2017 McDonald criteria[36] were designated as having MS or CIS if they did not fulfil the McDonald 2017 criteria for dissemination in space or in time and had no conversion to MS at the end of follow-up; they were excluded if they had another neurological disease diagnosis. All individuals designated as having CTL had a diagnosis of a non-inflammatory neurological disease, CSF oligoclonal bands (OCBs) were absent, and CSF white blood cell count (WCC) was < 3/mm^3. The clinical characteristics of the CTL group are presented in Supplementary Table S1.

Clinical follow-up and variables

At lumbar puncture (T0) patients underwent a standardized diagnostic evaluation including the Expanded Disability Status Scale (EDSS), brain and spinal MRIs, followed by regular clinical and MRI follow-up.

Time to clinical or MRI relapse (TTR) was defined as the time elapsed from treatment initiation to the occurrence of the first clinically confirmed relapse or MRI relapse, defined as a new T2 lesion on spinal or brain MRI.

Annualized clinical or MRI relapse rate (ARR) calculated as the ratio of the previously described TTR to the follow-up in years.

No Evidence of Disease Activity (NEDA) was defined as follows: (1) no clinical relapse, (2) no disability worsening (based on EDSS progression) and (3) no MRI activity (new T2 lesion or Gd+ lesion or enlarging T2 lesion), according to the NEDA-3 criteria[37] at 2 years and 5 years after the initiation of treatment in MS patients or the relapse onset for CIS patients.

The age-related MS severity score (ARMSS) was calculated by adjusting the EDSS scores for the patient's age at the time of assessment. Factoring age in severity assessment allows cross-sectional analysis of patient-level disability data.[38]

Descriptive statistics were performed on the demographic and clinical variables; two-sided Fisher's exact test or chi-squared test were used for contingency tables or two-sided Mann–Whitney U-test and Kruskal–Wallis' test were used for continuous variables, as appropriate. A correction for multiple comparisons was performed using the Benjamini–Hochberg method (*FDR*-value).

Ethics statement

This study was reviewed and approved by the Ethics Committee of Erasme Hopital Universitaire de Bruxelles under reference B406201733112-P2017/417. The participants provided written informed consent to participate in the prospective study.

Sample collection and handling

CSF samples were collected during the initial diagnostic assessment before intravenous steroid administration if needed, when possible, or any DMT. Retrospective leftover CSF samples were stored at -20°C whereas prospective CSF samples were processed according to published guidelines.[39]

CSFs were thawed and concentrated using a centrifugation column (Centricon 3kDa, Millipore Merck) at 4°C and 14.000g until 25 to 30µL.

The depletion of the top 14 most abundant proteins, corresponding to albumin and immunoglobulins, was performed using affinity columns (Pierce Top 14 Abundant Protein Depletion Spin Columns; Thermo Scientific), as described by the manufacturer. Protein concentration in the samples was measured after depletion using the Folin method.[40] Proteins were diluted in 25mM ammonium carbonate and dithiothreitol was added to obtain a final concentration of 25mM. Each sample was incubated for 1 h at 4°C. Alkylation of samples was performed by the addition of iodoacetamide with final concentration of 71mM before incubation for 1h at 4°C. All samples were diluted 4 times with acetone and incubated à -20°C during 1h for protein precipitation. After centrifugation at 13.000 rpm for 20 min at 4°C, acetone was removed and 40 ngr of trypsin (Gold Mass Spectrometry Grade V5280 Promega, USA WI) was added. Samples were incubated and shaked at 37°C during 30 min and then over night without shaking. Samples were incubated and shaked at 37°C during 30min and then over night without shaking.

Finally, the pH of each sample was adjusted to 2 with a 10% formic acid (FA) solution. Samples were then cleaned up using a 1 mL high-lipophilic balance column equilibrated with acetonitrile (ACN) 5%/0.1% FA, eluted with ACN 80%/0.1% FA, dried in a vacuum concentrator, and resuspended in buffer A (H2O 100%/HCOOH 0.1%).

Data dependent analysis

Proteins were identified from CSF samples using label-free DDA mass spectrometry (TripleTOF 5600 SWATH, AB Sciex) coupled with micro-high-performance liquid chromatography (µLC 425, Eksigent). Peptide separation was performed by applying a 90 min gradient hydrophobic separation on a column (ChromXP C18 CL, 150 mm x 0.3 mm, 3 µM, 120 A, Sciex) using a two-step ACN gradient with FA: 5–25% ACN / 0.1% FA in 48 min then 25%–60% ACN / 0.1% FA in 20 min and were sprayed online in the mass spectrometer. Subsequently, peptides were ionized by electrospray ionization and injected into the TripleTOF 5600 mass spectrometer (TT 5600, AB Sciex). The 20 most intense precursors with charge states of 2 to 4 were selected for fragmentation. MS1 spectra were collected in the range of 400–1250 m/z with an accumulation time of 250 ms, and MS2 spectra were collected in the range of 100–2000 m/z for 100 ms.

Generation of spectral libraries

The generated .WIFF files were directly analysed by ProteinPilot software (AB Sciex) and the recorded spectra were matched to peptides by the search algorithm Paragon (AB Sciex) using the database human Swiss-Prot part of UniProtKB.[41] Trypsin was selected as digestion enzyme. Oxidation at methionine was set as a dynamic modification and carbamidomethylation as a static modification at cysteine. A decoy database search was performed with a target false discovery rate of < 1%.

Data independent analysis

1 µg of peptides was injected using SWATH-MS acquisition on a Triple TOF 5600 mass spectrometer (Sciex, Concord, Canada) interfaced to an Eksigent NanoLC Ultra 2D HPLC System (Eksignet, Dublin, CA). Peptides were injected on a separation column (Eksigent ChromXP C18 CL, 150 mm x 0.3 mm, 3 µM, 120 A, Sciex) using a two steps acetonitrile gradient (5%–25% ACN / 0.1% HCOOH in 48 min then 25%–60% ACN / 0.1% HCOOH in 20 min) and were sprayed online into the mass spectrometer. Swath acquisitions were performed using 42 windows of a fixed effective isolation width to cover a mass range of 200–1500 m/z. SWATH MS2 spectra were collected from 200–1500 m/z. The collision energy for each window was determined according to the calculation for a charge 2+ ion centered upon the window with a spread of 15. An accumulation time of 76 ms was used for all fragment-ion scans in high-sensitivity mode and for the survey scans in high-resolution mode acquired at the beginning of each cycle, resulting in a duty cycle of ~3.3 s.

Raw data were converted to mzML format with MSConvert (Version 3.0) and their processing was carried out with DIA-NN[42] (version 1.8.2 beta 27) with search parameters set as follows: precursor *FDR* 1%; mass accuracy at MS1 and MS2 both set to 0; scan window set to 0; isotopologues and MBR turned on; protein inference at gene level; heuristic protein inference enabled; quantification strategy set to QuantUMS (high accuracy); neural network classifier double-pass mode; cross-run normalization RT-dependent. A universal library was used, and protein re-annotation was performed. Spectra were searched against a prior *in loco* generated library containing 9972 proteins covered by 174348550 peptides including for common contaminants.[43]

To generate the library, DDA and DIA data acquired on TT5600 were used to generate one universal library by FragPipe[44] (version 18.0) using a pre-defined workflow

DIA_SpecLib_Quant. Specifically, decoys were first added to the FASTA which contains human protein sequences (UP000005640, 78986 reviewed and unreviewed entries, downloaded from UniProtKB on 07th June 2022). Then, MSFragger (version 3.5) was used to search DDA raw data, with the following settings: precursor and fragment mass tolerance 30 ppm; strict trypsin with no more than five missed cleavages; peptide length 6–50; peptide mass 500–5000; C + 57.021464 as fixed modification; M + 15.9949, N-term +42.0106, nQ -17.0265, nE -18.0106, DN +0.984016 and carbamylation +43.005814 as variable modifications; min matched fragments 4; max fragment charge 2. In the validation step, MSBooster was implemented on both spectra and RT levels, and then Percolator and ProteinProphet integrated in Philosopher (version 4.4.0) were used for PSM validation and protein inference. Library generation was conducted using EasyPQP (version 0.1.50) with RT calibration based on ciRTs and Lowess fraction set to 0.04. Only fragment types b and y were included with a tolerance of 15 ppm with a max delta_unimod.

Differential expression analysis

MSstats (version 4.16.1)[45] in R (version 4.5.1)[46] was used to perform the data processing, model-based estimation of the quantity of each protein based on a relative $\log_2$ transformation, and group comparison following the workflow described previously.[47] All samples underwent median normalization. Only the top 100 features were selected for protein summarization, uninformative features and outliers were removed. No specific peptides were excluded, and proteins with only one feature or with features present in less than three measurements per run were excluded. No missing value imputation was performed. For the group comparison analysis, we selected proteins present in at least 90% of the runs in at least one group of comparisons. The group comparison was performed between MS vs CTL, MSCIS vs CTL, CIS vs CTL, MS vs CIS, EDA vs NEDA at 2 and 5 years of follow-up. A

correction for multiple comparisons was performed using the Benjamini–Hochberg method, and a threshold for adjusted *P-value* < 0.05 referred to as false discovery rate (*FDR*) in the figures or table was deemed significant.

Diagnostic and prognostic classification models

A four-fold cross-validation was performed to identify the most discriminative combination of proteins, as previously described.[48] Patient data were divided into four equal subgroups in the training set. For each group, each protein was fitted on a logistic regression model between the two groups, and its classification ability was evaluated using the area under the receiver operating characteristics (AUROC) curve. The most discriminative protein was selected as the first classifier. Additional discriminative proteins were repeatedly added while increasing AUROC values. The proteins selected at least twice within the four cross-validation steps in the training set were chosen as the characteristic classification signature for that training set. The best classification signature for each training set was fitted on a logistic regression model and applied to the validation set. The procedure from dividing the data into training and validation sets to fitting the logistic model with the best classification signature was repeated 500 times to assess the reproducibility of the classification ability. A final consensus model was comprised of the combination of proteins that were selected most frequently in the 500 repeats. To obtain the upper level for the predictive accuracy of the selected consensus proteins, the final model was fitted to the full dataset, and the predictive accuracy was quantified using the AUROC, sensitivity, specificity, and accuracy. The estimate of variability associated with the receiver operating characteristics (ROC) curve was obtained by plotting the 25th and the 75th quantile of the sensitivities for each value of 1-specificity obtained in the validation set over all the iterations for which each protein combination was selected. The

pROC package in R was used to draw ROCs, calculate AUROCs, and other performance measures (*i.e.*, sensitivity, specificity, and accuracy).

For each classification task, we identified the top three most frequently selected protein combinations, representing the most robust predictors across resampling. ROC curves of the top three selected combinations in each classification case are graphed (solid lines), representing performance on the full cohort and reported with their AUROC and confidence interval (CI) at 95% among the entire cohort. Dotted lines show the distribution of the ROC curves across the 500 validation iterations. To evaluate whether CSF proteomic signatures could predict short-and mid-term inflammatory activity (NEDA vs. EDA at 2 and 5 years) in multiple sclerosis, we applied the same iterative, cross-validated logistic regression framework used for diagnostic analyses.

To assess whether simpler protein signatures could retain strong discriminative power, we evaluated the performance of multivariate logistic models constructed using the unique proteins contributing to the top three combinations of each comparison. These models were evaluated on the entire cohort using bootstrap resampling and are depicted by the purple solid line.

Results

Differential clinical profiles of CIS and MS groups inform biomarker identification strategy

A total of 187 samples were initially considered. As outlined in Supplementary data Fig. S1, three samples failed the initial quality control, and 184 samples were processed using mass spectrometry. The second quality filter based on protein identification depth and signal intensity led to the exclusion of an additional 64 samples. Ultimately, 120 samples fulfilled all analytical requirements and were retained for biostatistical analyses, comprising 69 retrospective samples and 51 prospective samples (17 from HUB, 14 from CHUB, and 20 from CUSL). Of those 62 had MS, 15 had CIS and 43 were CTL.

The clinical characteristics of MS, CIS, and CTL patients at last follow-up are presented in Table 1 and statistics are reported in Supplementary File S1.

MS patients were significantly older than CTL individuals (30 [14–80] vs. 38 [20–82], *FDR*-value = 0.025). CSF OCBs were more frequent in MS than in CIS patients (95% vs. 60%, *FDR*-value < 0.001). Clinical onset differed substantially: optic neuritis being the presenting syndrome in 73% of CIS patients versus 15% of MS patients (*FDR*- value < 0.001).

Over half of the CIS patients (53%) had no brain lesions at baseline, whereas all MS patients presented with at least one lesion. None of the CIS patients exhibited more than 10 brain lesions, in contrast to 48% of the MS patients. Seventy seven percent of the MS patients presented with at least one spinal lesion on MRI, whereas this was observed in only 20% of the CIS patients. Although not statistically significant, MS patients tended to show a higher proportion of gadolinium-enhancing lesions (60% vs 27%, *FDR*-value = 0.067). Almost no CIS patients had received DMT at baseline. Markers of inflammatory disease activity,

including the number of clinical relapses and annualized relapse rate (ARR) were all higher in MS than in CIS (0 [0–2] vs. 0, *FDR*-value = 0.020 and 0.18 [0–2.35] vs 0, *FDR*-value < 0.001 respectively). Measures of disability progression (EDSS) showed limited discriminatory power, but the ARMSS score was significantly higher in MS patients (3 [0.3–8.9] vs. 1 [0.4–5], *FDR*-value = 0.030). Forty five percent of MS patients achieved NEDA at 2 years and 11% at 5 years.

CSF proteomic differences reflect disease stage and activity across the CIS–multiple sclerosis continuum

We performed an in-depth CSF proteomic analysis combining DDA- and DIA/SWATH-MS on 120 clinically well-characterized samples. We quantified 1,257 proteins, of which 593 were selected for downstream analysis, as described in the Methods section.

Considering multiple sclerosis as a continuum that begins prior to the CIS stage, we structured our analyses to compare the combined MS and CIS group (MSCIS) against CTL, complemented by separate comparisons of MS vs CTL and CIS vs CTL to identify potential diagnostic biomarkers. Building on the clinical differences distinguishing CIS from MS, as well as the NEDA outcomes, we also searched for differentially expressed proteins associated with prognosis at disease onset and at short- and mid-term follow-up. For this purpose, differential expression analysis was performed between MS and CIS patients, and within the overall MSCIS cohort according to NEDA status at 2 and 5 years as illustrated in Figure 1 and reported in Suppelmentary File S2.

Differential expression analysis revealed distinct proteomic signatures between MS, CIS, CTL, NEDA vs. EDA at 2 and 5 years. In total, 53 proteins were differentially expressed between MS and CTL, 32 between MSCIS and CTL, 13 between CIS and CTL and 4 between MS and CIS (Fig. 2A). The largest foldchanges were observed between MS and CTL. No

differentially expressed protein was common to the three comparisons of a single categories (MS vs CTL, CIS vs CTL and MS vs CIS), only one overlapped between the MS vs CIS and CIS vs CTL comparisons, two between MS vs CTL and MS vs CIS and seven (12%) between MS vs CTL and CIS vs CTL (Fig. 2B). In contrast, as many as 18 differentially expressed proteins overlapped between comparison of disease activity (NEDA status) at 2 and 5 years (Fig. 2B).

To visualize the relationships of these protein signatures across all comparisons, heatmaps in the Figure 2C displayed how the proteins identified in each comparison behave in others comparisons, revealing clusters of proteins that are tightly associated with specific disease states or progression profiles (Fig. 2C).

Altogether, the proteomic analysis of CSF from MS and CIS patients compared with CTL individuals revealed diagnostic specificity, minimal overlap at early disease stages, and greater prognostic consistency across the CIS–multiple sclerosis continuum.

Identification of integrated biomarkers correlating with disease progression and initial spinal cord pathology.

To better understand which molecular changes are shared between early multiple sclerosis stages and mid-term disease activity, we examined the overlap between proteins identified in the diagnostic comparisons (MS, CIS, CTL) and those associated with disease activity (NEDA vs EDA at 2 and 5 years). This revealed 15 proteins common to both diagnostic and prognostic contrasts (Fig. 3A). These proteins represent a core set of candidates that are not only dysregulated early in the CIS-multiple sclerosis continuum but also remain associated with short and mid-term disease activity, making them suitable candidates for further investigation.

Among these 15 shared proteins summarized in the Table 2, several have already been described in the literature as potential diagnostic or prognostic biomarkers in multiple sclerosis. These include A0A7P0Z497,[49] ADA2,[50,51] CH3L2,[52] CALR[53,54] and LAMP2.[55] Other proteins in this group have recognized roles in multiple sclerosis pathophysiology: for example, A0A8J8YUT9 belongs to a receptor family involved in immune-mediated neuroprotective interactions,[56] and CO3A1 is strongly induced in both active and inactive multiple sclerosis brain lesions.[57] AT2A2, another protein in this set, is known to be modified by dimethyl fumarate (DMF), a commonly used DMT in multiple sclerosis.[58] We also confirmed dysregulation of classic humoral immune markers such as IGHG1 and IGKC consistent with intrathecal B-cell activation and KFLC which is a fragment of IGKC, already used in the clinical routine and observed in multiple sclerosis. Importantly, we identified five proteins (DSC2, IRPL1, MMRN2, PCOC1, TSP2) for which no evidence has yet been reported in the multiple sclerosis literature, highlighting novel candidates that may contribute to early disease mechanisms or progression.

We next evaluated whether the abundance of these proteins correlated with baseline clinical features, including age, sex, clinical onset, MRI T2 lesion burden in the brain and spinal cord, gadolinium enhancement and OCB presence. CH3L2 was significantly increased in MSCIS patients presenting at least one spinal lesion at baseline (9.061 [8.763–9.299] vs. 8.501 [8.262–8.841], *FDR* = 0.004; Fig. 3B), suggesting a link with early spinal cord involvement.

We then assessed the association of these proteins with markers of disease severity and progression. Proteins were examined against the EDSS and ARMSS scores at the end of the follow-up, and the disease inflammatory activity with TTR, ARR, and NEDA status at 2 and 5 years. CALR levels were decreased in patients with a higher-than-median ARMSS score (9.539 [9.247–10.044] vs. 10.044 [9.886–10.309], *FDR* = 0.005; Fig. 3C), while IRPL1 and LAMP2 were elevated in the same subgroup (11.928 [11.553–12.409] vs. 11.195 [10.646–

12.108], *FDR* = 0.048 and 9.1 [8.578–9.437] vs. 8.138 [7.828–8.825], *FDR* = 0.005, respectively), suggesting potential links with disability progression.

In addition, CH3L2, CO3A1, IGHG1, IGKC and LAMP2 were all higher in patients who experienced disease activity at 5 years, compared with those who maintained NEDA (respectively 8.966 [8.768–9.245] vs. 8.501 [8.217–8.693], *FDR*-value = 0.024, 9.340 [9.096–9.542] vs. 9.076 [8.892–9.105], *FDR*-value = 0.024, 11.172 [10.208–11.961] vs. 9.221 [8.501–10.406], *FDR*-value = 0.024, 10.714 [9.768–11.156] vs. 9.457 [8.740–9.753], *FDR*-value = 0.032, 8.928 [8.027–9.366] vs. 8.028 [7.844–8.210], *FDR*-value = 0.024; Fig. 3D). These findings underscore the prognostic relevance of this protein subset and its potential role in predicting long-term disease activity.

Iterative cross-validated regression reveals highly discriminative CSF proteomic signatures along the CIS–multiple sclerosis continuum

To evaluate the diagnostic potential of the 15 proteins shared between diagnostic and prognostic comparisons, we applied the logistic regression framework combined with a fourfold cross-validation, repeated over 500 iterations (see the Methods section for further description).

For the MSCIS vs CTL classification (Fig. 4A), CH3L2 emerged as the most recurrently selected protein (113 selections), achieving an AUROC of 0.762 ($P = 8\times10^{\square\square}$) (Fig. 4A). The combination of CH3L2 and MMRN2 was selected 89 times and performed better (AUROC 0.809, $P = 4\times10^{\square\square}$). ADA2, selected 75 times, also showed strong discriminative ability (AUROC 0.748, $P = 1\times10^{\square\square}$). The results were similar when focusing on MS vs CTL (Fig. 4B). The most frequently selected combination, CH3L2 and MMRN2 (126 selections),

achieved an AUROC of 0.832 (P = 2×10□□). CH3L2 alone (96 selections) reached an AUROC of 0.783 (P = 5×10□□), while the combination CH3L2 + IGKC (58 selections) yielded an AUROC of 0.831 (P = 9×10□□). These results highlight CH3L2 as a central diagnostic protein across early multiple sclerosis phenotypes.

For the CIS vs CTL comparison (Fig. 5A), the combination CH3L2 + LAMP2 (44 selections) provided an AUROC of 0.824 (P = 0.0003). The combination ADA2 + AT2A2 (41 selections) slightly outperformed it (AUROC 0.837, P = 0.0002), while A0A8J8YUT9 + ADA2 (30 selections) also performed well (AUROC 0.816, P = 0.0004). These combinations suggest that CIS, although clinically milder, already harbors detectable proteomic deviations from CTL.

Finally, the MS vs CIS classification (Fig. 5B) revealed AT2A2 as the most robust discriminator (225 selections, AUROC 0.765, P = 0.004). The combinations AT2A2 + A0A8J8YUT9 (84 selections, AUROC 0.802, P = 0.0009) and AT2A2 + IGKC (52 selections, AUROC 0.782, P = 0.002) further improved classification, underscoring AT2A2 as a key protein distinguishing early MS from CIS.

To assess whether simpler protein signatures could retain strong discriminative power, we next evaluated the performance of multivariate logistic models constructed using the unique proteins contributing to the top three combinations of each comparison. These models were evaluated on the entire cohort using bootstrap resampling (Figures 4 and 5).

The multi-protein diagnostic signatures demonstrated strong and consistent performance across all comparisons. For distinguishing MSCIS from CTL (Fig. 4A), the three-protein combination CH3L2, MMRN2 and ADA2 achieved an AUROC of 0.826 (P = 1×10□□). In the MS vs CTL classification (Fig. 4B), the signature composed of CH3L2, MMRN2 and IGKC performed even better, reaching an AUROC of 0.863 (P = 1×10□□). The highest

diagnostic accuracy was observed for CIS vs CTL, where a five-protein panel (CH3L2, LAMP2, ADA2, AT2A2 and A0A8J8YUT9) yielded an AUROC of 0.898 ($P = 2\times10^{\square\square}$). Finally, the ability to discriminate MS from CIS (Fig. 5A) was supported by the combination AT2A2, A0A8J8YUT9 and IGKC, which reached an AUROC of 0.808 ($P = 0.0007$).

Collectively, these results demonstrate that multiple protein combinations, particularly those involving CH3L2, MMRN2, ADA2, IGKC, AT2A2 and LAMP2, provide strong and reproducible diagnostic power across the MS-CIS-CTL comparisons. These findings highlight a set of highly informative proteins capable of differentiating early multiple sclerosis phenotypes and controls with high accuracy.

Proteomic profiling of CSF reveals signatures that predict short- and mid-term inflammatory activity in multiple sclerosis

Focusing on comparisons between patients with EDA and those achieving NEDA, we identified a restricted set of proteins that consistently contributed to prognostic discrimination at both 2 and 5 years of follow-up (Fig. 6).

At 2 years (Fig. 6A), the three most frequently selected protein combinations across 500 cross-validation iterations were DSC2 + IGHG1 (46 occurrences), which achieved the highest performance with an AUROC of 0.960 ($P = 2\times10^{\square\square}$), followed by CO3A1 + LAMP2 (31 occurrences, AUROC 0.931, $P = 3\times10^{\square\square}$), and CH3L2 + LAMP2 (30 occurrences, AUROC 0.897, $P = 8\times10^{\square\square}$). These results reveal a strong and coherent prognostic signal, with multiple combinations converging on proteins associated with immune activity, extracellular matrix remodeling and inflammatory signaling.

At 5 years, the two leading combinations were identical to those identified at 2 years; DSC2 + IGHG1 (40 occurrences, AUROC 0.960, $P = 2\times10^{\square\square}$) and CO3A1 + LAMP2 (37 occurrences, AUROC 0.931, $P = 3\times10^{\square\square}$), demonstrating remarkable longitudinal stability of the prognostic proteomic substrate. The third-ranked combination differed only slightly, with CH3L2 + DSC2 emerging as a robust discriminator (29 occurrences, AUROC 0.956, $P = 3\times10^{\square\square}$).

To further evaluate the prognostic value of these markers, we aggregated the unique proteins from the top combinations at 2 and 5 years and generated a multivariate logistic regression model. The signature combining DSC2, IGHG1, CO3A1, LAMP2 and CH3L2, yielded an AUROC of 1.0 for discriminating between EDA and NEDA in the full dataset at 2 and 5 years.

Discussion

In this study, we took advantage of the strengths of the SWATH-MS approach—its reproducible quantification and broad proteomic coverage—to analyze 120 CSF samples and identify predictive proteomic signatures associated with diagnosis, onset and both short- and mid-term prognosis in early multiple sclerosis. Across all comparisons, we quantified 1,257 proteins and identified 162 differentially expressed proteins.

The comparison between MS and CTL showed the higher number and amplitude of changes in diagnostic comparisons, consistent with a more pronounced inflammatory and neurodegenerative substrate in established multiple sclerosis. Despite the biological overlap between CIS and multiple sclerosis (heatmap Fig. 2C), very few DEPs were shared across diagnostic categories. This limited overlap underscores the heterogeneity and stage-specific molecular alterations across early demyelinating presentations and highlights the potential for proteins to reflect distinct biological processes in CIS versus MS groups.

We observed a higher overlap when comparing EDA vs NEDA at 2 and 5 years. This suggests that biomarkers associated with inflammatory activity and clinical or MRI progression may be more stable and consistently detectable over time than those discriminating diagnostic categories.

Distinct expression patterns emerge for multiple sclerosis-specific markers, CIS-driven proteins, and those linked to disease activity, supporting the idea that different biological axes (diagnosis vs prognosis) rely on partially non-overlapping molecular pathways.

Among the identified differentially expressed proteins, 15 were common to diagnostic and prognostic group comparisons. Within these proteins, CH3L2 was significantly elevated in patients with at least baseline spinal T2 lesions, whereas LAMP2 was higher in patients with

greater disability progression. When assessing the diagnostic and prognostic potential of these 15 shared proteins, 10 proteins were identified with strong discriminative ability across comparisons. ADA2, AT2A2, A0A8J8YUT9, CH3L2, CO3A1, IGHG1, IGKC and LAMP2 contributed most consistently to diagnostic classification, whereas DSC2, IGHG1, CO3A1, LAMP2 and CH3L2—alone or in combination—demonstrated excellent discrimination between EDA and NEDA at both time points. As expected, such performance reflects an overfitting inherent to the aggregation of all top-ranked proteins and highlights the need for external validation. Nonetheless, the reproducibility of DSC2, IGHG1, CO3A1 and LAMP2 across complementary analyses and time points strongly supports their relevance as candidate prognostic biomarkers. This recurrence could also underscore their shared involvement in the biological processes associated with disease activity.

Chitinase-3-like protein 2 (CH3L2), is a well-documented biomarker for MS. Expressed by astrocytes and microglia, CH3L2 has been associated with the conversion of CIS to multiple sclerosis, shorter relapse-free intervals and disability progression.[52,59,60] In our analysis, CH3L2 appeared in both diagnostic signatures (MSCIS vs CTL, MS vs CTL, CIS vs CTL) and prognostic models differentiating EDA from NEDA-3 at T2 and T5, in agreement with existing literature, although we did not detect significant differences between MS and CIS.

Immunoglobulins are central to multiple sclerosis pathophysiology and are established biomarkers, notably through OCBs[36] and KFLC.[3,15–17,61,62] As KFLC derives from IGKC, its presence in our diagnostic and prognostic signatures aligns with previous findings. IGHG1, the constant region of IgG1, has recently been implicated in multiple sclerosis pathogenesis: plasma IgG1 aggregates induce complement-dependent apoptosis in neurons and astrocytes and discriminate multiple sclerosis from healthy control (HC) or other neurological disease controls with high accuracy (AUC 0.93).[63] IgG1 also differentiates SPMS from RRMS and PPMS (AUC 0.91–0.92).[64] IGHG1 and IGKC genes were also highly expressed in all lesion

types in the brain white matter of patients with progressive multiple sclerosis.[65] These data strongly support the biological plausibility of our IGHG1-based prognostic models.

Lysosome-associated membrane glycoprotein 2 (LAMP2) plays a critical role in autophagy, a pathway implicated in multiple sclerosis pathogenesis.[66] A previous study reported elevated serum LAMP2 in SPMS vs RRMS, though CSF LAMP2 lacked diagnostic utility for distinguishing multiple sclerosis from HC.[55] In our cohort, CSF LAMP2 levels were higher in patients with ARMSS progression and in those experiencing EDA at 5 years; LAMP2 also contributed to the top prognostic signatures for EDA vs NEDA-3 at both T2 and T5. Another study investigating iron metabolism disorders in multiple sclerosis identified LAMP2 within a multi-gene predictive model, showing that its expression in both blood and brain samples helped distinguish MS patients from controls.[67] In our cohort, CSF LAMP2 did not contribute to the classification of MS or MSCIS versus CTL, although it appeared among the top three protein combinations distinguishing CIS from CTL. Notably, a third publication examining X-linked genetic factors associated with multiple sclerosis susceptibility reported 20 SNP-linked genes that were significantly more abundant in MS patients, and LAMP2 was among them.[68] Consistent with this, LAMP2 up-regulation in MS group was also described in pediatric multiple sclerosis.[69]

Adenosine deaminase 2 (ADA2) regulates extracellular adenosine, a molecule involved in neurotransmission, inflammation, regeneration, and blood brain barrier (BBB) permeability.[70] Of interest, the purine nucleoside analogue 2-chlorodeoxyadenosine (cladribine), a treatment for active multiple sclerosis, is resistant to deamination by ADA. As a result, its nucleotide derivatives accumulate in lymphocytes and induce CD4□, CD8□ and CD19□ cell death.[71,72] In line with previous studies, ADA2 was differentially expressed between MS and CTL groups and could discriminate them (AUROC 0.748).[50,51]

For several proteins identified in our signatures, clinical biomarker data are lacking, but mechanistic studies support their involvement in multiple sclerosis. A0A8J8YUT9 is an isoform of NTRK2 which interacts with is involved in CNS homeostasis and multiple sclerosis pathogenesis.[56,73] CO3A1, encoding collagen alpha-1 (III) chain, is strongly induced in active and inactive multiple sclerosis lesions.[57,65] AT2A2 regulates cytosolic calcium via SERCA2 activity and modulates T and B cell receptor signaling; its modification by DMF in experimental autoimmune encephalomyelitis mouse models suggests functional relevance.[58]

Finally, our work describes two proteins not previously associated with multiple sclerosis: desmocollin-2 (DSC2) and multimerin-2 (MMRN2). DSC2 is a desmosomal adhesion protein implicated in congenital heart disease[74] and several cancers.[75,76] Notably, DSC2 was recently identified as a dominant entry receptor for Epstein-Barr virus (EBV) infection of epithelial cells,[77] a finding of interest given the strong evidence linking EBV to multiple sclerosis development.[78,79] MMRN2, an extracellular matrix glycoprotein involved in vascular stability and angiogenesis may be relevant in multiple sclerosis given the prominent role of BBB dysfunction in demyelinating lesions.[80–82]

Our study has limitations, including mixed retrospective and prospective sampling, heterogeneous pre-analytical handling, a limited number of patients with CIS, and the absence of an external validation cohort. To mitigate these issues, we applied volume and peptide normalization prior to mass spectrometry acquisition, selected samples based on protein depth and intensity thresholds, and restricted downstream analyses to proteins detected in at least 90% of runs within a comparison group and we employed cross-validation to strengthen internal robustness. Furthermore, the multicenter nature of the study increased its generalizability.

In summary, our study confirms several proteins as reliable diagnostic and prognostic biomarkers in multiple sclerosis (CH3L2, IGHG1, IGKC, LAMP2, ADA2), identifies

additional candidate biomarkers supported by pathophysiological evidence (A0A8J8YUT9, AT2A2, CO3A1), and introduces DSC2 and MMRN2 as novel proteins of interest. Although further validation is required, these findings provide new insights into multiple sclerosis biology and highlight promising avenues for biomarker development and therapeutic exploration.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[83] partner repository with the dataset identifier PXD073146.

Acknowledgements

We would like to thank Dr. Serena Borelli and the entire team at the CHUB Neurology Department, as well as Prof. Van Pesch and his team at the Neurology Department of the Cliniques Universitaires de Saint Luc for their collaboration in exchanging CSF samples.

Funding

LB received for this study a 2 years full-time researcher funding by the Erasmus Fund for Research, a grant from the Belgian League Against Multiple Sclerosis and the Minerva grant. This work was additionally supported by MSCA-IF-2018 (843107, recipient: X.B.), by Innoviris BB2B-Attract (RBC/BFB1, recipient: X.B), Fonds Paul Genicot (recipient: X.B.), Fonds Gaston Ithier (recipient: X.B), the FNRS (CDR n°40008450 and 40028133).

Competing interests

S.B. received speaker/consulting honoraria and/or travel grants from Sanofi, Roche, Janssen, Merck, Novartis, Alexion, and Amgen.

Prof. van Pesch has received travel grants from Merck Healthcare KGaA (Darmstadt, Germany), Sanofi, Servier and Roche.

His institution has received research grants and consultancy fees from Roche, Biogen, Sanofi, Merck Healthcare KGaA (Darmstadt, Germany), Bristol Meyer Squibb, Alexion, Amgen, Neuraxpharm and Novartis Pharma.

The remaining authors report no competing interests.

Supplementary material

Supplementary material is available at *Brain* online.

Authors contributions

LB and VI handled CSF samples and prepared samples for proteomics analysis. VI performed the MS analysis under the supervision of DC and XB. LB, EB and ES were involved in the data analysis of the SWATH-MS data. LB, GP and NG designed the clinical study. NG, DC and XB were involved in the overall design and supervised the study. LB with the help of GP, SB, SE and VVP recruited patients and compiled all the clinical data used in the study. LB and XB wrote the manuscript and all the authors have read and revised the article and approved the submitted version.

References

1. Muñoz González G, ´t Hart BA, Bugiani M, et al. A focus on the normal-appearing white and gray matter within the multiple sclerosis brain: a link to smoldering progression. *Acta Neuropathol (Berl)*. 2025;150(1):16. doi:10.1007/s00401-025-02923-1

2. Kuhlmann T, Moccia M, Coetzee T, et al. Multiple sclerosis progression: time for a new mechanism-driven framework. *Lancet Neurol*. 2023;22(1):78-88. doi:10.1016/S1474-4422(22)00289-7

3. Montalban X, Lebrun-Frénay C, Oh J, et al. Diagnosis of multiple sclerosis: 2024 revisions of the McDonald criteria. *Lancet Neurol*. 2025;24(10):850-865. doi:10.1016/S1474-4422(25)00270-4

4. Lebrun-Frenay C, Kantarci O, Siva A, et al. Radiologically Isolated Syndrome: 10-Year Risk Estimate of a Clinical Event. *Ann Neurol*. 2020;88(2):407-417. doi:10.1002/ana.25799

5. Tintoré M, Rovira A, Río J, et al. Baseline MRI predicts future attacks and disability in clinically isolated syndromes. *Neurology*. 2006;67(6):968-972. doi:10.1212/01.wnl.0000237354.10144.ec

6. Fisniku LK, Brex PA, Altmann DR, et al. Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis. *Brain*. 2008;131(3):808-817. doi:10.1093/brain/awm329

7. The Optic Neuritis Study Group. Multiple Sclerosis Risk After Optic Neuritis: Final Optic Neuritis Treatment Trial Follow-up. *Arch Neurol*. 2008;65(6):727-732. doi:10.1001/archneur.65.6.727

8. Brownlee WJ, Miller DH. Clinically isolated syndromes and the relationship to multiple sclerosis. *J Clin Neurosci Off J Neurosurg Soc Australas*. 2014;21(12):2065-2071. doi:10.1016/j.jocn.2014.02.026

9. Tintore M, Rovira À, Río J, et al. Defining high, medium and low impact prognostic factors for developing multiple sclerosis. *Brain*. 2015;138(7):1863-1874. doi:10.1093/brain/awv105

10. Comi G, Radaelli M, Soelberg Sørensen P. Evolving concepts in the treatment of relapsing multiple sclerosis. *Lancet Lond Engl*. 2017;389(10076):1347-1356. doi:10.1016/S0140-6736(16)32388-1

11. Montalban X, Gold R, Thompson AJ, et al. ECTRIMS/EAN Guideline on the pharmacological treatment of people with multiple sclerosis. *Mult Scler Houndmills Basingstoke Engl*. 2018;24(2):96-120. doi:10.1177/1352458517751049

12. Kappos L, Edan G, Freedman MS, et al. The 11-year long-term follow-up study from the randomized BENEFIT CIS trial. *Neurology*. 2016;87(10):978-987. doi:10.1212/WNL.0000000000003078

13. Lublin FD, Häring DA, Ganjgahi H, et al. How patients with multiple sclerosis acquire disability. *Brain*. 2022;145(9):3147-3161. doi:10.1093/brain/awac016

14. Okuda DT, Kantarci O, Lebrun-Frénay C, et al. Dimethyl Fumarate Delays Multiple Sclerosis in Radiologically Isolated Syndrome. *Ann Neurol*. 2023;93(3):604-614. doi:10.1002/ana.26555

15. Leurs C, Twaalfhoven H, Lissenberg-Witte B, et al. Kappa free light chains is a valid tool in the diagnostics of MS: A large multicenter study. *Mult Scler Houndmills Basingstoke Engl*. 2020;26(8):912-923. doi:10.1177/1352458519845844

16. Levraut M, Laurent-Chabalier S, Ayrignac X, et al. Kappa Free Light Chain Biomarkers Are Efficient for the Diagnosis of Multiple Sclerosis. *Neurol Neuroimmunol Neuroinflammation*. 2022;10(1):e200049. doi:10.1212/NXI.0000000000200049

17. Levraut M, Gavoille A, Landes-Chateau C, et al. Kappa Free Light Chain Index Predicts Disease Course in Clinically and Radiologically Isolated Syndromes. *Neurol Neuroimmunol Neuroinflammation*. 2023;10(6):e200156. doi:10.1212/NXI.0000000000200156

18. Freedman MS, Gnanapavan S, Booth RA, et al. Guidance for use of neurofilament light chain as a cerebrospinal fluid and blood biomarker in multiple sclerosis management. *eBioMedicine*. 2024;101:104970. doi:10.1016/j.ebiom.2024.104970

19. Comabella M, Fernández M, Martin R, et al. Cerebrospinal fluid chitinase 3-like 1 levels are associated with conversion to multiple sclerosis. *Brain J Neurol*. 2010;133(Pt 4):1082-1093. doi:10.1093/brain/awq035

20. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature*. 2003;422(6928):198-207. doi:10.1038/nature01511

21. Liu Y, Hüttenhain R, Collins B, Aebersold R. Mass spectrometric protein maps for biomarker discovery and clinical research. *Expert Rev Mol Diagn*. 2013;13(8):811-825. doi:10.1586/14737159.2013.845089

22. Gillet LC, Navarro P, Tate S, et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics MCP*. 2012;11(6):O111.016717. doi:10.1074/mcp.O111.016717

23. Collins BC, Hunter CL, Liu Y, et al. Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. *Nat Commun*. 2017;8(1):291. doi:10.1038/s41467-017-00249-5

24. Xu Z, Lee A, Nouwens A, Henderson RD, McCombe PA. Mass spectrometry analysis of plasma from amyotrophic lateral sclerosis and control subjects. *Amyotroph Lateral Scler Front Degener*. 2018;19(5-6):362-376. doi:10.1080/21678421.2018.1433689

25. Ludwig C, Gillet L, Rosenberger G, Amon S, Collins BC, Aebersold R. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol Syst Biol*. 2018;14(8):e8126. doi:10.15252/msb.20178126

26. Anjo SI, Santa C, Manadas B. SWATH Mass Spectrometry Applied to Cerebrospinal Fluid Differential Proteomics: Establishment of a Sample-Specific Method. *Methods Mol Biol Clifton NJ*. 2019;2044:169-189. doi:10.1007/978-1-4939-9706-0_11

27. Andrés-Benito P, Povedano M, Domínguez R, et al. Increased C-X-C Motif Chemokine Ligand 12 Levels in Cerebrospinal Fluid as a Candidate Biomarker in Sporadic Amyotrophic Lateral Sclerosis. *Int J Mol Sci*. 2020;21(22):8680. doi:10.3390/ijms21228680

28. Park SA, Jung JM, Park JS, et al. SWATH-MS analysis of cerebrospinal fluid to generate a robust battery of biomarkers for Alzheimer's disease. *Sci Rep*. 2020;10:7423. doi:10.1038/s41598-020-64461-y

29. Delbaz A, Chen M, Jen FEC, et al. Neisseria meningitidis Induces Pathology-Associated Cellular and Molecular Changes in Trigeminal Schwann Cells. *Infect Immun*. 2020;88(4):e00955-19. doi:10.1128/IAI.00955-19

30. Bakochi A, Mohanty T, Pyl PT, et al. Cerebrospinal fluid proteome maps detect pathogen-specific host response patterns in meningitis. *eLife*. 10:e64159. doi:10.7554/eLife.64159

31. Santacruz CA, Vincent JL, Duitama J, et al. The Cerebrospinal Fluid Proteomic Response to Traumatic and Nontraumatic Acute Brain Injury: A Prospective Study. *Neurocrit Care*. 2022;37(2):463-470. doi:10.1007/s12028-022-01507-1

32. Zafar S, Noor A, Younas N, et al. SWATH Mass Spectrometry-Based CSF Proteome Profile of GBA-Linked Parkinson's Disease Patients. *Int J Mol Sci*. 2022;23(22):14166. doi:10.3390/ijms232214166

33. Karayel O, Virreira Winter S, Padmanabhan S, et al. Proteome profiling of cerebrospinal fluid reveals biomarker candidates for Parkinson's disease. *Cell Rep Med*. 2022;3(6):100661. doi:10.1016/j.xcrm.2022.100661

34. Pandit AK, Misra S, Sengupta S, et al. Cerebrospinal fluid proteins in idiopathic intracranial hypertension: An exploratory SWATH proteomics analysis. *PROTEOMICS – Clin Appl*. 2024;18(1):2300021. doi:10.1002/prca.202300021

35. Torres Iglesias G, López-Molina M, Botella L, et al. Differential Protein Expression in Extracellular Vesicles Defines Treatment Responders and Non-Responders in Multiple Sclerosis. *Int J Mol Sci*. 2024;25(19):10761. doi:10.3390/ijms251910761

36. Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol*. 2018;17(2):162-173. doi:10.1016/S1474-4422(17)30470-2

37. Giovannoni G, Turner B, Gnanapavan S, Offiah C, Schmierer K, Marta M. Is it time to target no evident disease activity (NEDA) in multiple sclerosis? *Mult Scler Relat Disord*. 2015;4(4):329-333. doi:10.1016/j.msard.2015.04.006

38. Manouchehrinia A, Westerlind H, Kingwell E, et al. Age Related Multiple Sclerosis Severity Score: Disability ranked by age. *Mult Scler Houndmills Basingstoke Engl*. 2017;23(14):1938. doi:10.1177/1352458517690618

39. Teunissen CE, Tumani H, Engelborghs S, Mollenhauer B. Biobanking of CSF: international standardization to optimize biomarker development. *Clin Biochem*. 2014;47(4-5):288-292. doi:10.1016/j.clinbiochem.2013.12.024

40. Lowry OliverH, Rosebrough NiraJ, Farr AL, Randall RoseJ. PROTEIN MEASUREMENT WITH THE FOLIN PHENOL REAGENT. *J Biol Chem*. 1951;193(1):265-275. doi:10.1016/S0021-9258(19)52451-6

41. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*. 2019;47(Database issue):D506-D515. doi:10.1093/nar/gky1049

42. Demichev V, Messner CB, Vernardis SI, Lilley KS, Ralser M. DIA-NN: Neural networks and interference correction enable deep proteome coverage in high throughput. *Nat Methods*. 2020;17(1):41-44. doi:10.1038/s41592-019-0638-x

43. Frankenfield AM, Ni J, Ahmed M, Hao L. Protein Contaminants Matter: Building Universal Protein Contaminant Libraries for DDA and DIA Proteomics. *J Proteome Res*. 2022;21(9):2104-2113. doi:10.1021/acs.jproteome.2c00145

44. Demichev V, Yu F, Teo GC, et al. High sensitivity dia-PASEF proteomics with DIA-NN and FragPipe. *bioRxiv*. Preprint posted online March 9, 2021:2021.03.08.434385. doi:10.1101/2021.03.08.434385

45. Choi M, Chang CY, Clough T, et al. MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinforma Oxf Engl*. 2014;30(17):2524-2526. doi:10.1093/bioinformatics/btu305

46. R Core Team (2025). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

47. Kohler D, Staniak M, Yu F, Nesvizhskii AI, Vitek O. An MSstats workflow for detecting differentially abundant proteins in large-scale data-independent acquisition mass spectrometry experiments with FragPipe processing. *Nat Protoc*. Published online May 20, 2024:1-24. doi:10.1038/s41596-024-01000-3

48. Borràs E, Cantó E, Choi M, et al. Protein-Based Classifier to Predict Conversion from Clinically Isolated Syndrome to Multiple Sclerosis. *Mol Cell Proteomics MCP*. 2016;15(1):318-328. doi:10.1074/mcp.M115.053256

49. Fan G, Li G, Li L, Da Y. Pin1 maintains the effector program of pathogenic Th17 cells in autoimmune neuroinflammation. *J Autoimmun*. 2024;147:103262. doi:10.1016/j.jaut.2024.103262

50. Åkesson J, Hojjati S, Hellberg S, et al. Proteomics reveal biomarkers for diagnosis, disease activity and long-term disability outcomes in multiple sclerosis. *Nat Commun*. 2023;14:6903. doi:10.1038/s41467-023-42682-9

51. Hinsinger G, Du Trieu De Terdonck L, Urbach S, et al. CD138 as a Specific CSF Biomarker of Multiple Sclerosis. *Neurol Neuroimmunol Neuroinflammation*. 2024;11(3):e200230. doi:10.1212/NXI.0000000000200230

52. Møllgaard M, Degn M, Sellebjerg F, Frederiksen JL, Modvig S. Cerebrospinal fluid chitinase-3-like 2 and chitotriosidase are potential prognostic biomarkers in early multiple sclerosis. *Eur J Neurol*. 2016;23(5):898-905. doi:10.1111/ene.12960

53. Hejazi MS, Jafari S, Montazersaheb S, et al. Annexin A1, calreticulin and high mobility group box 1 are elevated in secondary progressive multiple sclerosis: Does immunogenic cell death occur in multiple sclerosis? *BioImpacts BI*. 2025;15:30264. doi:10.34172/bi.30264

54. Ní Fhlathartaigh M, McMahon J, Reynolds R, et al. Calreticulin and other components of endoplasmic reticulum stress in rat and human inflammatory demyelination. *Acta Neuropathol Commun*. 2013;1:37. doi:10.1186/2051-5960-1-37

55. Xu T, Shi Y, Zheng G, Zhang G. Diagnostic Potential of Two Novel Biomarkers for Neuromyelitis Optica Spectrum Disorder and Multiple Sclerosis. *Diagn Basel Switz*. 2023;13(9):1572. doi:10.3390/diagnostics13091572

56. Stadelmann C, Kerschensteiner M, Misgeld T, Brück W, Hohlfeld R, Lassmann H. BDNF and gp145trkB in multiple sclerosis brain lesions: neuroprotective interactions between immune and neuronal cells? *Brain J Neurol*. 2002;125(Pt 1):75-85. doi:10.1093/brain/awf015

57. Mohan H, Krumbholz M, Sharma R, et al. Extracellular Matrix in Multiple Sclerosis Lesions: Fibrillar Collagens, Biglycan and Decorin are Upregulated and Associated with Infiltrating Immune Cells. *Brain Pathol*. 2010;20(5):966-975. doi:10.1111/j.1750-3639.2010.00399.x

58. Herrmann AK, Wüllner V, Moos S, et al. Dimethyl fumarate alters intracellular Ca2+ handling in immune cells by redox-mediated pleiotropic effects. *Free Radic Biol Med*. 2019;141:338-347. doi:10.1016/j.freeradbiomed.2019.07.005

59. Comabella M, Sastre-Garriga J, Borras E, et al. CSF Chitinase 3–Like 2 Is Associated With Long-term Disability Progression in Patients With Progressive Multiple Sclerosis. *Neurol Neuroimmunol Neuroinflammation*. 2021;8(6):e1082. doi:10.1212/NXI.0000000000001082

60. Petržalka M, Meluzínová E, Libertínová J, et al. IL-2, IL-6 and chitinase 3-like 2 might predict early relapse activity in multiple sclerosis. *PLoS ONE*. 2022;17(6):e0270607. doi:10.1371/journal.pone.0270607

61. Rosenstein I, Axelsson M, Novakova L, et al. Intrathecal kappa free light chain synthesis is associated with worse prognosis in relapsing–remitting multiple sclerosis. *J Neurol*. 2023;270(10):4800-4811. doi:10.1007/s00415-023-11817-9

62. Duell F, Högelin KA, Vlad B, et al. Kappa Free Light Chain Index Correlates With Prognostic Biomarkers in Multiple Sclerosis and Decreases Slowly Following Treatment. *Eur J Neurol*. 2025;32(7):e70291. doi:10.1111/ene.70291

63. Zhou W, Graner M, Beseler C, et al. Plasma IgG aggregates as biomarkers for multiple sclerosis. *Clin Immunol Orlando Fla*. 2023;256:109801. doi:10.1016/j.clim.2023.109801

64. Wuhrer M, Selman MHJ, McDonnell LA, et al. Pro-inflammatory pattern of IgG1 Fc glycosylation in multiple sclerosis cerebrospinal fluid. *J Neuroinflammation*. 2015;12(1):235. doi:10.1186/s12974-015-0450-1

65. Elkjaer ML, Frisch T, Reynolds R, et al. Molecular signature of different lesion types in the brain white matter of patients with progressive multiple sclerosis. *Acta Neuropathol Commun*. 2019;7:205. doi:10.1186/s40478-019-0855-7

66. Liang P, Le W. Role of autophagy in the pathogenesis of multiple sclerosis. *Neurosci Bull*. 2015;31(4):435-444. doi:10.1007/s12264-015-1545-5

67. Tang C, Yang J, Zhu C, et al. Iron metabolism disorder and multiple sclerosis: a comprehensive analysis. *Front Immunol*. 2024;15:1376838. doi:10.3389/fimmu.2024.1376838

68. Borziak K, Finkelstein J. X-linked genetic risk factors that promote autoimmunity and dampen remyelination are associated with multiple sclerosis susceptibility. *Mult Scler Relat Disord*. 2022;66. doi:10.1016/j.msard.2022.104065

69. Liguori M, Nuzziello N, Licciulli F, et al. Combined microRNA and mRNA expression analysis in pediatric multiple sclerosis: an integrated approach to uncover novel pathogenic mechanisms of the disease. *Hum Mol Genet*. 2018;27(1):66-79. doi:10.1093/hmg/ddx385

70. Dunwiddie TV, Masino SA. The role and regulation of adenosine in the central nervous system. *Annu Rev Neurosci*. 2001;24:31-55. doi:10.1146/annurev.neuro.24.1.31

71. Beutler E. Cladribine (2-chlorodeoxyadenosine). *The Lancet*. 1992;340(8825):952-956. doi:10.1016/0140-6736(92)92826-2

72. Giovannoni G, Comi G, Cook S, et al. A Placebo-Controlled Trial of Oral Cladribine for Relapsing Multiple Sclerosis. *N Engl J Med*. 2010;362(5):416-426. doi:10.1056/NEJMoa0902533

73. Schirò G, Iacono S, Ragonese P, Aridon P, Salemi G, Balistreri CR. A Brief Overview on BDNF-Trk Pathway in the Nervous System: A Potential Biomarker or Possible Target in Treatment of Multiple Sclerosis? *Front Neurol*. 2022;13:917527. doi:10.3389/fneur.2022.917527

74. Lorenzon A, Pilichou K, Rigato I, et al. Homozygous Desmocollin-2 Mutations and Arrhythmogenic Cardiomyopathy. *Am J Cardiol*. 2015;116(8):1245-1251. doi:10.1016/j.amjcard.2015.07.037

75. Zeng J, Sun Y, Man Y, Tang H, Xie L, He M. Validation the role of desmocollin-2 in osteosarcoma based on single cell and bulk RNA seq and experimental analyses. *J Cancer*. 2023;14(14):2619-2632. doi:10.7150/jca.87411

76. Han JY, Che N, Mo J, et al. Desmoglein 2 and desmocollin 2 depletions promote malignancy through distinct mechanisms in triple-negative and luminal breast cancer. *BMC Cancer*. 2024;24:532. doi:10.1186/s12885-024-12229-2

77. Zhang H, Li YC, Pang D, et al. Desmocollin 2 is a dominant entry receptor for Epstein-Barr virus infection of epithelial cells. *Nat Microbiol*. Published online July 25, 2025. doi:10.1038/s41564-025-02067-8

78. Lanz TV, Brewer RC, Ho PP, et al. Clonally Expanded B Cells in Multiple Sclerosis Bind EBV EBNA1 and GlialCAM. *Nature*. 2022;603(7900):321-327. doi:10.1038/s41586-022-04432-7

79. Bjornevik K, Cortese M, Healy BC, et al. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science*. 2022;375(6578):296-301. doi:10.1126/science.abj8222

80. Fejza A, Poletto E, Carobolante G, et al. Multimerin-2 orchestrates the cross-talk between endothelial cells and pericytes: A mechanism to maintain vascular stability. *Matrix Biol Plus*. 2021;11:100068. doi:10.1016/j.mbplus.2021.100068

81. Girolamo F, Coppola C, Ribatti D, Trojano M. Angiogenesis in multiple sclerosis and experimental autoimmune encephalomyelitis. *Acta Neuropathol Commun*. 2014;2:84. doi:10.1186/s40478-014-0084-z

82. Nishihara H, Perriot S, Gastfriend BD, et al. Intrinsic blood–brain barrier dysfunction contributes to multiple sclerosis pathogenesis. *Brain*. 2022;145(12):4334-4348. doi:10.1093/brain/awac019

83. Perez-Riverol Y, Bandla C, Kundu DJ, et al. The PRIDE database at 20 years: 2025 update. *Nucleic Acids Res*. 2025;53(D1):D543-D553. doi:10.1093/nar/gkae1011

Figure Legends

Figure 1 Experimental design of the study

(A) Retrospective and prospective enrollment of multiple sclerosis (MS), clinically isolated syndrome (CIS) and controls (CTL) patients form 3 Belgium sites Hopital Universitaire de Bruxelles (HUB), Centre Hospitalier Universitaire de Brugmann (CHUB) and Cliniques Universitaires de Saint-Luc (CUSL). **(B)** Preparation and acquisition of the collected CSF samples by label-free data independent acquisition. **(C)** Clinical follow-up of MS and CIS patients up to five years. **(D)** Differential expression analysis performed with the MSstats R-package across diagnostic groups (MS vs. CTL, MSCIS vs. CTL, CIS vs. CTL and MS vs. CIS) and disease activity vs. no evidence of disease activity at 2 and 5 years of follow-up in the MSCIS group. **(E)** Diagnosis and prognosis performance of the 15 shared proteins of the 6 groups' comparisons evaluated by logistic regression model with cross-validation.

Figure 2. Overview of the proteome variations in relation to compared groups of patients

Differential expression analysis of CSF proteins in MS and CIS (n = 77) patients versus CTL subjects (n = 43) and among MS (n = 62) and CIS (n = 15) patients between active (n = 26/34) versus stable disease (n = 34/10) defined by the NEDA achievement at 2 and 5 years of follow-up. **(A)** Volcano plots showing differentially expressed proteins with a false discovery rate (FDR) < 0.05 in relation to the compared groups. The differential expression analysis was performed by the MSstats R-package based on linear mixed-effects models. The top 10 upregulated or downregulated proteins are marked. The proteins in purple are common differentially expressed proteins across the 6 comparisons. **(B)** Venn diagrams highlighting

the proteins overlap between the different groups' comparison. **(C)** Heatmaps based on the $\log_2$ fold change (FC) of all identified differentially expressed proteins grouped by condition comparisons and activity vs stable disease at 2 and 5 years of follow-up and showing their behavior in other groups comparisons. The asterisk represents the level of statistical significance: ** *FDR* < 0.05, * *P* < 0.05.

Figure 3. Identification of common differentially expressed proteins between diagnostic and prognostic groups

(A) Venn diagram showing the overlap between both groups' comparison: NEDA vs EDA representing all the identified proteins of the active vs stable disease at 2 and 5 years of follow-up and MS, CIS and CTL comparisons grouping the differentially expressed proteins of the 4 conditions' comparisons. Boxplots of the statistically relevant association assessed by a two-sided Mann–Whitney U-test corrected for multiple comparison with Benjamini–Hochberg method between these 15 proteins and **(B)** the presence or absence of spinal MRI lesion at the initial assessment, **(C)** ARMSS at the end of follow-up compared to the median (2.45 [0.3-8.9]) of the MSCIS group, and **(D)** the EDA versus NEDA at 5 years of follow-up. The boxplot represents the median, Q1, Q3 and the IQR.

Figure 4. MSCIS vs CTL and MS vs CTL classification performance of the 15 common proteins

Performance of the 15 proteins for predicting diagnosis of MSCIS vs CTL and MS vs CTL determined by a logistic regression model which uses the area under the receiver operating

characteristics curve (AUROC) to assess the predictive performance of a protein, with a forward selection of the proteins and a cross validation by splitting the cohort in training/validation groups (2/3-1/3) over 500 iterations. **(A)** Frequency plots represent the top three protein combination that was the most selected as the best classifier over the 500 iterations. **(B)** Predictive power of the top three protein combination in differentiating MSCIS (n = 77) vs CTL (n = 43) and MS (n = 62) vs CTL (n = 43), is analyzed by AUROC. The AUROC scores were assessed by a two-sided Mann–Whitney U-test. The P were in the order: 8×10^{-6}, 4×10^{-7}, 1×10^{-5}, 2×10^{-7}, 5×10^{-6} and 9×10^{-8}. The purple AUROC was obtained with a multivariate logistic regression with bootstrapping on the entire cohort for the unique proteins from the top three protein combinations (TOP3 unique proteins). The P were in the order 1×10^{-7} and 1×10^{-8}

Figure 5 CIS vs CTL and MS vs CIS classification performance of the 15 common proteins

Performance of the 15 proteins for predicting diagnosis of CIS vs CTL and MS vs CIS determined by a logistic regression model which uses the area under the receiver operating characteristics curve (AUROC) to assess the predictive performance of a protein, with a forward selection of the proteins and a cross validation by splitting the cohort in training/validation groups (2/3-1/3) over 500 iterations. **(A)** Frequency plots represent the top three protein combination that was the most selected as the best classifier over the 500 iterations. **(B)** Predictive power of the top three protein combination in differentiating CIS (n = 15) vs CTL (n = 43) and MS (n = 62) vs CIS (n = 15), is analyzed by AUROC. The AUROC scores were assessed by a two-sided Mann–Whitney U-test. The P were in the order: 0.0003, 0.0002, 0.0004, 0.004, 0.0009 and 0.002. The purple AUROC was obtained with a

multivariate logistic regression with bootstrapping on the entire cohort for the unique proteins from the top three protein combinations (TOP3 unique proteins). The *P* were in the order 2×10^{-5} and 0.0007.

Figure 6 NEDA vs EDA classification performance of the 15 common proteins at 2 and 5 years of follow-up

Performance of the 15 proteins for predicting prognosis by the disease activity at 2 and 5 years of follow-up after treatment start determined by a logistic regression model which use the area under the receiver operating characteristics curve (AUROC) to assess the predictive performance of a protein, with a forward selection of the proteins and a cross validation by splitting the cohort in training/validation groups (2/3-1/3) over 500 iterations. **(A)** Frequency plots represent the top three protein combination that was the most selected as the best classifier over the 500 iterations. **(B)** Predictive power of the top three protein combination in differentiating EDA (n =26) vs NEDA (n = 34) at 2 years and EDA (n = 34) vs NEDA (n = 10) at 5 years, is analyzed by AUROC. The AUROC scores were assessed by a two-sided Mann-Whitney U-test. The *P* were in the order: 8×10^{-4}, 3×10^{-4}, 4×10^{-5}, 2×10^{-5}, 3×10^{-5} , and 3×10^{-5}. The purple AUROC was obtained with a multivariate logistic regression with bootstrapping on the entire cohort for the unique proteins from the top three protein combinations (TOP3 unique proteins).

Table 1 Baseline and clinical characteristics of the patients with multiple sclerosis, clinically isolated syndrome and controls

Clinical characteristics of the cohort			MS	CIS	CTL	*FDR*-value*
Cohort size		*n*	**62**	**15**	**43**	
Number of MS and CIS diagnosis at T0		*n*	59	18		
Number of conversion CIS to MS		*n*		3		
Delay onset to diagnosis		Days Median (range)	60.5 (2–748)	30 (7–1095)		0.463
Age[a]		Years Median (range)	30 (14–80)	35 (19–55)	38 (20–82)	**0.045**
Sex		F / M	46 / 16	12 / 3	24 / 19	0.129
Follow-up		Years Median (range)	4 (0–10)	3.14 (2–10)	NA	0.705
CSF data[a]	*WCC*	Median (range)	4 (0–59)	3 (0–17)	1 (0–4.2)	**< 0.001**
	OCBs	Yes / No	59 / 3	9 / 6	0 / 40[b]	**< 0.001**
Clinical onset	*ON*	*n* (%)	9 (15%)	11 (73%)		**< 0.001**
	TM	*n* (%)	28 (46%)	1 (7%)		**0.014**
	PF	*n* (%)	14 (23%)	1 (7%)		0.352
	Enc	*n* (%)	10 (16%)	2 (13%)		1
Annualized clinical and MRI relapse rate			0.18 (0–2.35)	0		**< 0.001**
Number of clinical relapse		Median (range)	0 (0–2)	0		**0.020**
Time to MRI or clinical relapse after treatment start		Months	12.1 (-1.1–93.5)	NA		NA
EDSS	*Baseline*	Median (range)	2 (0–7.5)	2 (0–3,5)		0.585
	End	Median (range)	1.5 (0–8)	1 (–3)		0.176
MRI data	*Brain T2 lesion*	0 lesion	0	8 (53%)		**< 0.001**
		1 to 3 lesions	9 (15%)	5 (33%)		0.250
		4 to 9 lesions	23 (37%)	2 (13%)		0.176
		> 10 lesions or confluent	30 (48%)	0		**< 0.001**
	Spinal T2 lesion	No	15 (26%)	12 (80%)		**< 0.001**
		Yes	46 (77%)	3 (20%)		**< 0.001**
	Gd+ lesions	No	25 (40%)	11 (73%)		0.067
		Yes	37 (60%)	4 (27%)		0.067
Treatment baseline / End[b]		First line	42 / 30	0 / 1		**< 0.001 / < 0.001**
		Second line	5 / 18	0 / 0		0.661 / **0.030**
		None	11 / 10	15 / 14		**< 0.001 / < 0.001**
NEDA at 2y n = 60[b]		Yes / no	21 / 26	13 / 0		**<0.001**
NEDA at 5y n = 44[b]		Yes / no	4 / 34	6 / 0		**0.038**
ARMSS at the end of FU[b]		Median (range)	3 (0.3–8.9)	1 (0.4–5)		**0.030**

MS multiple sclerosis patients, CIS clinically isolated syndrome patients, CTL controls patients, CSF cerebrospinal fluid, WCC white cells count, OCBs Oligoclonal bands, ON optic neuritis, TM transverse myelitis, PF posterior fossae, Enc encephalitic MRI magnetic resonance imagery, EDSS expanded disability status scale, Gd+ gadolinium enhancement, NEDA no evidence of disease activity by no relapses, no increase in disability and no new or active lesion on MRI.

**P* was calculated for MS vs CIS except for age, sex and CSF data it was MS vs CIS vs CTL patients. Two-sided Fisher's exact test or Chi squared test when appropriate were used for contingency tables or two-sided Mann–Whitney U-test and Kruskal–Wallis's test when appropriate for continuous variables. A correction for multiple comparisons was done by Benjamini–Hochberg method (FDR-value).

[a]Age was significantly different only between MS and CTL patients. CSF white cells count was significantly different between MS vs CTL and CIS vs CTL patients but not between MS vs CIS. OCBs were significantly different between MS vs CIS patients.

[b] Missing data exists. Three CTL patients had no OCBs information. Treatment baseline and end (n = 58): For MS patients were lost of follow-up before starting a treatment during and one did not benefit DMT because of an ongoing cancer treatment and was classify as none DMT. ARMSS (n = 61).

Table 2 Description of the common 15 differentially expressed proteins between conditions MS-CIS-CTL and active versus stable disease comparisons.

Common proteins	Protein name	Previously described as potential biomarker	Reference
A0A7P0Z497	Peptidyl-prolyl cis-trans isomerase	No, but elevated in MS patients and in EAE mice	Fan et al. 2024
A0A8J8YUT9	Tyrosine-protein kinase receptor	No, but the receptor family is involved in immune-mediated neuroprotective interactions in MS	Stadelmann et al. 2002
ADA2	Adenosine deaminase 2	Yes, potential biomarker of the active phase of MS	Hinsinger et al. 2024
AT2A2	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	No, but modified by the dimethyl fumarate	Hermann et al. 2019
CALR	Calreticulin	Elevated in SPMS compared to control Increased in sera of MS patients compared to control	Hejazi et al. 2024 Ní Fhlathartaigh et al. 2013
CH3L2	Chitinase-3-like protein 2	Yes, for diagnosis and prognosis	Mollgaard et al. 2016
CO3A1	Collagen alpha-1(III) chain	No, but its gene is strongly induced in active and even more in inactive MS lesions	Mohan et al. 2010
DSC2	Desmocollin-2	No	
IGHG1	Immunoglobulin heavy constant gamma 1	Yes, locally synthesized IgG in the CSF of MS patients consists mainly of IgG1 Serum IgG1 can discriminate between MS vs control, SPMS vs RRMS and SPMS vs PPMS	Kashka et al. 1979 Zhou et al. 2023
IGKC	Immunoglobulin kappa constant	Yes, its fragment KFLC for diagnosis of MS and prognosis.	Levraut et al. 2022 Levraut et al. 2023
IRPL1	Interleukin-1 receptor accessory protein-like 1	No	
LAMP2	Lysosome-associated membrane glycoprotein 2	Yes, predictor of the MS progression in serum	Xu et al. 2023
MMRN2	Multimerin-2	No	
PCOC1	Procollagen C-endopeptidase enhancer 1	No	
TSP2	Thrombospondin-2	No	

Table summarizing the 15 proteins with their complete name and if they were already described as potential biomarkers in scientific literature.

Supplementary detailed proteomics method

Samples protocol preparation

CSFs were thawed and concentrated with a centrifugation column (Centricon 3kDa, Millipore Merck) at 4°C and 14.000g until 25 to 30µL.

The depletion of the top 14 most abundant proteins, was performed on affinity columns (Pierce Top 14 Abundant Protein Depletion Spin Columns; Thermo Scientific) then the protein concentration was measured with the Folin method.[1]

Proteins were diluted in 25mM ammonium carbonate and dithiothreitol was added for a final concentration of 25mM. Each sample was incubated during 1h at 4°C. Alkylation of samples was performed by the addition of iodoacetamide with final concentration of 71mM before incubation during 1h at 4°C. All samples were diluted 4 times with acetone and incubated à -20°C during 1h for protein precipitation. After centrifugation at 13.000 rpm during 20min at 4°C, acetone was removed and 40 ng of trypsin (Gold Mass Spectrometry Grade V5280 Promega, USA WI) was added. Samples were incubated and shaked at 37°C during 30min and then over night without shaking.

The pH of each sample was adjusted at 2 with a 10% formic acid (FA) solution. Finally, the clean-up of the fractions was achieved using a 1mL High Lipophilic Balance column equilibrated by acetonitrile (ACN) 80% and ACN 5%/0.1% FA, eluted by ACN 80%/0.1% FA and dried in a vacuum concentrator and resuspended in tampon A (H20 100%/HCOOH 0.1%).

Proteomic Analysis of CSF

Data Dependant Analysis

Proteins were identified from CSF samples by label-free DDA mass spectrometry (TripleTOF 5600 SWATH, AB Sciex) coupled to micro-high-performance liquid chromatography (µLC 425, Eksigent). Peptides separation was performed by applying a 90min gradient

hydrophobic separation on a column (ChromXP C18 CL, 150mm x 0.3mm, 3µM, 120A, Sciex) using a two-step ACN gradient with FA: 5-25% ACN / 0.1% FA in 48min then 25%-60% ACN / 0.1% FA in 20min and were sprayed online in the mass spectrometer. Subsequently, peptides were ionized by electrospray ionization (ESI) and injected into the TripleTOF 5600 mass spectrometer (TT 5600, AB Sciex). The 20 most intense precursors with charge state 2 to 4 were selected for fragmentation. MS1 spectra were collected in the range 400-1250 m/z with an accumulation of 250ms and MS2 spectra were collected in the range 100-2000 m/z for 100ms.

Generation of Spectral Libraries

The generated .WIFF files were directly analysed by ProteinPilot software (AB Sciex) and the recorded spectra were matched to peptides by the search algorithm Paragon (AB Sciex) using the database human Swiss-Prot part of UniProtKB.[2] Trypsin was selected as digestion enzyme. Oxidation at methionine was set as dynamic modification, carbamidomethylation as static modification at cysteine. Decoy database search was performed with target false discovery rate < 1% (FDR).

Data Independent Analysis

1µg of peptides was injected using SWATH-MSp acquisition on a Triple TOF 5600 mass spectrometer (Sciex, Concord, Canada) interfaced to an Eksigent NanoLC Ultra 2D HPLC System (Eksignet, Dublin, CA). Peptides were injected on a separation column (Eksigent ChromXP C18 CL, 150mm x 0.3mm, 3µM, 120A, Sciex) using a two steps acetonitrile gradient (5-25% ACN / 0.1% HCOOH in 48min then 25%-60% ACN / 0.1% HCOOH in 20min) and were sprayed online in the mass spectrometer. Swath acquisitions were performed using 42 windows of fix effective isolation width to cover a mass range of 200-1500m/z. The collision energy for each window was determined according to the calculation for a charge 2+ ion centered upon the window with a spread of 15. An accumulation time of 76ms was used for all fragment-ion scans in high-sensitivity mode and for the survey scans

in high-resolution mode acquired at the beginning of each cycle, resulting in a duty cycle of ~3.3s.

Raw data were converted to mzML format with MSConvert (Version 3.0) and their processing was carried out with DIA-NN[3] (version 1.8.2 beta 27) with search parameters set as follows: precursor FDR 1%; mass accuracy at MS1 and MS2 both set to 0; scan window set to 0; isotopologues and MBR turned on; protein inference at gene level; heuristic protein inference enabled; quantification strategy set to QuantUMS (high accuracy); neural network classifier double-pass mode; cross-run normalization RT-dependent. A universal library was used, and protein re-annotation was performed. Spectra were searched against a prior *in loco* generated library containing 9972 proteins covered by 174348550 peptides including for common contaminants[4].

To generate the library, DDA and DIA data acquired on TT5600 were used to generate one universal library by FragPipe[5] (version 18.0) using a pre-defined workflow DIA_SpecLib_Quant. Specifically, decoys were first added to the FASTA which contains human protein sequences (UP000005640, 78986 reviewed and unreviewed entries, downloaded from UniProtKB on 07th June 2022). Then, MSFragger (version 3.5) was used to search DDA raw data, with the following settings: precursor and fragment mass tolerance 30□ppm; strict trypsin with no more than five missed cleavages; peptide length 6–50; peptide mass 500–5000; C□+□57.021464 as fixed modification; M□+□15.9949, N-term +42.0106, nQ -17.0265, nE -18.0106, DN +0.984016 and carbamylation +43.005814 as variable modifications; min matched fragments 4; max fragment charge 2. In the validation step, MSBooster was implemented on both spectra and RT levels, and then Percolator and ProteinProphet integrated in Philosopher (version 4.4.0) were used for PSM validation and protein inference. Library generation was conducted using EasyPQP (version 0.1.50) with RT calibration based on ciRTs and Lowess fraction set to 0.04. Only fragment types b and y were included with a tolerance of 15□ppm with a max delta_unimod.

References

1. Lowry OliverH, Rosebrough NiraJ, Farr AL, Randall RoseJ. PROTEIN MEASUREMENT WITH THE FOLIN PHENOL REAGENT. *J Biol Chem*. 1951;193(1):265-275. doi:10.1016/S0021-9258(19)52451-6

2. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*. 2019;47(Database issue):D506-D515. doi:10.1093/nar/gky1049

3. Demichev V, Messner CB, Vernardis SI, Lilley KS, Ralser M. DIA-NN: Neural networks and interference correction enable deep proteome coverage in high throughput. *Nat Methods*. 2020;17(1):41-44. doi:10.1038/s41592-019-0638-x

4. Frankenfield AM, Ni J, Ahmed M, Hao L. Protein Contaminants Matter: Building Universal Protein Contaminant Libraries for DDA and DIA Proteomics. *J Proteome Res*. 2022;21(9):2104-2113. doi:10.1021/acs.jproteome.2c00145

5. Demichev V, Yu F, Teo GC, et al. High sensitivity dia-PASEF proteomics with DIA-NN and FragPipe. *bioRxiv*. Preprint posted online March 9, 2021:2021.03.08.434385. doi:10.1101/2021.03.08.434385

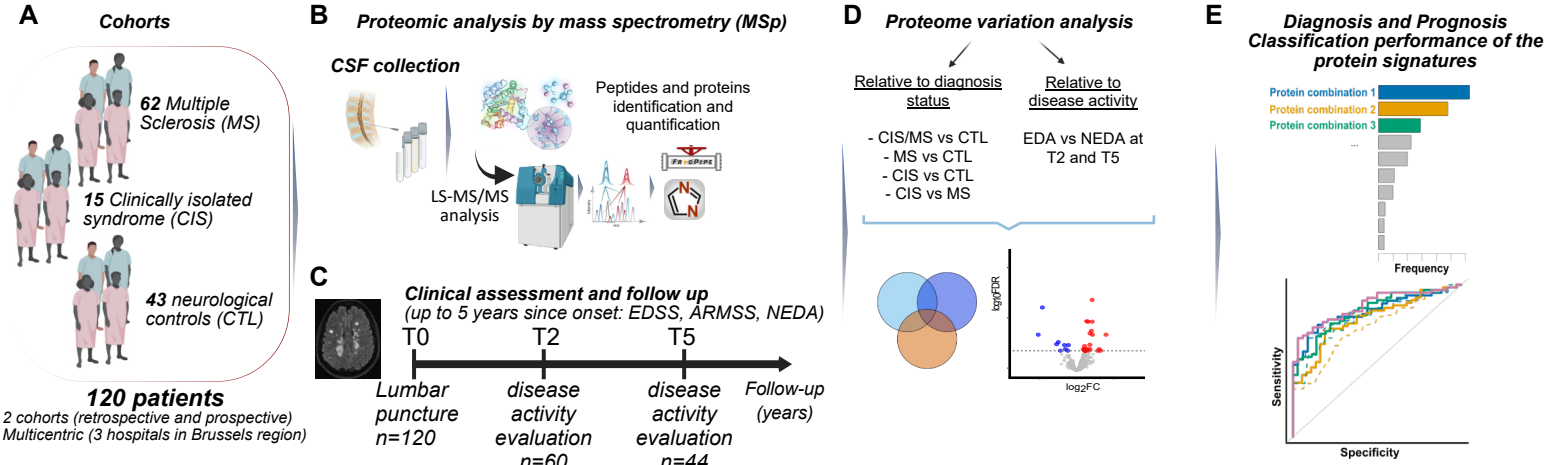
A
Cohorts
62 Multiple Sclerosis (MS)
15 Clinically isolated syndrome (CIS)
43 neurological controls (CTL)
120 patients
2 cohorts (retrospective and prospective)
Multicentric (3 hospitals in Brussels region)
B
Proteomic analysis by mass spectrometry (MSp)
CSF collection
Peptides and proteins identification and quantification
LS-MS/MS analysis
C
Clinical assessment and follow up
(up to 5 years since onset: EDSS, ARMSS, NEDA)
T0
T2
T5
Lumbar puncture n=120
disease activity evaluation n=60
disease activity evaluation n=44
Follow-up (years)
D
Proteome variation analysis
Relative to diagnosis status
Relative to disease activity
- CIS/MS vs CTL
- MS vs CTL
- CIS vs CTL
- CIS vs MS
EDA vs NEDA at T2 and T5
log2FDR
log2FC
E
Diagnosis and Prognosis Classification performance of the protein signatures
Protein combination 1
Protein combination 2
Protein combination 3
Frequency
Sensitivity
Specificity

Figure 1

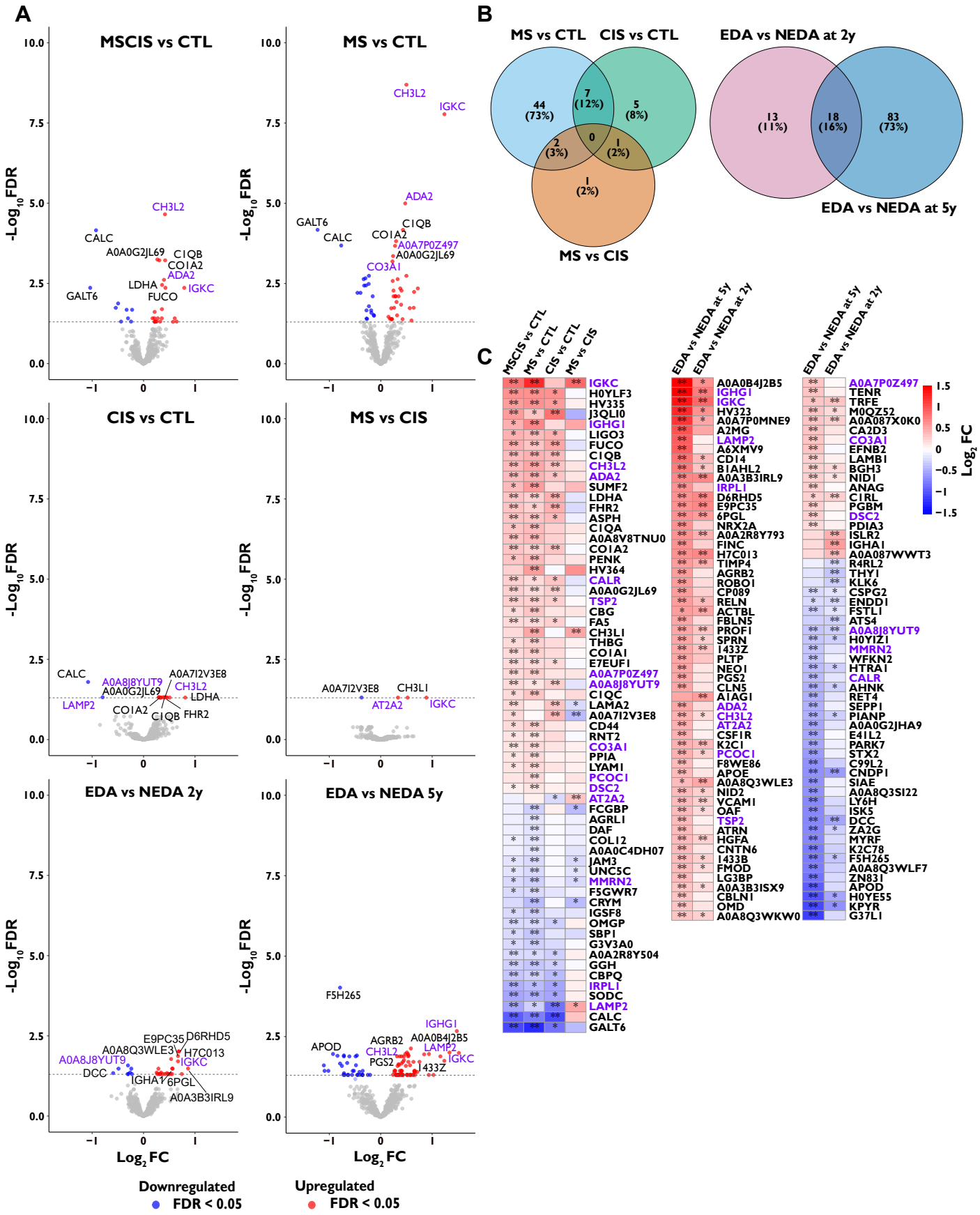
A
MSCIS vs CTL
MS vs CTL
CIS vs CTL
MS vs CIS
EDA vs NEDA 2y
EDA vs NEDA 5y
-Log10 FDR
Log2 FC
Downregulated
FDR < 0.05
Upregulated
FDR < 0.05
B
MS vs CTL
CIS vs CTL
MS vs CIS
EDA vs NEDA at 2y
EDA vs NEDA at 5y
44 (73%)
7 (12%)
5 (8%)
2 (3%)
0
1 (2%)
1 (2%)
13 (11%)
18 (16%)
83 (73%)
C
MSCIS vs CTL
MS vs CTL
CIS vs CTL
MS vs CIS
EDA vs NEDA at 5y
EDA vs NEDA at 2y
Log2 FC

Figure 2

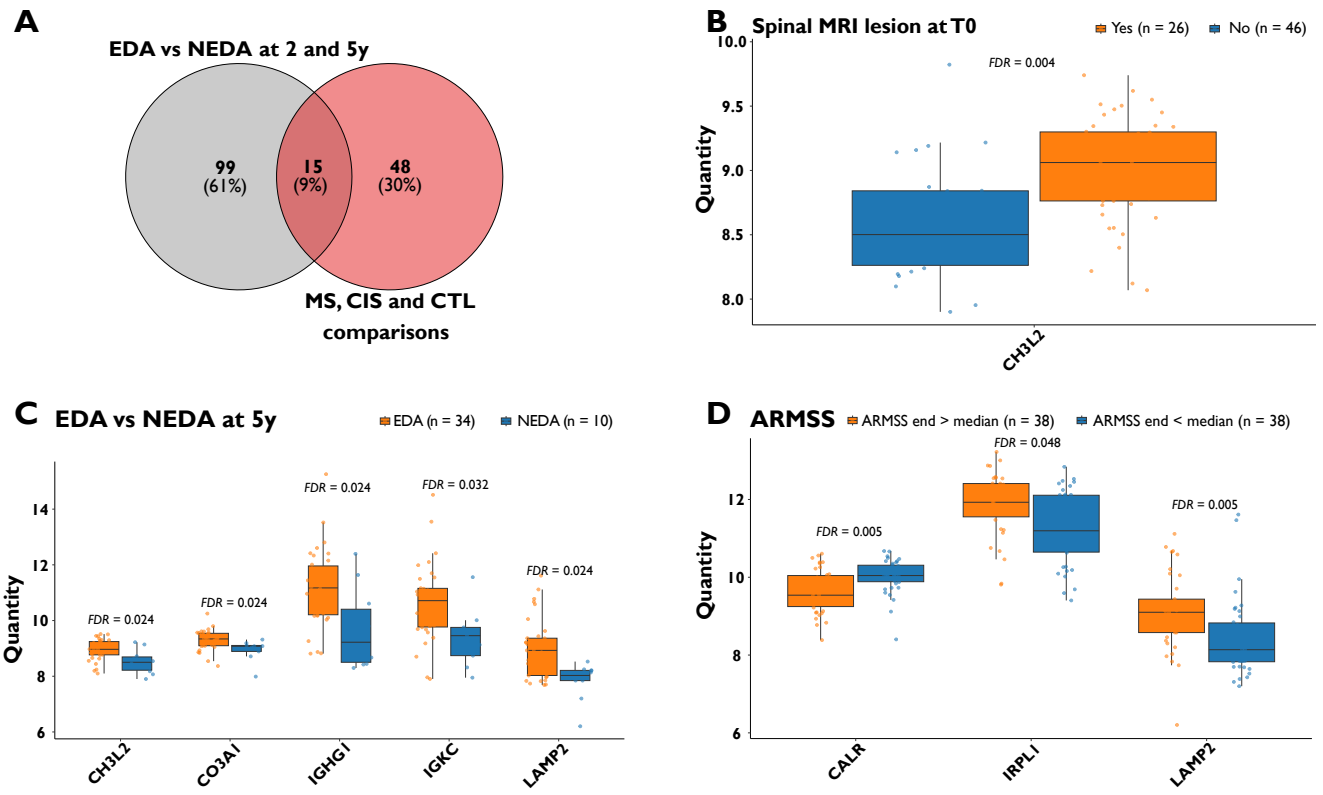
A
EDA vs NEDA at 2 and 5y
99
(61%)
15
(9%)
48
(30%)
MS, CIS and CTL
comparisons
B
Spinal MRI lesion at T0
Yes (n = 26)
No (n = 46)
FDR = 0.004
Quantity
CH3L2
C
EDA vs NEDA at 5y
EDA (n = 34)
NEDA (n = 10)
FDR = 0.024
FDR = 0.024
FDR = 0.024
FDR = 0.032
FDR = 0.024
Quantity
CH3L2
CO3A1
IGHG1
IGKC
LAMP2
D
ARMSS
ARMSS end > median (n = 38)
ARMSS end < median (n = 38)
FDR = 0.005
FDR = 0.048
FDR = 0.005
Quantity
CALR
IRPL1
LAMP2

Figure 3

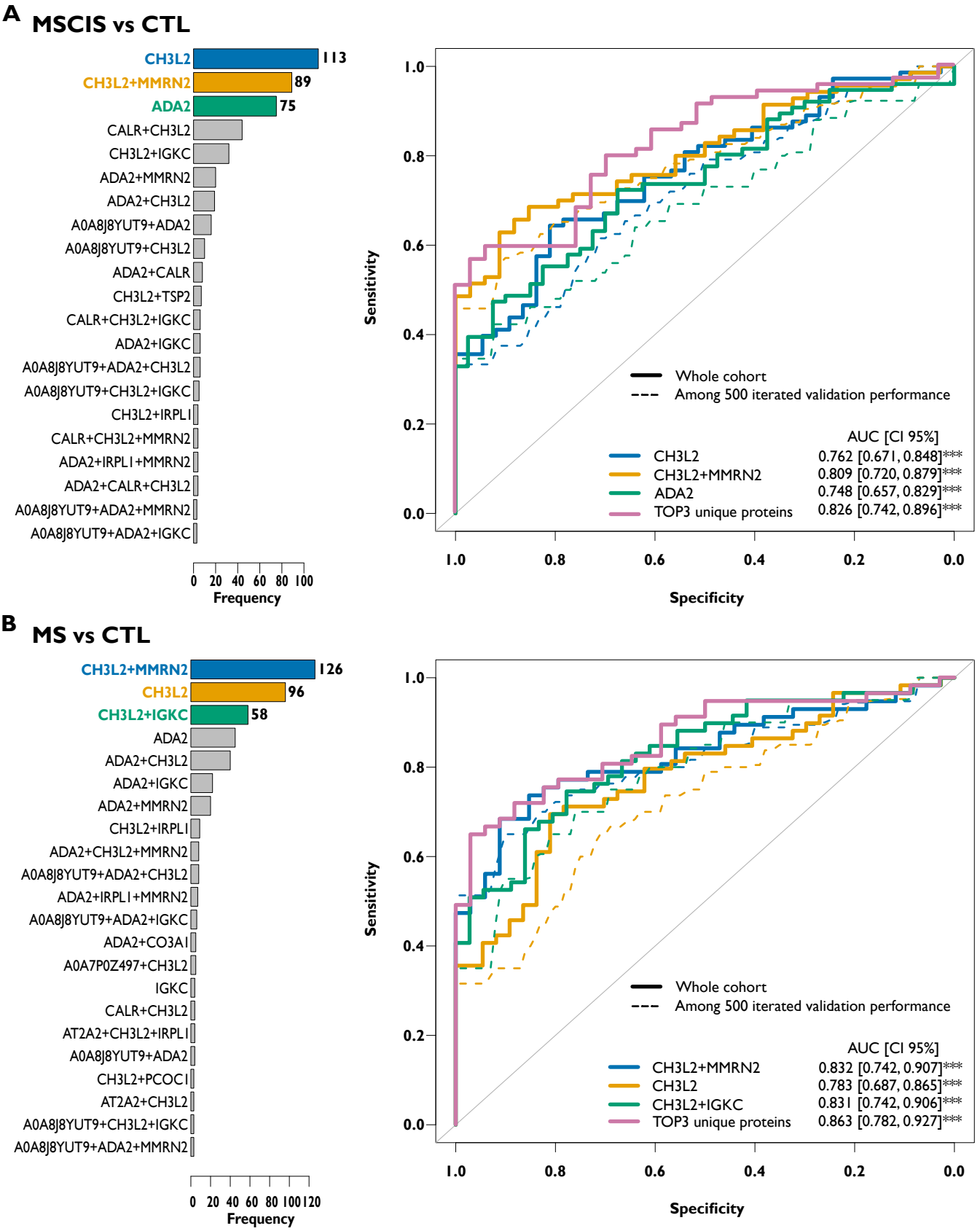
A
MSCIS vs CTL
CH3L2
CH3L2+MMRN2
ADA2
CALR+CH3L2
CH3L2+IGKC
ADA2+MMRN2
ADA2+CH3L2
A0A8J8YUT9+ADA2
A0A8J8YUT9+CH3L2
ADA2+CALR
CH3L2+TSP2
CALR+CH3L2+IGKC
ADA2+IGKC
A0A8J8YUT9+ADA2+CH3L2
A0A8J8YUT9+CH3L2+IGKC
CH3L2+IRPL1
CALR+CH3L2+MMRN2
ADA2+IRPL1+MMRN2
ADA2+CALR+CH3L2
A0A8J8YUT9+ADA2+MMRN2
A0A8J8YUT9+ADA2+IGKC
113
89
75
0 20 40 60 80 100
Frequency
Sensitivity
Specificity
1.0 0.8 0.6 0.4 0.2 0.0
Whole cohort
Among 500 iterated validation performance
AUC [CI 95%]
CH3L2 0.762 [0.671, 0.848]***
CH3L2+MMRN2 0.809 [0.720, 0.879]***
ADA2 0.748 [0.657, 0.829]***
TOP3 unique proteins 0.826 [0.742, 0.896]***
B
MS vs CTL
CH3L2+MMRN2
CH3L2
CH3L2+IGKC
ADA2
ADA2+CH3L2
ADA2+IGKC
ADA2+MMRN2
CH3L2+IRPL1
ADA2+CH3L2+MMRN2
A0A8J8YUT9+ADA2+CH3L2
ADA2+IRPL1+MMRN2
A0A8J8YUT9+ADA2+IGKC
ADA2+CO3A1
A0A7P0Z497+CH3L2
IGKC
CALR+CH3L2
AT2A2+CH3L2+IRPL1
A0A8J8YUT9+ADA2
CH3L2+PCOC1
AT2A2+CH3L2
A0A8J8YUT9+CH3L2+IGKC
A0A8J8YUT9+ADA2+MMRN2
126
96
58
0 20 40 60 80 100 120
Frequency
Sensitivity
Specificity
1.0 0.8 0.6 0.4 0.2 0.0
Whole cohort
Among 500 iterated validation performance
AUC [CI 95%]
CH3L2+MMRN2 0.832 [0.742, 0.907]***
CH3L2 0.783 [0.687, 0.865]***
CH3L2+IGKC 0.831 [0.742, 0.906]***
TOP3 unique proteins 0.863 [0.782, 0.927]***

Figure 4

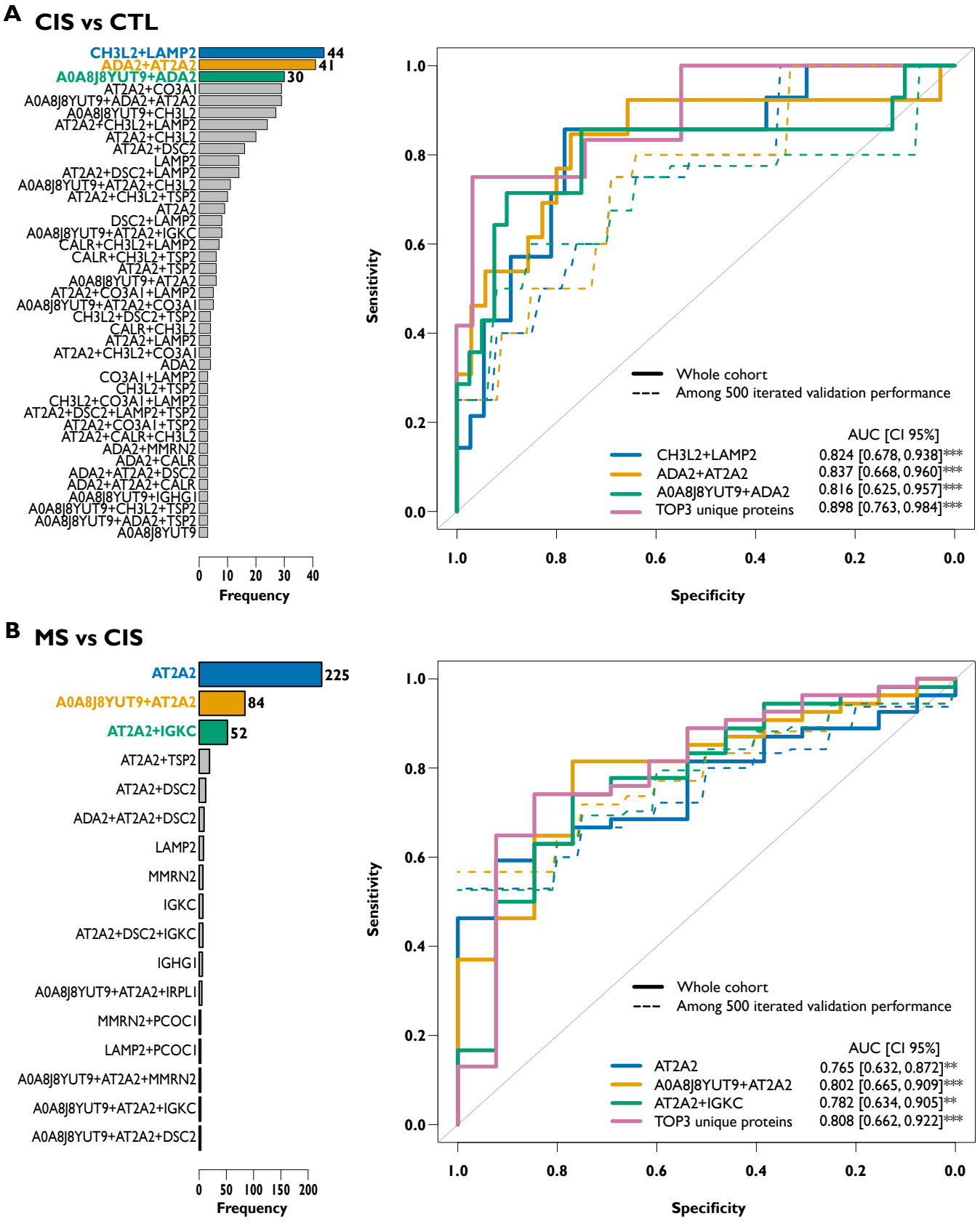
A
CIS vs CTL
CH3L2+LAMP2
ADA2+AT2A2
A0A8J8YUT9+ADA2
AT2A2+CO3A1
A0A8J8YUT9+ADA2+AT2A2
A0A8J8YUT9+CH3L2
AT2A2+CH3L2+LAMP2
AT2A2+CH3L2
AT2A2+DSC2
LAMP2
AT2A2+DSC2+LAMP2
A0A8J8YUT9+AT2A2+CH3L2
AT2A2+CH3L2+TSP2
AT2A2
DSC2+LAMP2
A0A8J8YUT9+AT2A2+IGKC
CALR+CH3L2+LAMP2
CALR+CH3L2+TSP2
AT2A2+TSP2
A0A8J8YUT9+AT2A2
AT2A2+CO3A1+LAMP2
A0A8J8YUT9+AT2A2+CO3A1
CH3L2+DSC2+TSP2
CALR+CH3L2
AT2A2+LAMP2
AT2A2+CH3L2+CO3A1
ADA2
CO3A1+LAMP2
CH3L2+TSP2
CH3L2+CO3A1+LAMP2
AT2A2+DSC2+LAMP2+TSP2
AT2A2+CO3A1+TSP2
AT2A2+CALR+CH3L2
ADA2+MMRN2
ADA2+CALR
ADA2+AT2A2+DSC2
ADA2+AT2A2+CALR
A0A8J8YUT9+IGHG1
A0A8J8YUT9+CH3L2+TSP2
A0A8J8YUT9+ADA2+TSP2
A0A8J8YUT9
44
41
30
0 10 20 30 40
Frequency
Sensitivity
Specificity
1.0 0.8 0.6 0.4 0.2 0.0
Whole cohort
Among 500 iterated validation performance
AUC [CI 95%]
CH3L2+LAMP2 0.824 [0.678, 0.938]***
ADA2+AT2A2 0.837 [0.668, 0.960]***
A0A8J8YUT9+ADA2 0.816 [0.625, 0.957]***
TOP3 unique proteins 0.898 [0.763, 0.984]***
B
MS vs CIS
AT2A2
A0A8J8YUT9+AT2A2
AT2A2+IGKC
AT2A2+TSP2
AT2A2+DSC2
ADA2+AT2A2+DSC2
LAMP2
MMRN2
IGKC
AT2A2+DSC2+IGKC
IGHG1
A0A8J8YUT9+AT2A2+IRPL1
MMRN2+PCOC1
LAMP2+PCOC1
A0A8J8YUT9+AT2A2+MMRN2
A0A8J8YUT9+AT2A2+IGKC
A0A8J8YUT9+AT2A2+DSC2
225
84
52
0 50 100 150 200
Frequency
Sensitivity
Specificity
1.0 0.8 0.6 0.4 0.2 0.0
Whole cohort
Among 500 iterated validation performance
AUC [CI 95%]
AT2A2 0.765 [0.632, 0.872]**
A0A8J8YUT9+AT2A2 0.802 [0.665, 0.909]***
AT2A2+IGKC 0.782 [0.634, 0.905]**
TOP3 unique proteins 0.808 [0.662, 0.922]***

Figure 5

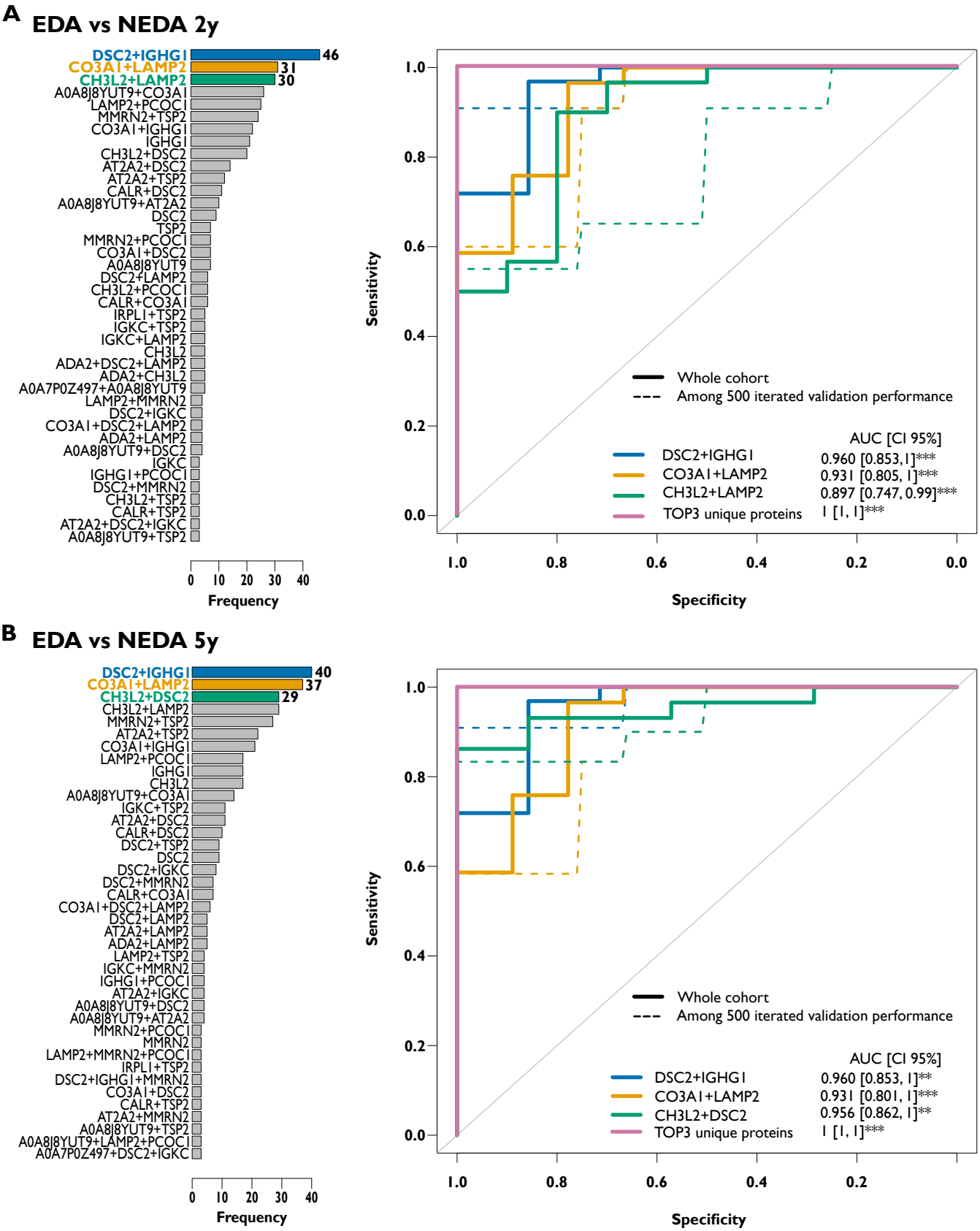
A
EDA vs NEDA 2y
DSC2+IGHG1 46
CO3A1+LAMP2 31
CH3L2+LAMP2 30
A0A8J8YUT9+CO3A1
LAMP2+PCOC1
MMRN2+TSP2
CO3A1+IGHG1
IGHG1
CH3L2+DSC2
AT2A2+DSC2
AT2A2+TSP2
CALR+DSC2
A0A8J8YUT9+AT2A2
DSC2
TSP2
MMRN2+PCOC1
CO3A1+DSC2
A0A8J8YUT9
DSC2+LAMP2
CH3L2+PCOC1
CALR+CO3A1
IRPL1+TSP2
IGKC+TSP2
IGKC+LAMP2
CH3L2
ADA2+DSC2+LAMP2
ADA2+CH3L2
A0A7P0Z497+A0A8J8YUT9
LAMP2+MMRN2
DSC2+IGKC
CO3A1+DSC2+LAMP2
ADA2+LAMP2
A0A8J8YUT9+DSC2
IGKC
IGHG1+PCOC1
DSC2+MMRN2
CH3L2+TSP2
CALR+TSP2
AT2A2+DSC2+IGKC
A0A8J8YUT9+TSP2
0 10 20 30 40
Frequency
Sensitivity
Specificity
1.0 0.8 0.6 0.4 0.2 0.0
Whole cohort
Among 500 iterated validation performance
AUC [CI 95%]
DSC2+IGHG1 0.960 [0.853, 1]***
CO3A1+LAMP2 0.931 [0.805, 1]***
CH3L2+LAMP2 0.897 [0.747, 0.99]***
TOP3 unique proteins 1 [1, 1]***
B
EDA vs NEDA 5y
DSC2+IGHG1 40
CO3A1+LAMP2 37
CH3L2+DSC2 29
CH3L2+LAMP2
MMRN2+TSP2
AT2A2+TSP2
CO3A1+IGHG1
LAMP2+PCOC1
IGHG1
CH3L2
A0A8J8YUT9+CO3A1
IGKC+TSP2
AT2A2+DSC2
CALR+DSC2
DSC2+TSP2
DSC2
DSC2+IGKC
DSC2+MMRN2
CALR+CO3A1
CO3A1+DSC2+LAMP2
DSC2+LAMP2
AT2A2+LAMP2
ADA2+LAMP2
LAMP2+TSP2
IGKC+MMRN2
IGHG1+PCOC1
AT2A2+IGKC
A0A8J8YUT9+DSC2
A0A8J8YUT9+AT2A2
MMRN2+PCOC1
MMRN2
LAMP2+MMRN2+PCOC1
IRPL1+TSP2
DSC2+IGHG1+MMRN2
CO3A1+DSC2
CALR+TSP2
AT2A2+MMRN2
A0A8J8YUT9+TSP2
A0A8J8YUT9+LAMP2+PCOC1
A0A7P0Z497+DSC2+IGKC
0 10 20 30 40
Frequency
Sensitivity
Specificity
1.0 0.8 0.6 0.4 0.2 0.0
Whole cohort
Among 500 iterated validation performance
AUC [CI 95%]
DSC2+IGHG1 0.960 [0.853, 1]**
CO3A1+LAMP2 0.931 [0.801, 1]***
CH3L2+DSC2 0.956 [0.862, 1]**
TOP3 unique proteins 1 [1, 1]***

Figure 6