Published in final edited form as:

Ann Neurol. 2017 November ; 82(5): 795–812. doi:10.1002/ana.25083.

Molecular-based diagnosis of Multiple Sclerosis and its progressive stage

Christopher Barbour^{1,2,*}, Peter Kosa^{1,*}, Mika Komori^{1,\$}, Makoto Tanigawa¹, Raturaj Masvekar¹, Tianxia Wu³, Kory Johnson⁴, Panagiotis Douvaras⁵, Valentina Fossati⁵, Ronald Herbst⁶, Yue Wang⁶, Keith Tan⁷, Mark Greenwood², and Bibiana Bielekova^{1,#}

¹Neuroimmunological Diseases Unit, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD

²Department of Mathematical Sciences, Montana State University, Bozeman, MT

³Clinical Trials Unit, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD

⁴Bioinformatics Section, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD

⁵The New York Stem Cell Foundation Research Institute, New York, NY

⁶Department of Oncology Research, MedImmune LLC, Gaithersburg, MD

⁷Translational Medicine, Neuroscience, Innovative Medicines and Early Development; AstraZeneca, Cambridge, UK

Abstract

Objective—Biomarkers aid diagnosis, allow inexpensive screening of therapies and guide selection of patient-specific therapeutic regimens in most internal medicine disciplines. In contrast, neurology lacks validated measurements of the physiological status, or dysfunction(s) of cells of the central nervous system (CNS). Accordingly, patients with chronic neurological diseases are often treated with a single disease-modifying therapy without understanding patient-specific drivers of disability.

Therefore, using multiple sclerosis (MS) as an example of a complex polygenic neurological disease, we sought to determine if cerebrospinal fluid (CSF) biomarkers are intra-individually stable, cell type-, disease- and/or process-specific and responsive to therapeutic intervention.

[#]To whom correspondence should be addressed: Bibi Bielekova, MD, Neuroimmunological Diseases Unit (NDU), National Institute of Neurological Disorders and Stroke (NINDS), National Institutes of Health (NIH), Building 10, Room 5C103, Bethesda, Maryland 20892, USA (Bibi.Bielekova@nih.gov).

^{*}Equal contribution

^{\$}MK contributed to this work as a former employee of NINDS, and the opinions expressed in this work do not represent her current affiliation - Eli Lilly Japan K.K., Kobe, Japan

Author contributions: study concept and design: B.B.; data acquisition and analysis: C.B., P.K., M.K., M.T., R.M., T.W., K.J., P.D., V.F., R.H., Y.W., K.T., M.G. and B.B.; drafting the manuscript and figures: C.B., P.K. and B.B.

Potential conflicts of interest: BB, PK, MK, CB, and MG are co-inventors of US patent application number 62/038,530: Biomarkers for Diagnosis and Management of Neuro-immunological Diseases, which pertains to the results of this paper. BB, PK, and MK have assigned their patent rights to the US Department of Health and Human Services.

Methods—We used statistical learning in a modeling cohort (n=225) to develop diagnostic classifiers from DNA-aptamer-based measurements of 1128 CSF proteins. An independent validation cohort (n=85) assessed the reliability of derived classifiers. The biological interpretation resulted from in-vitro modeling of primary or stem cell-derived human CNS cells and cell lines.

Results—The classifier that differentiates MS from CNS diseases that mimic MS clinically, pathophysiologically and on imaging, achieved a validated area under receiver-operator characteristic curve (AUROC) of 0.98, while the classifier that differentiates relapsing-remitting from progressive MS achieved a validated AUROC of 0.91. No classifiers could differentiate primary-from secondary-progressive MS better than random guessing. Treatment-induced changes in biomarkers greatly exceeded intra-individual- and technical variabilities of the assay.

Interpretation—CNS biological processes reflected by CSF biomarkers are robust, stable, and disease- or even disease-stage specific. This opens opportunities for broad utilization of CSF biomarkers in drug development and precision medicine for CNS disorders.

Introduction

Biomarkers play a critical role in diagnostic and therapeutic decisions in many areas of internal medicine. Cell specific analytes (such as liver function tests) provide essential information about functionality in their cells of origin and represent the basis of molecular diagnosis. Molecular dissection of complex disorders allows selection of optimal, individualized therapy. Such “precision” therapy consists of simultaneous application of (multiple) drugs that collectively target all pathological processes that underlie expression of a disease in particular patient.

In contrast, neurologists lack tools that provide reliable information about the dysfunction of constituent cells of the CNS. This ambiguity leads to 20–40% diagnostic errors (1, 2), slow therapeutic progress (3) and suboptimal clinical outcomes. Complex neurological disorders such as multiple sclerosis (MS) are generally treated by a single disease modifying treatment (DMT), without understanding patient-specific drivers of disability. The multiplicity of mechanisms in neurodegenerative diseases and heterogeneity within patient populations makes successful treatment by a single therapy unlikely. Conversely, proving clinical efficacy of a single therapy is difficult precisely because of limited contribution of the targeted mechanism to the overall disease process.

Thus, reliable quantification of diverse pathogenic processes in the CNS of living subjects is a prerequisite for broad therapeutic progress in neurology. Although cerebrospinal fluid (CSF), an outflow for CNS interstitial fluid (4) is an ideal source for molecular biomarkers, remarkably few CSF biomarkers have reached clinical practice or drug development (5). This reality is partly based on a circular argument: CSF examinations are not implemented in clinical trials or clinics because of a lack of validated, commercially-available biomarker measurements, while reliable data on surrogacy of biomarkers to clinical outcomes can be obtained only from clinical trials or wide clinical use.

Consequently, the goal of this proof-of-concept study was to investigate on the example of MS the following hypotheses: 1. A subset of CSF biomarkers are intra-individually stable in

the absence of disease process or therapeutic intervention, and such biomarkers can be assembled into clinically useful tests; 2. A subgroup of CSF biomarkers have restricted cellular origin and can be used to develop clinically-useful classifiers; 3. Healthy and different disease states of the CNS are sufficiently dissimilar on a molecular level that CSF biomarker-based classifiers can differentiate a specific disease from those that have similar clinical phenotype, pathophysiology, or imaging features; 4. CSF biomarker-based classifiers can also quantify evolution of a single disease process, thus differentiating its stages; and 5. Therapy-induced changes in CSF biomarkers can be readily distinguished from intra-individual variability, demonstrating that CSF biomarkers could serve as pharmacodynamic markers in drug development.

Methods

Subjects

Subjects were prospectively recruited (5/2009–3/2015) as part of a Natural History protocol “Comprehensive Multimodal Analysis of Neuromimmunological Diseases of the Central Nervous System” (ClinicalTrials.gov Identifier: NCT00794352). The patients’ eligibility criteria included age 18–75 years and presentation with a clinical syndrome consistent with immune-mediated CNS disorder, or neuroimaging consistent with inflammatory or demyelinating CNS disease. The inclusion criteria for healthy donors (HD) were age 18–75 years and vital signs within normal range at the time of the screening visit. The diagnostic workup included a neurological exam, MRI of the brain and laboratory tests (blood, CSF) as described (6). Diagnoses of relapsing-remitting MS (RRMS), primary progressive MS (PPMS) and secondary progressive MS (SPMS) were based on 2010 revised McDonald diagnostic criteria (7). The remaining subjects were classified as either other inflammatory neurological disorders (OIND; e.g., meningitis/encephalitis, Susac’s Syndrome, CNS vasculitis, Systemic Lupus Erythematosus and genetic immunodeficiencies with CNS inflammation) or non-inflammatory neurological disorders (NIND; e.g., epilepsy, vascular/ischemic disorders, leukodystrophy) based on the evidence of intrathecal inflammation as published (6, 8). The final clinical diagnostic classification was based on longitudinal follow-up, but reached prior to development of SOMAscan-based diagnostic classifiers. The vast majority of subjects (with few OIND exceptions described elsewhere (6)) were not treated by any disease-modifying treatments (DMT) at the time of CSF collection.

Clinical information of the validation cohort (n=85) were not available to developers of the diagnostic classifiers, while the results of the molecular diagnostic tests were not available to clinicians determining diagnoses.

CSF collection and processing

CSF was collected on ice and processed according to a written standard operating procedure. Research CSF aliquots were assigned prospective alpha-numeric codes, and centrifuged (335 g for 10 minutes at 4°C) within 15 minutes of collection. The CSF supernatant was aliquoted and stored in polypropylene tubes at –80 °C until use.

SOMAScan

SOMAScan (SomaLogic, Inc., Boulder, CO) is a relative quantification of 1128 proteins (i.e., SOMAScan version available between 6/2012 and 10/2016; (9)) or 1300 protein (i.e., SOMAScan version available after 10/2016) using single-stranded DNA molecules synthesized from chemically modified nucleotides (SOMAmers; Slow Off-rate Modified DN Aptamers). Chemical modifications enhance affinity binding to specific proteins. SOMAmers play a dual role of protein affinity-binding reagents and a DNA sequence recognized by complementary DNA probes. This enables quantification of individual protein concentration using a DNA concentration quantified by hybridization (10–12). The raw data (relative fluorescent units [RFU]) are normalized and calibrated: hybridization normalization uses a set of twelve hybridization controls and a common pooled calibrator corrects for plate-to-plate variation. SOMAScan focuses on secreted soluble proteins, using a single discovery platform for all research applications.

Assessment of cellular origin of tested biomarkers from a subset of human immune cells and CNS cells

Fresh peripheral blood mononuclear cells (PBMC) of two HD were obtained from Ficoll gradient-treated lymphocytophoresis samples. Monocytes, B-cells, CD4+ T cells, CD8+ T cells, natural killer cells, innate lymphoid cells, and myeloid dendritic cells were Fluorescence-activated cell sorted (FACS). Each purified cell subtype was cultured (1×10^6 cells/ml) in serum-free X-VIVO 15 medium (Lonza, Walkersville, MD) with or without 10 μ g/mL phorbol 12-myristate 13-acetate (PMA) and 1 μ M Ionomycin. Supernatants were collected after 48 hours and frozen until use.

Isolated primary human CNS cells or cell lines: human neurons (ScienCell, Carlsbad, CA), human astrocytes (ScienCell), human brain endothelial cell line (HCMEC/D3; provided by Pierre-Olivier Couraud, PhD, INSERM, France (13)), human microglia cell line (CHME5; provided by Nazira El-Hage, PhD, Florida International University, USA), and human choroid plexus epithelial cells (hCPEpiC; ScienCell) were plated (10^5 cells/ml; 10 ml/flask). Cells were treated with PBMC culture media (control) and an inflammatory mediators (supernatant from lipopolysaccharide- and CD3/CD28 beads-stimulated human PBMCs; 50% v/v). Oligodendrocytes were differentiated from the NIH approved human embryonic stem cell line RUES1 using published protocol (14). Cell-culture supernatants were collected after 24-hour incubation and frozen until use.

Measuring signal-to-noise ratio (SNR) in biomarkers

Differences in biomarker measurements in identical samples analyzed blindly at different times ($n=29$; 88 samples) quantified the technical variability. Similarly, differences in longitudinal samples of HDs ($n=11$; 24 samples) collected ~1 year apart quantified the biological variation. From each technical or biological replicate, we calculated the relative percent change for each SOMAmer as the difference in the repeated measures divided by the average of the replicates.

To constrain the number of biomarkers used for statistical learning, we calculated signal-to-noise ratios (SNR) using R statistical software (15) as exemplified in Fig 1A. Briefly, we

estimated the residual variance for each log-transformed biomarker from a linear mixed model (16), after accounting for subject-to-subject variation using the random intercept adjustment. The residual variance measured when identical samples were analyzed repeatedly, represents “technical” variation (i.e., variation caused by differences in assay runs). Analogously, the variance measured in multiple different CSF samples derived from individual HDs represents “biological” variation (i.e., intra-individual variation in healthy state). A third type of variance, which we call “clinical”, reflects how individual biomarkers vary in the presence of different disease states (e.g., variation across all disease states). This variance was estimated by using one observation for each biomarker from each of the $n=225$ subjects in the training dataset. The SNR was calculated as the clinical variance divided by the sum of all (the clinical, biological, and technical) variances. Thus, SNR reflects the proportion of the total variability that is attributable to variation among subjects from different diagnostic categories. In other words, biomarkers with high SNR (i.e., values close to 1) are biomarkers with large differences across subjects in the training cohort, but with low variation in longitudinal sampling of healthy people that can be detected with minimal variations across different assay runs. These markers have a high diagnostic potential. In contrast, biomarkers with SNR close to 0 represent proteins that are either not detectable in the CSF, or offer low diagnostic value because the physiological (intra-individual) variation or assay noise are comparable to the difference between diagnostic categories we desire to measure. An analogous SNR procedure restricted the number of biomarker *ratios* used for statistical learning.

Area under the Receiver Operating Characteristics Curve (AUROC)

The AUROC (R statistical software using the roc function in the pROC package (17)) quantified the ability of biomarkers, biomarker ratios, and diagnostic classifiers to differentiate diagnostic categories. Higher AUROC values imply a larger potential for separating diagnostic groups. The AUROCs were calculated for the 124750 ratios formed from the top 500 SNR SOMAmers in the modeling cohort ($n=225$) and these were used to restrict the number of biomarker ratios used for statistical learning. Graphical exploration of the distributions of the SNR and AUROCs were used to determine cutoffs for the best ratios on the two criteria in each situation. For differentiating MS (RRMS, PPMS, SPMS) from non-MS subjects (OIND, NIND, HD), cutoffs of AUROC >0.65 and SNR >0.75 were selected. Similarly, cutoffs of AUROC >0.7 and SNR >0.7 were used in differentiating progressive (PPMS, SPMS) MS from RRMS. Finally, cutoffs of AUROC >0.65 and SNR >0.65 were used for differentiating SPMS from PPMS. Ratios meeting these cutoffs were used in statistical learning to assemble diagnostic classifiers.

The performance of resulting classifiers was assessed by AUROC, along with its 95% bootstrapped confidence interval (CI), in the independent validation cohort.

Statistical learning to develop diagnostic classifiers using a Random Forest methodology

Random forests (18) (randomForest R package (19); see description in Fig 2) were built by sequentially estimating multiple classification trees (between 500–1500; selected based on the stability of the out-of-bag [OOB] error; Fig 2) using bootstrapped training cohort samples ($n=225$) with a random subset of predictors for each node. The trees in the “forest”

are averaged together, providing more reliable predictions than are possible using a single classification tree. Biomarkers and biomarker ratios were natural log-transformed prior to classifier construction. Variable importance measures (average decrease in accuracy from permuting each covariate across all trees) (20, 21) evaluated the contribution of individual biomarkers to the classifier. The R code for constructing these classifiers is provided as Supplementary file 1.

Results

SOMAscan Assay on CSF samples

Using SOMAscan we analyzed CSF samples in blinded fashion in two independent cohorts of subjects (Table 1): 1. A training cohort (n=225) consisted of untreated subjects from six diagnostic groups (RRMS, n=40; PPMS, n=40; SPMS, n=40; NIND, n=39; OIND, n=41; and HD, n=25). 2. An independent validation cohort (n=85 untreated subjects) consisted of 14 subjects per non-MS diagnostic categories (OIND and NIND), 16 subjects per RRMS and PPMS group, 15 SPMS subjects and 10 HD. The raw SOMAscan data for both cohorts are available in Supplementary table 1.

In addition, to assess the technical and biological variability of the SOMAscan, we analyzed 88 CSF samples representing technical replicates (identical CSF aliquots analyzed at different time-points) and 24 longitudinal CSF samples collected from 11 HD over one-year span, serving as biological replicates. The average technical and biological relative percent change (Fig 1B and Supplementary table 2) were 11.9% and 12.9%, respectively. To determine the effect of immunomodulatory DMT on SOMAmers, we analyzed 10 longitudinal CSF samples from five OIND patients collected before and after therapy with high dose methylprednisolone. The resulting average relative percent change of 52.1% exceeded the highest technical and biological relative percent change. Therefore, we concluded that the SOMAscan reliably measures CSF biomarkers and can detect an effect of DMT in subjects over time.

Development of diagnostic tests for MS

Considering the complex biological mechanisms underlying MS, a single biomarker cannot reliably differentiate MS from all other CNS diseases. Machine learning strategies (18), such as random forests (19), combine multiple biomarkers using a statistical algorithm that enhances sensitivity and specificity (22), resulting in clinically-useful classifiers (See Fig 2 for explanation of the random forest algorithm).

We employed random forests trained using the modeling (n=225) cohort to develop three classifiers: 1. One that differentiates MS from all other diagnostic groups, 2. One that differentiates RRMS from progressive MS (i.e., PPMS+SPMS) and, 3. One that differentiates PPMS from SPMS. Because the validity of the classifier must be tested in an independent cohort not used for the model construction (23), we assessed the performance of the diagnostic tests in the independent (n=85) cohort by predicting AUROCs and their CIs (Fig 3).

Classifiers that used all 1128 SOMAmers achieved validated AUROC=0.91 (CI: 0.84 – 0.97) for the MS versus non-MS test, AUROC=0.73 (CI: 0.57–0.90) for the RRMS versus progressive MS test and AUROC 0.64 (CI: 0.44–0.84) for SPMS versus PPMS test (Fig 3).

We expected that not all 1128 proteins measured by SOMAscan will be detectable in the CSF. We filtered out noise stemming from the poorly-detectable biomarkers, by restricting the number of SOMAmers to the 500 with the highest SNR (see Methods for details). We reasoned that the most useful biomarkers will vary greatly among subjects from different diagnostic categories, whereas they will have stable physiological levels (i.e., they will have low variance in biological replicates of HDs) and can be measured with high precision (i.e., low variance in technical replicates). This reduced set of 500 biomarkers (Supplementary table 3) improved performance of the classifiers: AUROC for the RRMS versus progressive MS classifier increased to 0.80 (CI: 0.65–0.95); the performance of the diagnostic test for MS marginally improved (AUROC=0.92, CI: 0.86–0.98), and remained unchanged for SPMS versus PPMS test (Fig 3).

While biomarkers are secreted by diverse cells under physiological or pathological states, many biomarkers are biologically related; e.g., they physically interact or belong to the same network. However, random forests consider biomarkers only *sequentially* within any given tree. For related biomarkers, such as receptor-ligand pairs, the pathogenic process may depend more on their stoichiometry than on their respective concentrations. Therefore, we hypothesized that considering related biomarkers *simultaneously*, for example as ratios, will add discriminatory value. Mathematically, this corresponds to broadening the biomarker-based random forests from partitioning the predictor space based on absolute concentrations of individual markers (represented by blue dotted perpendicular lines in Fig 2C) to considering predictors built from the relative proportion of biomarkers (represented by diagonal orange line in Fig 2C).

Consequently, we used the 500 high-SNR SOMAmers to generate 124750 biomarker ratios. To build random forests only from ratios with highest clinical utility, we combined AUROC with SNR values in the modeling (n=225) cohort, selecting logical cut-offs described in Methods that allowed for enough diversity to capture the biological processes while limiting the dimension of the search space. This led to 5401 retained ratios for MS versus non-MS, 3626 retained ratios for progressive versus relapsing MS, and 1504 retained ratios for SPMS versus PPMS. Using these sets of ratios strongly enhanced the performance of random forest models distinguishing MS from non-MS and RRMS from progressive MS (validated AUROC=0.95; CI: 0.91–0.99 and AUROC=0.88; CI: 0.76–1.00, respectively). However, the performance of SPMS versus PPMS diagnostic test remained low (AUROC=0.45; CI: 0.24–0.67).

A clinical test has to fulfill technical requirements of reproducible measurement across different laboratories, often achieved by using standard curves; this makes measuring several hundreds of proteins prohibitive. Therefore, we sought to identify the smallest number of biomarkers that can be assembled into the random forest classifiers without a significant loss of accuracy. To achieve this, we examined the out-of-bag (OOB) AUROC estimates from random forests generated from the modeling (n=225) cohort by sequentially adding ratios

with the highest variable importance (Fig 2D). The point of inflection where the OOB AUROC appeared to stabilize was selected to achieve models with high predictive ability at a lower complexity (Fig 3). Interestingly, the reduction of the number of ratios further improved the performance of the classifiers in the validation cohort: the 22 most important ratios in the MS vs non-MS classifier led to validated AUROC=0.98 (CI: 0.94–1.00) and the 21 most important ratios distinguished RRMS from progressive MS with AUROC=0.91 (CI: 0.80–1.00) (Fig 3). The 33 most important ratios distinguishing SPMS from PPMS led to a classifier with performance comparable to random guessing (AUROC=0.58; CI:0.37–0.79) and therefore this model was abandoned from further analyses.

In all three classifiers, variable importance measures were dominated by ratios in models that also included single SOMAmers (data not shown), validating our mathematical and biological foundation of the ratio-based hypothesis.

The clinical properties of the validated models are summarized in Figure 4. When using 50% cut-off to convert continuous probabilities into dichotomous classifiers, the MS molecular diagnostic test shows 87.2% sensitivity (CI: 77.7%–96.8%) and 94.7% specificity (CI: 87.6%–100.0%) with a diagnostic odds ratio of 123.0. The progressive MS classifier differentiates RRMS from progressive MS with 93.5% sensitivity (CI: 84.9%–100.0%) and 81.3% specificity (CI: 62.1%–100.0%), reaching a diagnostic odds ratio of 62.8.

Deconvolution of biomarkers' cell of origin

To investigate whether statistical learning preferentially selected biomarkers with restricted cellular origin into clinically-useful tests, we used in-vitro modeling on human primary immune and CNS cells (see Methods), complemented with data from the public domain, such as the RNA sequencing database of human CNS cells (24, 25) and The Human Protein Atlas (26) to assess cellular origin of the biomarkers in the optimized random forest classifiers (Fig 5–6). The results of SOMAscan analysis of supernatants from human CNS cell lines and from freshly sorted human peripheral immune cells are depicted in Supplementary table 4. We analyzed cell cultured media in resting state and upon activation: CNS cells/cell lines were exposed to supernatants from LPS- and CD3/CD28 beads-stimulated human PBMCs to mimic inflammatory conditions, while immune cells were activated by PMA/Ionomycin to achieve robust activation of all immune cells. Supernatants were analyzed by SOMAscan assay at time 0 and after 24-hour incubation/stimulation. Results are shown as the stimulation index - a ratio of RFUs at 24 hour and time 0 for each condition for 48 SOMAmers that form the two diagnostic classifiers. These results are the basis for the biological interpretation of the diagnostic classifiers.

Biological interpretation of MS versus non-MS diagnostic test

The 22 most important ratios of the MS diagnostic test were dominated by immune cell-specific biomarkers (Fig 5). Twenty-one of these contain plasma cell-specific biomarkers TNFRSF17 (BCMA) or IGG. The main insight from the MS diagnostic classifier is that in MS, activation of humoral immunity represented by plasma cells and plasmablasts is out of proportion to activation of all other cellular components of innate or adaptive immunity. Combining cellular origin and known biological functions, the biomarkers constituting the

MS diagnostic test can be divided into 9 subgroups, where plasma cell activation/levels is compared to: overall intrathecal inflammation (Fig 5; group 1a; PDCD1LG2, SLAMF6, CD48, CSF3, CXCL13, TNFRSF4) and amount of activation of myeloid lineage (Fig 5; groups 1b and 1c; PLA2G7, CCL7, TLR4 LY96, PRTN3). MS patients also show higher plasma cell activation/levels in comparison to vascular injury and/or ongoing CNS stress (Fig 5; group 1d, 2a, and 2c; FLT4, CDKN1B, TNFRSF6B, DSG2, CRK, PGK1, MAPK14, F9, DCTPP1). A ratio of IgG and IgM (Fig 5; group 2b) points to differences in immunoglobulin subtypes between MS and non-MS subjects. Higher plasma cell immunoglobulin secretion in comparison to levels of CNS injury and higher plasma cell activation in comparison to astrocyte activation (Fig 5; group 2d; TNC) was also observed in MS. Lastly, MS subjects have increased epithelial stress compared to activation of neutrophils (Fig 5; group 3; MMP7, PRTN3).

Biological interpretation of progressive MS versus RRMS diagnostic test

The 21 top ratios distinguish RRMS from progressive MS (Fig 6). Seven were ratios of EDA2R or EDAR with markers released predominantly by CNS cells, especially neurons and oligodendrocytes: STX1A, EPHA5, JAM3, NTRK3, RGMA, BOC, and UNC5 (Fig 6; group 1a); all of these ratios demonstrate relative loss of CNS-specific markers in progressive MS. Moreover, three ratios show proportional loss of neuronal and oligodendroglial markers in relation to ICOSLG, expressed on activated antigen-presenting cells, in progressive MS (Fig 6; group 1b; EPHA5, JAM3, TYRO3). Similarly, the ratio INHBA/JAM3 measures relative loss of an oligodendroglial marker in comparison to a marker secreted predominantly by myeloid cells during wound healing and tissue remodeling (Fig 6; group 1c). Progressive MS patients also show increased epithelial stress in comparison to overall intrathecal inflammation (Fig 6; group 1d; SELL, EDAR). Another group of biomarker ratios points out enhanced alternative pathways of complement activation in comparison to overall intrathecal inflammation (Fig 6; groups 2a and 2b; CFD, GZMA, SELL, SERPING1, IL22). Finally, a group of ratios relates to dysregulation of LTA/LTB and IL22 pathways, which play an important role in the formation of tertiary lymphoid follicles in progressive MS (Fig 6; groups 2c, 2d, 2e, and 3; LILRB2, SELL, LTA, LTB, IL10, PRTN3, ETHE1, GP6, CLEC1B) and also in platelet aggregation (Fig 6; group 3).

Discussion

Exploring recent advances in proteomics, we asked whether CSF biomarkers can reliably measure intrathecal processes and thus facilitate diagnosis, drug development, and clinical management of patients with complex CNS diseases. We hypothesized and confirmed that intra-individually stable CSF biomarkers with restricted cellular origin are over-represented in clinically-useful classifiers. Collected data provide proof-of-principle evidence that molecular diagnosis of polygenic CNS diseases is feasible with current technologies.

In contrast to internal medicine disciplines that utilize molecular biomarkers (27), contemporary diagnostic process and therapeutic decisions for polygenic neurological diseases are based on clinical findings and structural imaging, both of which lack molecular

specificity. This may contribute to high misclassification rate in neurodegenerative diseases against pathology (1, 2). Indeed, finding that one out of the first three “MS” subjects who succumbed to natalizumab-induced progressive multifocal leukoencephalopathy demonstrated no pathological evidence of MS (28) suggests that a >20% misdiagnosis rate may also be applicable to MS, despite advances provided by imaging. While we observed approximately 10% discrepancy between clinical- and biomarker-based MS diagnosis (Fig 7A), the absence of pathological evidence prevents determining which classification is correct. The majority of SOMAscan-misclassified “MS” patients lacked defining biological features of MS: intrathecal activation of plasma cells and adaptive immunity, validated by alternative assays (Fig 7B). Therefore, patients misclassified by molecular classifiers either exhibited a non-inflammatory form of MS, observed by pathologists at frequencies analogous to our MS misclassification rate (29), or had alternative conditions. Regardless of what we call such ailments, these patients lack targets of immunomodulatory DMTs and are unlikely to reap their benefit. Thus, providing therapeutically-relevant information represents the first advantage of molecular diagnosis. The second advantage is reporting diagnostic probabilities as a continuous variable that captures the strength of biological evidence, in comparison to a dichotomous clinical diagnosis. Indeterminate results close to 50% probability (which represented 88% of subjects with discrepant clinical and molecular diagnosis; Fig 7A) should be repeated, ideally after cessation of the acute process that prompted diagnostic testing.

Diagnostic specificity is of high clinical importance, because false positive results expose subjects to potential harms of unnecessary therapies. Because the MS diagnostic classifier is antigen-nonspecific, dysregulated immunity targeting non-MS antigens may be misclassified as MS if it elicits qualitatively similar intrathecal inflammation. Indeed, we observed that two OIND patients (one with CTLA-4 haploinsufficiency and another with chronic aseptic meningitis) were misclassified as MS. It is plausible that analogously to mutations shared among different cancers, conditions with pathogenic mechanisms similar to MS may respond to MS treatments. This is the third advantage of molecular taxonomy, as it could promote CNS therapeutics from disease-specific monotherapies to process-specific therapies, where treatments are shared among pathophysiologically-related conditions and rationally assembled into patient-specific polypharmacy regimens.

Our study has following limitations: 1. SOMAscan represents a selection of proteins that are not specifically targeted to the CNS. This drawback, however, also proves that molecular signatures of distinct diseases are sufficiently robust that sampling ~1% of the relevant proteome can reliably differentiate among them. Our observation that unbiased statistical learning selected virtually all available SOMAmers with restricted CNS cellular origin suggests that deliberate broadening of the sampled proteome to more CNS-relevant biomarkers has a potential to improve classifications, and expand understanding of disease mechanisms. 2. SOMAscan is a discovery platform, routinely optimized and expanded, and therefore lacking standards of clinical applications. We dealt with this problem by embedding many technical replicates that allowed normalization between different assay runs. However, even after normalizing and focusing on biomarkers with high SNR, the inter-assay variability (measured by technical replicates) decreased the performance of the classifiers. Furthermore, during the submission and review of this article, Somalogic updated

SOMAscan from an assay that quantifies 1128 proteins to an assay that uses different buffers, dilutions, and quantifies 1300 proteins. We subsequently used the original validation cohort to test whether random forests constructed from the selected set of biomarker ratios depicted in Figures 5 and 6 can still reliably differentiate MS from other diseases and RRMS from progressive MS using the new version of SOMAscan. We observed OOB AUROC of 0.89 (CI: 0.82–0.97) for MS versus non-MS classifier and OOB AUROC of 0.84 (CI: 0.67–1.00) for RRMS versus progressive MS classifier. This shows that our results are not assay-dependent, but reflect true biological processes. Nevertheless, biomarker-based precision neurology cannot be achieved without the biotech industry, which needs to develop fully quantitative, CSF-targeted assays that conform to technical requirements of clinical tests. To facilitate this, we considered “assay economy”, as an optimum between assay cost (dependent on the number of proteins that need absolute and relative quantification) and accuracy. Biomarker ratios simplify assay commercialization, by limiting the need to run standard curves for every analyte and providing internal normalization that avoids false positive results caused by, e.g., high protein levels. However, absolute quantification of at least the dominant biomarker partners, such as BCMA and EDA2R will likely be necessary for quality assurance for clinical applications.

The creative use of SNR screened out biomarkers of low clinical value; this improved the efficacy of the random forest algorithm, which searched for optimal biomarkers in the lower dimensional search space. Although such use of external data and domain-expert knowledge is encouraged in statistical learning as it typically improves performance (as it did in our case), we acknowledge that this methodology leads to some arbitrariness in the selection of markers and thus it may not work in other settings or for other cohorts. To assure adequate representation of all MS subtypes and both inflammatory and non-inflammatory MS mimics in the classifier construction, we designed this study to have approximately equal representation from all patient groups. This may not be representative of the rates in the real population of patients, where e.g., RRMS/SPMS patients are much more frequently encountered than PPMS subjects. Because SNR is dependent on the composition of the training cohort, we acknowledge that different compositions of the training cohort may lead to selection of different biomarkers, potentially better or worse for separating certain disease states. This behavior is inherent to any statistical learning process and, therefore, sampling and population structure must be considered carefully in the study design. We view our population selection as appropriate for the stated goals, because in addition to healthy donors, controls included subjects with varied inflammatory and non-inflammatory CNS diseases, who presented for the diagnostic work-up of MS or related neuroimmunological diseases and who must be differentiated from all three MS subtypes.

Supporting the notion that CSF biomarkers can expand understanding of CNS diseases, the following knowledge was gained from the current study: The essential difference between MS and its mimics is selective expansion/activation of B cell/plasma-cell lineages, out of proportion to the activation of other immune cells and to the resultant injury/stress of CNS-resident cells. An ancillary pathway that helps to diagnose MS is linked to a marker of tissue remodeling and repair, MMP7. These features are shared by all MS subtypes, indicating that PPMS is not a pathophysiologically-distinct “non-inflammatory” entity (30), but rather an equivalent disease stage to SPMS. This conclusion is supported by the observed inability to

validate a molecular classifier that differentiates PPMS from SPMS with accuracy higher than random guessing and by therapeutic response of PPMS to immunomodulation by ocrelizumab (31).

The dominance of plasma cell biomarkers in the molecular classifier poses a question of its value against current CSF tests such as IgG index and oligoclonal bands (OCB). We have included IgG index and MS classifier prediction rates to Figure 7C to demonstrate the superiority of the classifier. Similar data were obtained for OCB; in the cohort of patients with available OCB data the sensitivity (93.9%; CI: 90.3% – 97.6%) of OCB test was comparable to the sensitivity of molecular classifier (96.4%; CI: 93.5% – 99.2%), the specificity of the OCB test (80.0%; CI: 72.8% – 87.2%) was highly outperformed by the specificity of the molecular classifier (98.3%; CI: 96.0% – 100.0%).

Statistical learning also enhanced our understanding of progressive MS by demonstrating that PPMS and SPMS are biologically indistinguishable. These data argue for merging PPMS and SPMS cohorts in future drug development and clinical considerations. Features that differentiate progressive MS from RRMS are greater CNS tissue destruction, including more wide-spread endothelial/epithelial cell stress and reactive gliosis with increased permeability of CNS barriers and greater activation of innate immunity. In addition to proportional loss of oligodendroglial and neuronal biomarkers that likely reflect injury or loss of their cells of origin, there are also immunological differences between RRMS and progressive MS. These relate to innate immunity (complement, myeloid lineage, and antigen presentation), and to pathways involved in the formation of tertiary lymphoid follicles, such as lymphotoxin complex and IL-22. This is consistent with the pathological evidence of tertiary lymphoid follicles in progressive MS (32) and with a recent report that the level of compartmentalization of immune responses to the CNS can differentiate RRMS from two progressive MS subtypes (6). Finally, it is intriguing that 4 out of 21 ratios that differentiate progressive MS from RRMS (i.e., containing SERPING1 and CFD, which is essential in alternative complement activation by cleaving C3) are linked to “neurotoxic reactive astrocytes”, recently shown to mediate neuronal death in MS and other neurodegenerative diseases (33).

One may ask to what degree the identified MS progression-specific processes reflect aging. Re-analyzing probabilities of progressive MS in patients younger and older than 45 years demonstrated that the molecular classifier correctly differentiates RRMS from progressive MS irrespective of age (Fig 7D). Thus, the biological interpretation of MS classifiers offers the following unifying hypothesis for future longitudinal studies: while aberrant activation of B/plasma cell lineage is essential for development of MS, the complex response of CNS tissue, exemplified by microglial activation, toxic astrogliosis and endothelial/epithelial stress, determines the extent and irreversibility of demyelination and neuronal death, which underlie progressive accumulation of disability in MS.

Although longitudinal data represented only a small part of the current study, they were instrumental for selecting high SNR biomarkers, which improved the accuracy of the molecular classifiers. They also demonstrated the ability of CSF biomarkers to measure broad biological effects of applied therapies in the intrathecal compartment. Expanding CSF

biomarker studies to longitudinal cohorts could identify molecular signatures that forecast therapeutic efficacy, as well as biological synergisms among different treatments. Longitudinal cohorts are also required to determine the extent and stability of pathogenic heterogeneity (29). Implementation of CSF biomarkers to Phase I/II trials can guide dose and patient selection, and eliminate unpromising agents without accruing excessive costs and sequestering large numbers of available patients (34). Such biomarker-supported trials thus offer the promise to propel CSF-biomarker-based precision-medicine into neurology practice. While the presented results make these prospects realistic, they cannot be achieved without broader, visionary investment of efforts and resources to exploit the full potential of CSF biomarkers in neurology.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. Pierre-Olivier Courad (Institute Cochin, INSERM, France) for providing the HCMC/D3 cell line. We thank Elena Romm for processing of CSF samples. We thank clinicians Peter Williamson, Anil Panackal, Alison Wichman, Jamie Cherup, Irene Cortese, Joan Ohayon, Kaylan Fenton, Camilo Toro, Dennis Landis, Adeline Vanderver, Elisabeth Wells, Carlos Pardo, Lauren Krupp and research nurse Jenifer Dwyer for expert patient care, regulatory nurse Rosalind Hayden for help with regulatory paperwork and schedulers Anne Mayfield and Kewounie Pumphrey for patient scheduling. We thank Dragan Maric for help with cell sorting. Finally, we thank the patients, their caregivers and healthy volunteers, without whom this work could not be possible. The study was supported by the intramural research program of the National Institute of Neurological Disorders and Stroke (NINDS) of the National Institutes of Health (NIH) and the Material Transfer Agreement (MTA) between NINDS and Medimmune, LLC (A member of the AstraZeneca Group) that partially funded the SOMAscan assays. P.K. received fellowship support from the Myelin Research Foundation and M.K. received post-doctoral fellowship support from the Japan Society of the Promotion of Science.

Abbreviations

AUROC	area under the receiver operating characteristic curve
CI	confidence interval
CNS	central nervous system
CSF	cerebrospinal fluid
DMT	disease-modifying therapy
FACS	Fluorescence-activated cell sorting
ER	error rate
HD	healthy donor
LP	lumbar puncture
MS	multiple sclerosis
NIND	non-inflammatory neurological disorders
OIND	other inflammatory neurological disorders

OCB	oligoclonal bands
OOB	out-of-bag
PBMC	peripheral blood mononuclear cells
PCP	parallel coordinate plot
PMA	phorbol 12-myristate 13-acetate
PPMS	primary progressive multiple sclerosis
RFU	relative fluorescent unit
ROC	receiver operating characteristic curve
RRMS	relapsing-remitting multiple sclerosis
SNR	signal-to-noise ratio
SPMS	secondary progressive multiple sclerosis

References

1. Koga S, Aoki N, Uitti RJ, van Gerpen JA, Cheshire WP, Josephs KA, et al. When DLB, PD, and PSP masquerade as MSA: an autopsy study of 134 patients. *Neurology*. 2015; 85(5):404–12. [PubMed: 26138942]
2. Rizzo G, Copetti M, Arcuti S, Martino D, Fontana A, Logroscino G. Accuracy of clinical diagnosis of Parkinson disease: A systematic review and meta-analysis. *Neurology*. 2016; 86(6):566–76. [PubMed: 26764028]
3. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov*. 2004; 3(8):711–5. [PubMed: 15286737]
4. Johanson CE, Duncan JA 3rd, Klinge PM, Brinker T, Stopa EGS, GD. Multiplicity of cerebrospinal fluid functions: New challenges in health and disease. *Cerebrospinal fluid research*. 2008; 5:10. [PubMed: 18479516]
5. Vlassenko AG, McCue L, Jasielec MS, Su Y, Gordon BA, Xiong C, et al. Imaging and cerebrospinal fluid biomarkers in early preclinical alzheimer disease. *Ann Neurol*. 2016; 80(3):379–87. [PubMed: 27398953]
6. Komori M, Blake A, Greenwood M, Lin YC, Kosa P, Ghazali D, et al. Cerebrospinal fluid markers reveal intrathecal inflammation in progressive multiple sclerosis. *Ann Neurol*. 2015; 78(1):3–20. [PubMed: 25808056]
7. Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Annals of neurology*. 2011; 69(2):292–302. [PubMed: 21387374]
8. Bielekova B, Komori M, Xu Q, Reich DS, Wu T. Cerebrospinal Fluid IL-12p40, CXCL13 and IL-8 as a Combinatorial Biomarker of Active Intrathecal Inflammation. *PLoS ONE*. 2012; 7(11):e48370. [PubMed: 23226202]
9. Rohloff JC, Gelinas AD, Jarvis TC, Ochsner UA, Schneider DJ, Gold L, et al. Nucleic Acid Ligands With Protein-like Side Chains: Modified Aptamers and Their Use as Diagnostic and Therapeutic Agents. *Molecular therapy Nucleic acids*. 2014; 3:e201. [PubMed: 25291143]
10. Gold L, Walker JJ, Wilcox SK, Williams S. Advances in human proteomics at high scale with the SOMAscan proteomics platform. *New biotechnology*. 2012; 29(5):543–9. [PubMed: 22155539]
11. Kraemer S, Vaught JD, Bock C, Gold L, Katilius E, Keeney TR, et al. From SOMAmer-based biomarker discovery to diagnostic and clinical applications: a SOMAmer-based, streamlined multiplex proteomic assay. *PloS one*. 2011; 6(10):e26332. [PubMed: 22022604]

12. Gold L, Ayers D, Bertino J, Bock C, Bock A, Brody EN, et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PloS one*. 2010; 5(12):e15004. [PubMed: 21165148]
13. Weksler BB, Subileau EA, Perriere N, Charneau P, Holloway K, Leveque M, et al. Blood-brain barrier-specific properties of a human adult brain endothelial cell line. *Faseb J*. 2005; 19(13): 1872–4. [PubMed: 16141364]
14. Douvaras P, Fossati V. Generation and isolation of oligodendrocyte progenitor cells from human pluripotent stem cells. *Nat Protocols*. 2015; 10(8):1143–54. [PubMed: 26134954]
15. R Core Team. R: A Language and Environment for Statistical Computing. 2016.
16. Pinheiro J, Bates D, DebRoy S, Sarkar D. R Core Team. nlme: Linear and Nonlinear Mixed Effects Models. R package Version 31–128 [Internet]. 2016
17. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC bioinformatics*. 2011; 12:77. [PubMed: 21414208]
18. Hastie, T., Tibshirani, R., Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer; 2009.
19. Liaw A, Wiener M. Classification and Regression by randomForest. *R News*. 2002; 2(3):18–22.
20. Breiman L. Random Forests. *Machine Learning*. 2001; 45(1):5–32.
21. Friedman JH. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*. 2001; 29:1189–232.
22. Bielekova B, Vodovotz Y, An G, Hallenbeck J. How implementation of systems biology into clinical trials accelerates understanding of diseases. *Front Neurol*. 2014; 5:102. [PubMed: 25018747]
23. Ioannidis JP. A roadmap for successful applications of clinical proteomics. *Proteomics Clin Appl*. 2011; 5(5–6):241–7. [PubMed: 21523915]
24. Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O’Keeffe S, et al. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J Neurosci*. 2014; 34(36):11929–47. [PubMed: 25186741]
25. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*. 2012; 489(7416):391–9. [PubMed: 22996553]
26. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015; 347(6220):1260419. [PubMed: 25613900]
27. Morgan P, Van Der Graaf PH, Arrowsmith J, Feltner DE, Drummond KS, Wegner CD, et al. Can the flow of medicines be improved? Fundamental pharmacokinetic and pharmacological principles toward improving Phase II survival. *Drug discovery today*. 2012; 17(9–10):419–24. [PubMed: 22227532]
28. Kleinschmidt-DeMasters BK, Tyler KL. Progressive multifocal leukoencephalopathy complicating treatment with natalizumab and interferon beta-1a for multiple sclerosis. *N Engl J Med*. 2005; 353(4):369–74. [PubMed: 15947079]
29. Lucchinetti C, Bruck W, Parisi J, Scheithauer B, Rodriguez M, Lassmann H. Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination. *Ann Neurol*. 2000; 47(6):707–17. [PubMed: 10852536]
30. Stys PK, Zamponi GW, van Minnen J, Geurts JJ. Will the real multiple sclerosis please stand up? *Nat Rev Neurosci*. 2012; 13(7):507–14. [PubMed: 22714021]
31. Montalban X, Hemmer B, Rammohan K, Giovannoni G, De Seze J, Bar-Or A, et al. Efficacy and Safety of Ocrelizumab in Primary Progressive Multiple Sclerosis: Results of the Phase III Double-Blind, Placebo-Controlled ORATORIO Study (S49.001). *Neurology*. 2016; 86(16 Supplement)
32. Magliozzi R, Howell O, Vora A, Serafini B, Nicholas R, Puopolo M, et al. Meningeal B-cell follicles in secondary progressive multiple sclerosis associate with early onset of disease and severe cortical pathology. *Brain*. 2007; 130(Pt 4):1089–104. [PubMed: 17438020]
33. Liddelow SA, Guttenplan KA, Clarke LE, Bennett FC, Bohlen CJ, Schirmer L, et al. Neurotoxic reactive astrocytes are induced by activated microglia. *Nature*. 2017

34. Komori M, Lin YC, Cortese I, Blake A, Ohayon J, Cherup J, et al. Insufficient disease inhibition by intrathecal rituximab in progressive multiple sclerosis. *Annals of clinical and translational neurology*. 2016; 3(3):166–79. [PubMed: 27042677]
35. Komori M, Lin YC, Cortese I, Blake A, Ohayon J, Cherup J, et al. Insufficient disease inhibition by intrathecal rituximab in progressive multiple sclerosis. *Annals of clinical and translational neurology*. 2016; 3(3):166–79. [PubMed: 27042677]

Figure 1. SNR calculation and differences in technical and biological replicates versus pre- and post-treatment samples

(A) Graphical example of technical (top graphs) and biological (bottom graphs) variance calculation for SOMAmer SL004672. The x-axes correspond to the patient/sample that the measurements were produced from. The upper panels show technical replicates (n=88) and the lower panels show biological replicates (n=24). The left panels show the raw measurements (natural-log scale RFU) for SOMAmer SL004672 for each sample on the y-axes. The right panels show the identical raw observations with the random intercept effect subtracted to account for subject-to-subject variation. These residuals (after subtracting means of technical or biological replicates) were used to estimate the technical and biological variance, respectively. The horizontal black line is the estimated mean from the technical (top graphs) and biological (bottom graphs) variance models.

(B) Differences in biomarker measurements in identical samples analyzed repeatedly (technical replicates, n=88, left), in longitudinal HD samples measured at different time points (biological replicates, n=24, middle), and in patient samples before and after application of immunomodulatory therapy (biological changes, n=10, right) were quantified in two ways: (i) An average of Spearman rho values calculated across 500 high-signaling SOMAmers with high SNR, and (ii) an average of variabilities (a median of relative percent changes calculated as absolute difference of RFUs for each of the 500 high-signaling SOMAmers between two replicates divided by the average of the two RFUs) for all pairs of replicates in each respective category. Examples of the strongest and the weakest correlations between two samples in each category are visualized on 500 high-signaling SOMAmers. The axes show log10 scales of relative fluorescent units (RFU) of SOMAmers.

Figure 2. Highly simplified artificial example explaining the principles of random forests

(A) A decision tree differentiates *groups* of observations (*elements*) using selected *features*. E.g., to differentiate RRMS from progressive MS useful features may be MRI contrast-enhancing lesions (CELs) and T2 lesions, IgG index, and Disability. (B) Assembling features into a decision tree provides better results than classifying based on any single feature.

A decision tree algorithm selects from available features one that best differentiates diagnostic categories and computes its optimal threshold (i.e., the value on which to split the elements). The algorithm then finds the next best feature to split the categories, and this process repeats itself until meeting termination criterion; e.g., when a certain number of splits has occurred. The number of *splits* corresponds to the *depth* of a tree.

A random forest algorithm mitigates problem of unreliable predictions caused by *overfitting*. A random forest is a collection of decision trees, each generated slightly differently, using a random subset of features and elements. First, the algorithm restricts the number of features from which each new tree is constructed: if testing p features, the algorithm randomly selects $\sqrt{p}$ features available for every split in a tree. Second, each tree is constructed from a random sample of patients (with replacement) of the same size as the original training cohort (*bootstrapping*). The observations withheld from each tree due to bootstrapping are used to calculate *out of bag* (OOB) misclassification error.

In our example (C) only $\sqrt{4}=2$ features are used for each split in the decision trees of depth 2, with each of the decision trees generated from a bootstrapped subset of the training dataset. Panel C illustrates possible partitions for CELs-T2 lesions (upper) and CELs-Disability (lower) combinations of features, while panel (D) represents corresponding examples of decision trees, with a total of four trees in the forest. The final prediction is

derived as an average prediction from all randomly generated trees. For example, if one tree classified a patient as progressive MS but other 3 trees classified the patient as RRMS, the subject will be classified as RRMS with 75% probability. Because of the high variability in individual trees, the algorithm is typically run for many trees until the OOB predictions stabilize. Therefore it cannot be described by a mathematical equation or a single decision tree. The randomness assures that the algorithm searches the entire p-dimensional partition space (**E**; only 3 dimensions shown, but the search space is 4-dimensional) for the best features, and by averaging the partitioning thresholds in the training cohort, the classifier also effectively derives optimal global thresholds. By calculating the average OOB error when a feature is omitted from the construction of a tree, we can generate global “variable importance” metric (**F**) that reflect decrease in accuracy of the random forest classifier in the absence of the specific feature.

Figure 3. Schematic diagram of the SOMAscan analysis leading to molecular diagnostic tools

The SOMAscan assay comprises 1128 SOMAmers (solid black line curves). Calculation of technical and biological SNR reduced the number of SOMAmers considered for further analysis to 500 (dashed red line curves). Using the 500 high-signaling SOMAmers, 124750 biomarker ratios were generated that were subsequently tested for their SNR and their ability to differentiate two diagnostic groups (based on AUC in the modeling cohort), resulting in 5401 high-signaling biomarker ratios for MS versus non-MS diagnostic test, 3626 biomarker ratios for progressive versus RRMS diagnostic test, and 1504 biomarker ratios for SPMS versus PPMS diagnostic test (green dotted line curves). Out-of-bag AUC estimates (bottom graphs) examined from different random forests generated by sequentially adding ratios with the highest variable importance led to a logical cut-off (marked by solid red line) of 22 SOMAmer ratios for MS versus non-MS diagnostic comparison, 21 SOMAmer ratios for RRMS versus progressive MS diagnostic comparison, and 33 SOMAmer ratios for SPMS

versus PPMS (blue dash-dot line curves). Restriction of SOMAmer ratios to the most important ones resulted in validated AUROC =0.98 (MS versus non-MS, CI: 0.94–1.00), AUROC=0.91 (RRMS versus progressive MS, CI: 0.80–1.00), and AUROC=0.58 (SPMS versus PPMS, CI: 0.37–0.79). CI – 95% confidence interval, ER – error rate.

Figure 4. STARD (Standards for Reporting Diagnostic accuracy studies) diagrams and confusion matrices reporting the flow of subjects used for validation of the molecular diagnostic test
(A) STARD diagram and (C) confusion matrix for 85 subjects used for validation of the MS molecular diagnostic test and (B) STARD diagram and (D) confusion matrix for 47 subjects used for validation of the progressive MS molecular diagnostic test.

Figure 5. MS versus non-MS molecular diagnostic test

A parallel coordinate plot (PCP) for the 22 most important features that distinguish MS from non-MS. The plot displays individual patients from combined modeling (n=225) and validation cohort (n=85) divided into MS group (RRMS, PPMS, SPMS; thin red lines) and non-MS group (HD, NIND, OIND; thin blue lines). A group average is shown as thick yellow line for PPMS, thick red line for RRMS, thick orange line for SPMS, thick purple line for HD, thick green line for NIND, thick blue line for OIND. The y-axis shows SOMAmer natural log ratios scaled to 0–1 range. SOMAmer ratios were grouped based on the cellular origin and known functions of the individual components into nine groups. Different cell types are shown above the PCP to highlight the cell origin of individual SOMAmers. MS patients show higher plasma cell/plasmablast activation/levels compared to overall intrathecal inflammation (group 1a), myeloid lineage (groups 1b and 1c), epithelial damage (group 1d), CNS destruction and epithelial injury (group 2a), differences in immunoglobulin subtypes (group 2b), CNS and endothelial damage (group 2c), astrocyte

activation (group 2d) and higher epithelial injury compared to neutrophil activation (group 3). *ratios in the classifier are inverted.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 6. Progressive vs RRMS molecular diagnostic test

A parallel coordinate plot (PCP) for the 21 most important variables that distinguish progressive MS from relapsing MS. The plot displays individual MS patients from combined modeling (n=120) and validation cohort (n=47) divided into progressive MS group (PPMS and SPMS; thin blue lines) and relapsing MS group (RRMS; thin red lines). A group average is shown as thick purple line for PPMS, thick blue line for SPMS, and thick red line for RRMS. The y-axis shows SOMAmer natural log ratios scaled to 0–1 range. SOMAmer ratios were grouped based on the cellular origin and known functions of the individual components into nine groups. Different cell types are shown above the PCP to highlight the

cell origin of individual SOMAmers. Progressive MS patients show increased loss of neuronal, oligodendroglial, astrocytic, and neuroprotective markers (groups 1a and 1b), proportional loss of oligodendroglial marker compared to myeloid lineage and epithelial marker (group 1c), increased epithelial injury in comparison to overall immune activation (1d), enhanced complement activation (groups 2a and 2b), dysregulation of pathways linked to formation of tertiary lymphoid follicles (groups 2c, 2d, 2e, 3) and to platelet aggregation (3). *ratios in the classifier are inverted.

Figure 7. Comparison of the performance of clinical and molecular diagnostic tests

(A) The MS diagnostic probability (on the y-axis) of 85 subjects from the validation cohort is shown in the graph. Blue circles represent subjects with original non-MS diagnosis (HD, OIND, NIND) and orange circles represent subjects with original MS diagnosis (RRMS, PPMS, SPMS). The red line represents an arbitrary cut-off at 50%. The pink background marks an area between 30% and 70% where the certainty of the molecular classification is weak (contains 22.4% of the validation cohort's subjects). The orange background highlights 70.0% of the validation cohort's MS subjects with highly probable MS molecular diagnosis (>70%) and the blue background labels 86.8% of the validation cohort's non-MS subjects with high probability of non-MS molecular diagnosis (<30%). The gray bars represent a frequency distribution bar chart with the bin size of 5%. (B) Misdiagnosed subjects (pink circles) were evaluated for non-SOMALogic biomarkers of inflammation: IgG index, BCMA, sCD27, and CHI3L1 using alternative assays (for details on methodology, see (35)). The group medians are shown for MS subjects as an orange line and for non-MS subjects as a blue line. The seven MS subjects that were classified as non-MS by the molecular diagnostic test show a non-inflammatory type of disease, whereas the two non-MS (OIND) subjects that were categorized as MS according to the SOMALogic MS molecular classifier show significant levels of inflammatory markers, overlapping with MS.

(C) Comparison of IgG index data (left) and molecular MS diagnostic probability (right) in the combined modeling and validation cohort shows distributions of non-MS (blue circles) and MS subjects (orange circles) (D) Separation of RRMS (green circles) and progressive MS (PMS, purple circles) subjects into two age-categories (<45 years; left side, and >45 years; right side) shows that age does not affect performance of the progressive MS classifier.

Table 1

Demographic data

<i>Modeling cohort (n=225)</i>						
Diagnosis	HD (a)	NIND (b)	OIND (c)	RRMS (d)	PPMS (e)	SPMS (f)
N (female/male)	9/16	29/10*	17/24	26/14	17/23	23/17
Average	40.5	47.9	46.9	39.6	51.7	53.6
Age (SD)	(11.7)	(10.8)	(15.8)	(9.2)	(9.6)	(10.7)
range	22.4–58.4	18.2–70.6	15.4–74.2	18.0–59.9	28.2–70.4	27.4–69.6
Average	NA	5.4	4.2	4.4	10.6	20.4
Disease duration (SD)	NA	(7.6)	(3.9)	(6.0)	(7.5)	(10.7)
range	NA	0.2–34.5	0.5–13.1	0.0–20.7	0.5–30.3	1.5–42.4
Average	0.5	2.0	3.7	1.6	5.2	5.8
EDSS (SD)	(0.5)	(2.1)	(2.8)	(1.3)	(1.7)	(1.4)
range	0.0–1.5	0.0–6.5	0.0–9.0	0.0–6.0	2.0–8.5	2.0–8.0
Average	97.4	90.4	76.6	92.1	67.4	60.1
SNRS (SD)	(3.1)	(13.3)	(19.6)	(8.9)	(15.6)	(15.9)
range	87–100	51–100	42–100	65–100	24–94	29–87
Average	51.2	49.8	45.3	53.2	40.6	37.6
SDMT (SD)	(11.5)	(14.3)	(14.7)	(10.5)	(12.5)	(12.4)
range	32–71	19–80	26–77	32–76	4–58	12–59
<i>Validation cohort (n=85)</i>						

<i>Modeling cohort (n=225)</i>						
Diagnosis	HD (a)	NIND (b)	OIND (c)	RRMS (d)	PPMS (e)	SPMS (f)
Diagnosis	HD (a)	NIND (b)	OIND (c)	RRMS (d)	PPMS (e)	SPMS (f)
N (female/male)	5/5	12/2	6/8	10/6	9/7	11/4
Average	38.8	46.3	45.9	38.3	55.5	54.0
Age (SD)	(28.9)	(13.3)	(15.0)	(9.6)	(7.4)	(7.2)
range	28.9–56.2 ef	21.0–66.0 NA	22.1–71.0 NA	27.9–56.1 ef	36.0–64.5 ad	38.4–66.0 ad
Average	NA	5.6	4.1	2.6	10.0	25.0
Disease duration (SD)	NA	(5.3)	(4.1)	(5.1)	(5.9)	(7.7)
range	NA	0.4–14.9 f	0.4–14.9 ef	0.1–20.1 ef	1.6–24.2 cdf	9.5–38.4 bcde
Average	0.3	3.8	2.6	1.7	5.0	6.3
EDSS (SD)	(0.5)	(2.2)	(2.4)	(0.8)	(1.6)	(0.5)
range	0.0–1.0 bdef	1.0–6.5 af	0.0–6.5 f	1.0–4.0 aef	1.5–6.5 ad	5.0–7.0 abcde
Average	98.6	79.6	82.8	93.8	68.5	57.5
SNRS (SD)	(2.0)	(13.3)	(17.4)	(4.4)	(12.0)	(7.1)
range	95–100 bdef	57–98 af	57–100 f	80–98 aef	50–92 adf	49–75 abcde
Average	49.1	47.4	44.7	52.1	43.4	37.4
SDMT (SD)	(14.4)	(10.3)	(11.9)	(9.7)	(11.0)	(10.4)
range	32–69 NA	31–62 NA	28–65 NA	36–70 f	28–68 NA	20–51 d

Demographic data for six diagnostic categories represented by columns (a) – (f). Age, disease duration, EDSS, SNRS, and SDMT are shown as averages with standard deviation (SD) and a range. For each of the five continuous demographic variables: age, EDSS, SNRS, SDMT and disease duration, one-way analysis of variance (ANOVA) was used to compare the variable means across the six diagnosis groups: HD, NIND, OIND, PPMS, RR-MS, SP-MS (for disease duration without HD). Based on Levene's test for homogeneity of variances, ANOVA with equal or unequal variances was applied. Multiple comparisons between all pair-wise means were performed using Tukey's method. Normality was evaluated by the Shapiro-Wilk test based on the residuals. Natural log-transformation was applied to disease

duration. SAS version 9.4 was used for the statistical analyses. Small letters below SD identify diagnostic categories that show statistically significant difference with adjusted $p < 0.05$, NS = not statistically significant. EDSS = expanded disability status scale, SNRS = Scripps neurological rating scale, SDMT = symbol digit modality test, HD = healthy donors, NIND = non-inflammatory neurological disorders, OIND = other inflammatory neurological disorders, RRMS = relapsing-remitting multiple sclerosis, PPMS = primary progressive multiple sclerosis, SPMS = secondary progressive multiple sclerosis.

*. Fisher's exact test shows statistically significant difference in gender ($p = 0.005$) that disappears when the NIND group is excluded ($p = 0.0738$).