Multivariate Proteomic MS Disease Activity Test Score Performance Evaluated in an Independent Focal Inflammation Cohort

J. Kuhle[1], J. Oechtering[1], E. Willemse[1], D. Leppert[1], C. Granziera[1], P. Benkert[1], A. Maleska Maceski[1], S. Schaedelin[1], F. Zhang[2], F. Rubio da Costa[2], V. Gehman[2], F. Qureshi[2]
[1]University Hospital Basel, Switzerland, [2]Octave Bioscience, Menlo Park, CA, United States of America

P045

Background

- Qualitative and subjective assessments are relied upon in the current standard of care to monitor disease activity (DA) and disease progression (DP) for MS patients. A validated biological tool to quantitatively measure the level of disease activity in MS patients can therefore help address a significant unmet medical need.
- Previously, a serum based assay that measures 18 proteins used to determine 4 disease pathway scores (immune modulation, neuroinflammation, myelin biology and neuroaxonal integrity) and an overall disease activity (DA) score was analytically (Hu et al., 2021) and clinically validated (Chitnis et al., 2021) as previously reported. The primary endpoint for these analyses was the count of gadolinium enhancing (Gd+) lesions.
- Evaluating performance of the model in independent cohorts is important for ongoing characterization of model generalizability.

Objective

- To characterize performance of the MS Disease Activity (MSDA) algorithm in an independent cohort of MS patients to discriminate between patients with and without focal inflammation. Statistical metrics including sensitivity, NPV, accuracy and odds ratio that were used to establish the DA score thresholds corresponding to low (L), moderate (M) and high (H) disease activity categories were evaluated.

Methods

- Paired samples (n=138) from 69 patients recruited at the University Hospital Basel with RRMS were analyzed. For each sample pair, 1 was collected while experiencing a relapse and/or MRI focal inflammation and the other was collected while in remission. Relapse and/or MRI focal inflammation samples were drawn within 30 days after clinical relapse onset or with ≥ 1 Gd+ lesions on an MRI administered <30 days before the sample. Remission samples had no relapse 365 days before or 180 days after and no Gd+ lesions on an MRI <30 days before the sample. Patient sample demographics and characteristics are summarized in Table 1.
- Univariate analysis of 20 individual proteins included on custom assay panel used to run the MSDA Test was first performed to classify Relapse samples from Remission samples. P-values and average concentration difference were calculated to determine significance and establish directionality of the protein's association with this endpoint.
- The MSDA Test algorithm is a stacked classifier logistic regression model. Protein concentrations are demographically corrected for both age and sex using fixed coefficients that were established in the clinical validation study. The first layer of the model consists of 4 Disease Pathway Algorithms (restricted to subsets of 18 proteins pathophysiologically associated with one another). The second layer of the model utilizes the 4 Disease Pathway Algorithms as meta-features to determine an overall DA Score reflecting the likelihood of Gd positivity (see Figure 1).
- The primary output of the MSDA Test consists of the overall DA Score (scale = 1.0 to 10.0 with 0.5 intervals). The 138 samples were analyzed in the MSDA Test and assigned to Low (1.0 - 4.0), Moderate (4.5 - 7.0) or High (7.5 - 10.0) DA categories based on thresholds that were established in the previous clinical validation study cohort (n=617). The count of Gd+ lesions was used to optimize the established categorization strategy (0 lesions = Low DA, 1 lesion = Moderate DA, and ≥2 lesions = High DA). The Low versus Moderate/High threshold was selected based on sensitivity and negative predictive value (NPV). The Low/Moderate versus High threshold was selected based on accuracy.
- Confusion matrices were created for the 138 MRI focal inflammation samples based on their determined DA Score categorization (Low, Moderate, High) and their observed count of Gd+ lesions (0, 1, ≥2) to calculate the sensitivity, NPV, accuracy and odds ratio for comparison to the clinical validation study. 6 samples that did not have Gd+ lesion counts available were removed from this portion of the analysis.

Results

- In the univariate analysis of individual proteins to classify patients in a state of relapse or remission (which included both the presence of Gd enhancing lesions as well as clinically defined relapse status), 6 proteins were found to be significantly ($p< 0.05$) regulated: NfL, GFAP, MOG, VCAN, CXCL13 and TNFSF13B (BAFF). All of these with the exception of TNFSF13B (BAFF) were positively related with relapse status (i.e. higher concentrations correspond with the relapse state). Results are summarized in Figure 1 and Table 2.
- The sensitivity and NPV of the MSDA Test model for classifying 0 Gd Lesion samples vs ≥1 Gd lesion samples at the Low vs Moderate/High threshold were found to be 0.810 and 0.864 respectively compared to 0.821 and 0.845 in the previous clinical validation study. The accuracy for classifying ≤1 Gd Lesion samples vs ≥2 Gd lesion samples at the Low/Moderate versus High threshold was observed to be 0.856 compared to 0.851 in the previous clinical validation study (see Table 3).
- An odds ratio was determined indicating that a patient with a M/H score is 5.56 times more likely (95% CI: (2.31, 13.34) and $p<0.001$) to have ≥ 1 Gd lesions than a patient with a L score compared to 6.88 in the clinical validation study. An odds ratio was determined indicating that a patient with a H score is 11.44 times more likely (95% CI: (3.77, 34.73) and $p<0.001$) to have ≥2 Gd lesions than a patient with a L/M score compared to 10.39 in the clinical validation study (see Table 3).

Conclusions

- The MSDA Test algorithm performance demonstrated strong replication in this independent post-validation cohort for all metrics (sensitivity, NPV, accuracy and odds ratio) used to establish the score thresholds corresponding to Low, Moderate and High DA categories.
- Additional studies may be analyzed in the MSDA Test to further characterize model generalizability.
- The MSDA Test holds promise to be a sensitive and accurate tool to detect disease activity in clinical practice.

References:

1) Hu et al. 2021: Analytical Validation of a Multivariate Proteomic Serum Based Assay for Disease Activity Assessments in Multiple Sclerosis, P010 ACTRIMS 2021 W. Hu[1], L. Loh[1], H. Patel[1], M. DeGuzman[1], M. Becich[1], F. Rubio da Costa[1], V. Gehman[1], E. Assarsson[2], S.Ohlsson[2], M. Lundberg[2], J. Bergman[2], N. Nordberg[2], F. Qureshi[1] [1]Octave Bioscience, Inc., Menlo Park, CA, USA; [2]Olink Proteomics, Uppsala, Sweden

2) Chitnis et al. 2021: Clinical Validation Study Results of a Multivariate Proteomic Serum Based Assay for Disease Activity Assessments in Multiple Sclerosis, P574 ECTRIMS 2021 T. Chitnis[1], J. Foley[5], C. Ionete[2], N. El Ayoubi[3], S. Saxena[1], P. Gaitan-Walsh[1], H. Lokhande[1], A. Paul[1], F. Saleh[1], H. Weiner[1], F. Qureshi[4], M. J. Becich[4], F. Rubio da Costa[4], V. M. Gehman[4], S. J. Khoury[3] [1]Brigham and Women's Hospital, Harvard Medical School, Boston, MA; [2]University of Massachusetts Medical School, Worcester, MA; [3]American University of Beirut, Beirut, Lebanon; [4]Octave Bioscience, Menlo Park, CA; [5]Rocky Mountain Multiple Sclerosis Clinic, Salt Lake City, UT

Disclosures: J.Kuhle received speaker fees, research support, travel support, and/or served on advisory boards by ECTRIMS, Progressive MS Alliance, Swiss MS Society, Swiss National Research Foundation (320030_189140/1), University of Basel, Bayer, Biogen, Celgene, Merck, Novartis, Roche, Sanofi. D. Leppert is CMO of GeNeuro; he has received personal compensation for consulting and speaking, and travel reimbursement from Quanterix, Roche, Novartis, Orion,GeNeuroand Sanofi. Cristina Granziera has received the following fees which were used exclusively for research support: (i) advisory board and consultancy fees from Actelion, Novartis, Genzyme-Sanofi and F. Hoffmann-La Roche; (ii) speaker fees from Biogen, Genzyme-Sanofi and F. Hoffmann-La Roche; (iii) research support by F. Hoffmann-La Roche Ltd and prior to employment at USB has also received speaker honoraria and travel funding by Novartis. Johanna Oechtering, Eline Willemse, Pascal Benkert, Aleksandra Maleska and Sabine Schaedelin have nothing to disclose. F. Qureshi, F. Rubio da Costa, F. Zhang, and V. Gehman are employees of Octave Bioscience.

Figure 1. MSDA Stacked Classified Meta-Feature Algorithm

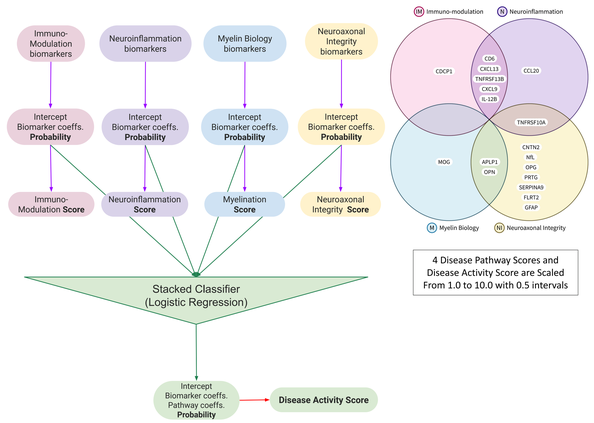

Table 1. Cohort Characteristics

Variable	Relapse (n=69)	Remission (n=69)
Gender = Women (%)	53 (76.8)	53 (76.8)
Age [Q1, Q3]	40.9 [30.1, 46.2]	41.7 [30.4, 46.7]
Disease Duration [Q1, Q3]	7.8 [4.0, 14.9]	8.6 [4.5, 15.8]
Disease Subtype		
CIS	1	0
RRMS	63	63
SPMS	4	4
PPMS	1	1
EDSS [Q1, Q3]	2.0 [2.0, 3.0]	2.0 [1.5, 3.0]
Clinical Relapse < 30d	36	0
Gd+ lesion count		
0	21	69
1	23	0
2	8	0
3	7	0
4	3	0
5	1	0
N/A	6	0
T2w lesion volume* [Q1, Q3]	5.7 [2.7, 17.4]	5.0 [2.1, 14.6]
T2w lesion number* [Q1, Q3]	34.0 [23.5, 49.5]	28.5 [17.0, 53.5]
DMT Category		
Blinded Phase 3 Trial	1	0
Monoclonal antibodies	3	13
Orals	32	42
Platform	9	5
Untreated	24	9

* T2w lesion information missing for 34/138 (24.6%) of the samples

Figure 2. Box and Whisker Plots of the Individual Protein Biomarkers

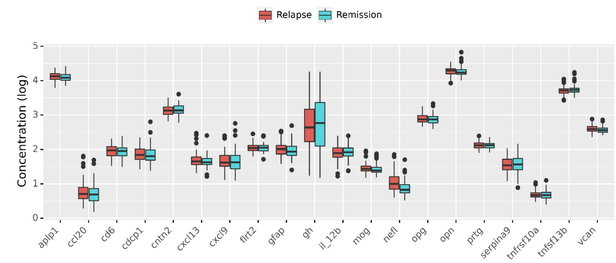

Table 2. Univariate Analysis Summary for Relapse vs. Remission Endpoint

Protein	Difference Relapse vs. Remission $\log_{10}$ [pg/mL]	p-value	Protein	Difference Relapse vs. Remission $\log_{10}$ [pg/mL]	p-value
APLP1	0.017	0.326	IL-12B	-0.016	0.558
CCL20	0.069	0.051	MOG	0.036	0.013
CD6	0.011	0.479	NfL	0.185	< 0.001
CDCP1	-0.003	0.866	OPG	0.016	0.222
CNTN2	-0.008	0.582	OPN	0.016	0.355
CXCL13	0.070	0.022	PRTG	0.012	0.218
CXCL9	0.016	0.668	SERPINA9	-0.012	0.674
FLRT2	0.000	0.970	TNFRSF10A	0.010	0.418
GFAP	0.052	0.004	TNFSF13B	-0.041	0.031
GH*	-0.056	0.579	VCAN*	0.030	0.019

*GH and VCAN are included on the custom assay panel however are not utilized in the MSDA Test algorithm.

Table 3. MSDA Test Score Confusion Matrices and Statistical Performance Metric Comparison at Score Thresholds

Low vs Moderate/High Score Thresholds Applied to 0 Gd Lesions vs ≥ 1 Gd Lesion

Focal Inflammation Cohort (n=138)	0 Gd	≥ 1 Gd	Sensitivity	Specificity	PPV	NPV	Accuracy	Odds Ratio
L (1.0-4.0)	51	8	0.810	0.567	0.466	0.864	0.644	5.56
M/H (4.5-10.0)	39	34						
Clinical Validation Study (n=617)	**0 Gd**	**≥ 1 Gd**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Accuracy**	**Odds Ratio**
L (1.0-4.0)	229	42	0.821	0.599	0.558	0.845	0.684	6.88
M/H (4.5-10.0)	153	193						

Low/Moderate vs High Score Thresholds Applied to 0 and 1 Gd Lesions vs ≥ 2 Gd Lesions

Focal Inflammation Cohort (n=138)	0/1 Gd	≥ 2 Gd	Sensitivity	Specificity	PPV	NPV	Accuracy	Odds Ratio
L/M (1.0-7.0)	103	9	0.526	0.912	0.500	0.920	0.856	11.44
H (7.5-10.0)	10	10						
Clinical Validation Study (n=617)	**0/1 Gd**	**≥ 2 Gd**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Accuracy**	**Odds Ratio**
L/M (1.0-7.0)	482	35	0.551	0.894	0.430	0.932	0.851	10.39
H (7.5-10.0)	57	43						