Clinical Validation Study Results of a Multivariate Proteomic Serum Based Assay for Disease Activity Assessments in Multiple Sclerosis

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], T. Hoyt[5], F. Qureshi[4], M. J. Becich[4], F. Rubio da Costa[4], V. M. Gehman[4], F. Zhang[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.

Objective

- To clinically validate a blood based multiplex proteomic disease activity test for associations with gadolinium enhancing (Gd+) lesions, New and Enlarging T2 lesions (N/E T2) and Active/Stable status (combination of Gd+, N/E T2 and clinical relapse status) using serum samples obtained from 4 sites: Brigham and Women's Hospital (BWH), University of Massachusetts (UMASS), American University of Beirut (AUB) and Rocky Mountain Multiple Sclerosis Clinic (RMMSC).

Introduction

- The current standard of care to evaluate disease activity (DA) and disease progression (DP) in Multiple Sclerosis (MS) patients relies primarily on qualitative radiographic and clinical assessments. Validated biological tools to quantitatively measure the level of disease activity will help to address a significant unmet medical need.
- A custom immunoassay panel that measures the concentrations of 18 proteins representing 4 primary pathways involved in MS pathophysiology was developed on the Olink™ platform utilizing Proximity Extension Assay technology.
- Proteins were selected for inclusion into the panel based on results observed in previously reported feasibility studies. Prior to the completion of this clinical validation study, the Multiple Sclerosis Disease Activity (MSDA) Test was analytically validated.

Methods

- 617 serum samples were assayed in the immunoassay panel. The samples were split into two subsets: 70% were included into a training subset (for algorithm development) and 30% into a blinded holdout-test subset (analysis was performed by an independent statistician). The subsets were stratified to ensure a balanced distribution across demographic characteristics, sample counts per site, and for the Gd+ lesion counts (see Table 1).
- The presence and count of Gd+ lesions, obtained via a matched MRI administered within 60 days of the blood draw was used for the primary DA endpoint analysis (Gd+ status determination was classified to include all samples with or without N/E T2 Lesions and Clinical Relapses). N/E T2 Lesions and Active/Stable status were analyzed as exploratory DA endpoints.
- Protein concentrations were $\log_{10}$ transformed and demographically adjusted as needed for both age and sex prior to utilization in algorithms. The final algorithm developed (based on analysis restricted to the training subset) utilized a stacked classifier logistic regression model. The first layer of the model consists of 4 Disease Pathway Algorithms (restricted to subsets of the 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 that reflects both the likelihood and severity of disease activity (see Figure 1).
- Thresholds were established for the DA Score scale (1.0 to 10.0 with 0.5 unit intervals) corresponding to Low (L), Moderate (M) and High (H) levels of Disease Activity and evaluated for sensitivity, specificity, positive predictive value, negative predictive value, accuracy and odds ratio based on the count of Gd lesions present on the associated MRI (L = 0 lesions, M = 1 lesion, H = ≥ 2 lesions). Sensitivity and NPV were selected as optimization metrics for the L vs M/H cutoff and accuracy was selected as the optimization metric for the L/M vs H cutoff. Analysis was performed at both score threshold cutoffs (L vs M/H and L/M vs H).

Results

- The multivariate model developed on the training subset was applied to the holdout-test subset and achieved an AUC of 0.765 relative to the Gd+ lesion endpoint, 0.734 relative to the N/E T2 endpoint and 0.773 relative to the Active/Stable endpoint. In each case, the multivariate model was found to be superior (p-value = <0.001) as compared to the top performing univariate biomarker NfL (see Figure 2).
- 2 X 2 confusion matrices were created to evaluate performance at the established score level thresholds in the training set, test set and for the entire study. The sensitivity and NPV of the DA score at the L vs M/H cutoff was determined to be 0.724 and 0.779 respectively in the test set. The accuracy at the L/M vs H cutoff was determined to be 0.883. A diagnostic odds ratio was determined at the L vs M/H cutoff indicating that a patient with a M/H score is 5.10 times more likely to have one or more Gd lesions than a patient with a L score. A diagnostic odds ratio was determined at the L/M vs H cutoff indicating that a patient with a H score is 15.79 times more likely to have two or more Gd lesions than a patient with a L/M score (see Table 2).
- A waterfall plot of the results for the entire study cohort demonstrates that the calculated DA Score reflects both the likelihood and severity of radiographic disease activity based on the presence (or lack thereof) and count of Gd+ lesions (see Figure 3).

Conclusions

- The MSDA Test has been successfully clinically validated. For all disease activity endpoints (Gd+, N/E T2 and Active/Stable), the multivariate model was significantly superior (p<0.001) to the top performing univariate biomarker.
- Additional analysis has been conducted to characterize performance relative to the disease progression endpoints in this study (see P029), and to investigate performance of the MSDA Test relative to the patient's current DMT (see P043).
- This validated multivariate proteomic blood-based assay for disease activity assessments can serve as a quantitative and objective tool to enhance the care for MS patients.

Disclosures: T. Chitnis has received compensation for consulting from Biogen, Novartis Pharmaceuticals, Roche Genentech, and Sanofi Genzyme. She has received research support from the National Institutes of Health, National MS Society, US Department of Defense, EMD Serono, I-Mab Biopharma, Mallinckrodt ARD, Novartis Pharmaceuticals, Octave Bioscience, Roche Genentech, and Tiziana Life Sciences. This research was conducted in part with the support of the Department of Defense through the Multiple Sclerosis Research Program under Award No. W81XWH-18-1-0648 (to T. Chitnis). C. Ionete has received research support from Biogen, Serono, Genentech, NMSS, and Department of Defense. She received compensation for advisory board activity from Sanofi-Genzyme. N. Ayoubi has received support to attend scientific educational courses from the following companies: Novartis, Merck Serono, Sanofi, Biologix. Has received speaker honoraria for scientific presentations on Multiple Sclerosis from the following companies: Biologix, Sanofi, Merck Serono, Novartis. S.i Saxena, P. Gaitan-Walsh, A. Paul, F. Saleh and T. Hoyt have no disclosures. H. Lokhande has received research support from the US Department of Defense and Octave Bioscience. H. Weiner has received research support from the Department of Defense, Genentech, Inc., National Institutes of Health, National Multiple Sclerosis Society, Novartis and Sanofi Genzyme. He has received compensation for consulting from Genentech, Inc, IM Therapeutics, IMAB Biopharma, MedDay Pharmaceuticals, Tiziana Life Sciences and vTvTherapeutics. S. Khoury has received compensation for scientific advisory board activity from Merck and Roche. J. Foley has received research support from Biogen, Novartis, Adamas, Octave, Genentech, and Mallinckrodt. He received speakers' honoraria and acted as a consultant for EMD Serono, Genzyme, Novartis, Biogen, and Genentech. He has equity interest in Octave. He is the founder of InterPro Biosciences. F. Qureshi, M. Becich, F. Zhang, F. Rubio da Costa, and V. Gehman are employees of Octave Bioscience.

Table 1. Clinical Validation Study Train vs Test Subset Stratification

	TRAIN		TEST	
	N	%	N	%
Sample Size	429	70%	188	30%
Gd+ Status				
0 Lesions	270	63%	112	60%
1 Lesions	102	24%	55	29%
≥ 2 Lesions	57	13%	21	11%
N/E T2 Lesion Presence				
No	229	64%	100	65%
Yes	128	36%	53	35%
Active (Gd+, N/E T2+, or Clinical Relapse) vs. Stable				
Stable	253	59%	107	57%
Active	176	41%	81	43%
Site				
AUB	143	33%	59	31%
BWH	134	31%	61	32%
RMMSC	117	28%	52	28%
UMASS	35	8%	16	9%
Age (Mean ± SD)	41.6 ± 12.7		42.5 ± 13.5	
Disease Duration (Mean ± SD)	9.56 ± 8.53		9.19 ± 8.83	
Sex				
Female	302	70%	134	71%
Male	127	30%	54	29%

Figure 1. MSDA Stacked Classified Meta-Feature Algorithm

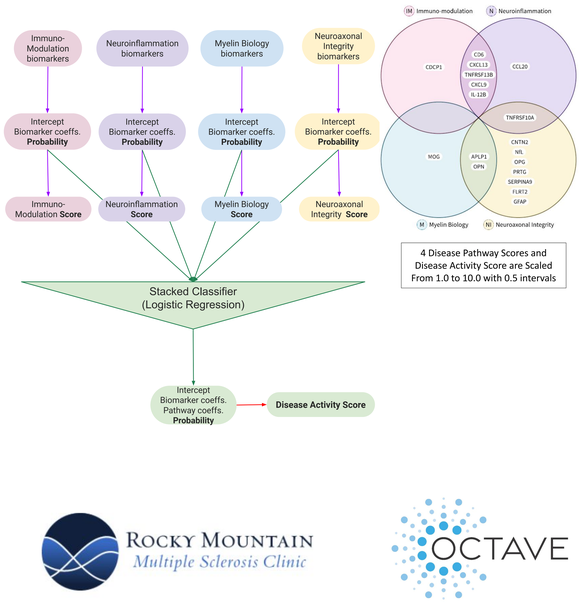
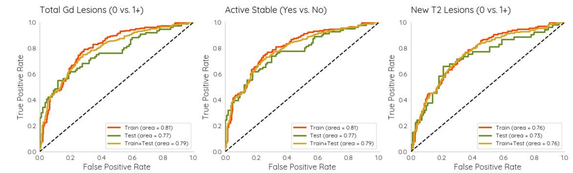


Figure 2. AUC for All DA Endpoints Compared to Top Performing Univariate Protein



Gd Lesion Holdout (Test Set)		Active/Stable Holdout (Test Set)		New Enlarging T2 Holdout (Test Set)	
Multivariate AUC	0.765	Multivariate AUC	0.773	Multivariate AUC	0.734
NFL AUC	0.694	NFL AUC	0.663	NFL AUC	0.618
p-value = <0.001*		p-value = <0.001*		p-value = <0.001*	

Train (n=429), Test (n=188), CV = Entire Clinical Validation Study (n=617)
*p-value determined via Fisher's Exact test

Table 2. MSDA Test Score Confusion Matrices and Statistical Performance Metrics

Low vs Moderate/High Score Thresholds Applied to 0 Gd Lesions vs ≥ 1 Gd Lesion								
Test (n=188)	**0 Gd**	**≥ 1 Gd**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Accuracy**	**Odds Ratio**
L (1.0-4.0)	74	21	0.724	0.661	0.591	0.779	0.686	5.10
M/H (4.5-10.0)	38	55						
Entire 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								
Test (n=188)	**0/1 Gd**	**≥ 2 Gd**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Accuracy**	**Odds Ratio**
L/M (1.0-7.0)	154	9	0.571	0.922	0.480	0.945	0.883	15.79
H (7.5-10.0)	13	12						
Entire 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						

Figure 3. Waterfall Plot of DA Scores for All Clinical Validation Study Samples