Classification of High Versus Low Annualized Relapse Rate Status in Subjects with Relapsing-Remitting Multiple Sclerosis Using Multivariate Serum Protein Biomarker Models

Neda Sattarnezhad[1], Shrishti Saxena[1], Cindy Gonzalez[1], Hrishikesh Lokhande[1], Bonnie Glanz[1], Ferhan Qureshi[2], Michael Becich[2], Remus Osan[2], Howard Weiner[1], Tanuja Chitnis[1]

[1]Partners MS Center, Department of Neurology, Ann Romney Center for Neurologic Diseases, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; [2] Octave Bioscience, Menlo Park, CA, USA

INTRODUCTION

BACKGROUND:

•Multiple sclerosis (MS) is a chronic inflammatory demyelinating disease of the central nervous system with various phenotypes and heterogenous disease course. [1]

•While exact pathophysiology of MS remains elusive, both inflammatory and degenerative processes are believed to play a role in the disease mechanism and disability progression. [1,2]

•Identifying disease-specific biomarkers may assist with predicting the diverse disease course and classifying patients to high risk versus low risk for disease activity and progression. [3,4]

•Use of multivariate models reflecting multiple biological pathways that are involved in the complex pathophysiology of MS will most likely increase predictive accuracy of these biomarkers.[5]

•Annualized Relapse Rate (ARR) is a quantifiable outcome measurement which has been strongly correlated with disability score and disease progression in relapsing forms of MS.[6,7]

•Multivariate biomarker models will most likely correlate strongly with clinical outcome measurements including ARR status.

OBJECTIVE:

To investigate the performance of multiple protein biomarker models to classify samples from Relapsing-Remitting (RR) MS subjects with High and Low ARR from the Comprehensive Longitudinal Investigation of Multiple Sclerosis at Brigham and Women's Hospital (CLIMB) study.

METHODS and MATERIALS

SUBJECTS:

Serum samples from 30 RRMS patients with Low ARR (≤0.2 relapses/year) were compared to 30 age, sex and treatment matched subjects with High ARR (≥1.0 relapses/year). All the patients were in the disease remission phase and hadn't received steroids within 30 days prior to the sample collection. All samples were measured for 1104 proteins, including serum levels of neurofilament light chain (sNfL), using Proximity Extension Assays (PEA) from Olink to quantify protein biomarker expression.

STATISTICAL ANALYSIS:

Univariate/multivariate machine learning-driven biostatistical techniques were applied to the proteomic data. Cross-validation and bootstrapping were performed to minimize overfitting and ensure best generalizability for predicting the relapse frequency status of new samples. Area under the curve (AUC) and accuracy were selected as the key metrics for evaluating model performance. Biomarkers that were identified as important features in multivariate models were investigated for biological relevance using pathway and network models to evaluate their involvement in MS pathophysiology.

RESULTS-I

1104 protein biomarkers were measured on all n=60 serum samples to screen for potential biomarkers. Measurements were filtered to exclude quality control warnings, high coefficient of variation, and assays frequently imputed at the limit of detection. To reduce dimensionality, multiple-hypothesis-adjusted univariate statistics and accuracy-weighted multivariate performance across 100,000 simulated models (ranging in type: logistic regression, support vector machine, random forest, stochastic gradient descent) were used to select for the best features. Incremental model-building strategies under regularization optimized for AUC as measured by cross-validation and bootstrapping. While several model types achieved a steady-state AUC > 0.90, a 7-feature logistic regression function was the most parsimonious and generalizable to future datasets.

Protein	P-value	AUC	Coefficient 95% C.I.	|Importance|
CD6	1.95E-04	0.76	(0.799,1.412)	0.196
IL-1RT2	1.77E-03	0.706	(0.466, 1.186)	0.146
COL4A1	2.05E-03	0.679	(-1.258, -0.702)	0.273
LEPR	2.16E-03	0.694	(-1.548,-0.945)	0.141
AGR2	8.15E-03	0.666	(-1.393, -0.799)	0.288
BCAN	2.59E-02	0.662	(0.900, 1.472)	0.313
CSTB	3.25E-02	0.658	(0.318, 0.957)	0.194
sNfL	3.16E-02	0.611	N/A	N/A

Figure 1. The 7 proteins in the logistic regression classification model (sNfL shown for comparison). Exploratory data analysis revealed normality across each protein signal distribution, and with n > 50, this assumption was safe to apply throughout. 2-sample, 1-sided homoscedastic t-tests were used for calculation of p-values. AUC is a trapezoidal integration of TPR, FPR across all 60 samples. The 95% confidence interval for each coefficient was computed by a 50/50 bootstrap procedure (B=100,000) to determine if explanatory variables in this optimal model are significant (nonzero). Standardized coefficients were used to compute relative model importance.

RESULTS-II

A multivariate logistic regression model that included 7 biomarkers achieved performance of 0.923 ± 0.072 AUC and 0.831 ± 0.024 Accuracy for classifying High vs. Low relapse rate specimens. The model was evaluated using leave-one-out cross-validation and a 50/50 train/test split simulated B=100,000 trials to establish confidence margins. The multivariate model significantly outperformed all univariate biomarkers including sNfL (0.611 AUC, 0.583 ± 0.0177 Accuracy). Cluster of Differentiation 6 (CD6), a protein associated with T-cell activation and proliferation as well as being an identified risk gene for MS[8], was identified as the strongest separating biomarker.

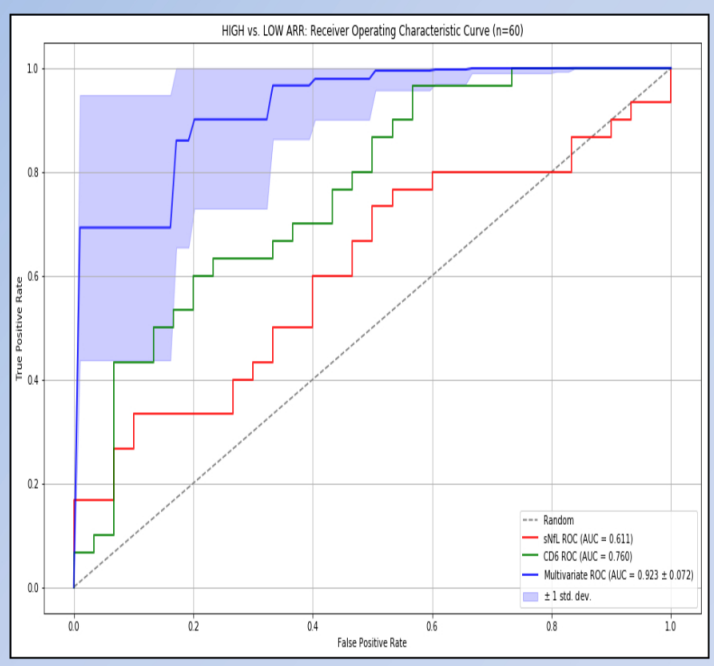

Figure 2. The Receiver Operating Characteristic visualizes the true and false positive rates of various thresholds to separate the normalized protein signals across samples. In this cohort, sNfL marginally outperformed random guessing for separating relapse rates in MS patients. CD6 was able to distinguish patients more sensitively and specifically, and the 7-feature multivariate model was able to significantly outperform all univariate markers. Correlational analysis between all features in the top-performing model (Pearson's R < 0.37 for all pairwise combinations) demonstrates the lack of collinearity in our system.

CONCLUSIONS

- Multivariate serum protein biomarker models representing several biological pathways were able to effectively classify subjects with high vs. low annualized relapse rates.
- Multivariate models outperformed all univariate approaches with a single biomarker, including CD6 and sNfL.
- These models may assist with predicting disease course and risk of disability progression to guide appropriate treatment planning and patient counseling.
- Further investigation with larger sample numbers and from additional cohorts is warranted.

REFERENCES

1)Wingerchuk, Dean M., and Jonathan L. Carter. "Multiple sclerosis: current and emerging disease-modifying therapies and treatment strategies." *Mayo Clinic Proceedings*. Vol. 89. No. 2. Elsevier, 2014. 2) Macaron, Gabrielle, and Daniel Ontaneda. "Diagnosis and Management of Progressive Multiple Sclerosis." *Biomedicines* 7.3 (2019): 56. 3) Comabella, Manuel, and Xavier Montalban. "Body fluid biomarkers in multiple sclerosis." *The Lancet Neurology* 13.1 (2014): 113-126. 4) Barro, Christian, et al. "Serum neurofilament as a predictor of disease worsening and brain and spinal cord atrophy in multiple sclerosis." *Brain* 141.8 (2018): 2382-2391. 5)Martinelli, V., et al. "Multiple biomarkers improve the prediction of multiple sclerosis in clinically isolated syndromes." *Acta Neurologica Scandinavica* 136.5 (2017): 454-461. 6)Jokubaitis, Vilija G., et al. "Predictors of long-term disability accrual in relapse-onset multiple sclerosis." *Annals of neurology* 80.1 (2016): 89-100. 7) Vercellino, M., et al. "Multiple sclerosis relapses: a multivariable analysis of residual disability determinants." *Acta Neurologica Scandinavica* 119.2 (2009): 126-130. 8) Yan Li et al. "CD6 as a potential target for treating multiple sclerosis." *Proceedings of the National Academy of Sciences* Mar 2017, 114 (10) 2687-2692.

DISCLOSURES

Dr. Sattarnezhad has received research support from Serono and Verily; has received travel grant from ECTRIMS-2019 scientific program committee for this presentation. Ms Saxena received research support from Octave, Serono and Verily. Ms. Gonzalez, Mr. Lokhande and Dr. Glanz have received research support from Serono and Verily. Mr. Qureshi, Mr. Becich and Mr. Qureshi are employees of Octave Bioscience. Dr Weiner reports grants from National Institutes of Health, grants from National Multiple Sclerosis Society, grants from Verily, grants from EMD Serono, grants from Biogen, grants from Teva Pharmaceuticals, grants from Sanofi, grants from Novartis, grants and personal fees from Genentech, Inc, grants and personal fees from Tilos Therapeutics, personal fees from Tiziana Life Sciences, personal fees from IM Therapeutics, personal fees from MedDay Pharmaceuticals, personal fees from vTv Therapeutics, outside the submitted work. Dr. Chitnis has served on advisory boards for Biogen, Novartis, and Sanofi-Genzyme; has participated in clinical trials sponsored by Sanofi-Genzyme and Novartis; has received research support from the Department of Defense, National MS Society, Guthy Jackson Charitable Foundation, Novartis, Octave, Serono and Verily-Genzyme.

SUPPORT

This study was supported by Octave Bioscience. U.S. Department of Defense (T.C). The CLIMB study has received support from EMD-Serono, the National MS Society, the Nancy Davis Center without Walls and Philanthropy.

CLIMB PATHWAYS TO CURING MS

octave BIOSCIENCE

ECTRIMS 2019

Contact email addresses:

n.sattarnejad@gmail.com

tchitnis@bwh.harvard.edu