

**Metabolomics approach to human brain spectroscopy identifies associations between clinical features and the frontal lobe metabolome in multiple sclerosis**

Lisa K. Vingara <sup>1,a</sup>, Hui Jing Yu <sup>2,b</sup>, Mark E. Wagshul <sup>3,c</sup>, Dana Serafin <sup>4</sup>, Christopher Christodoulou <sup>4</sup>, István Pelzcer <sup>1,\*</sup>, Lauren B. Krupp <sup>4,\*</sup>, and Mirjana Maletić-Savatić <sup>5,d,\*</sup>

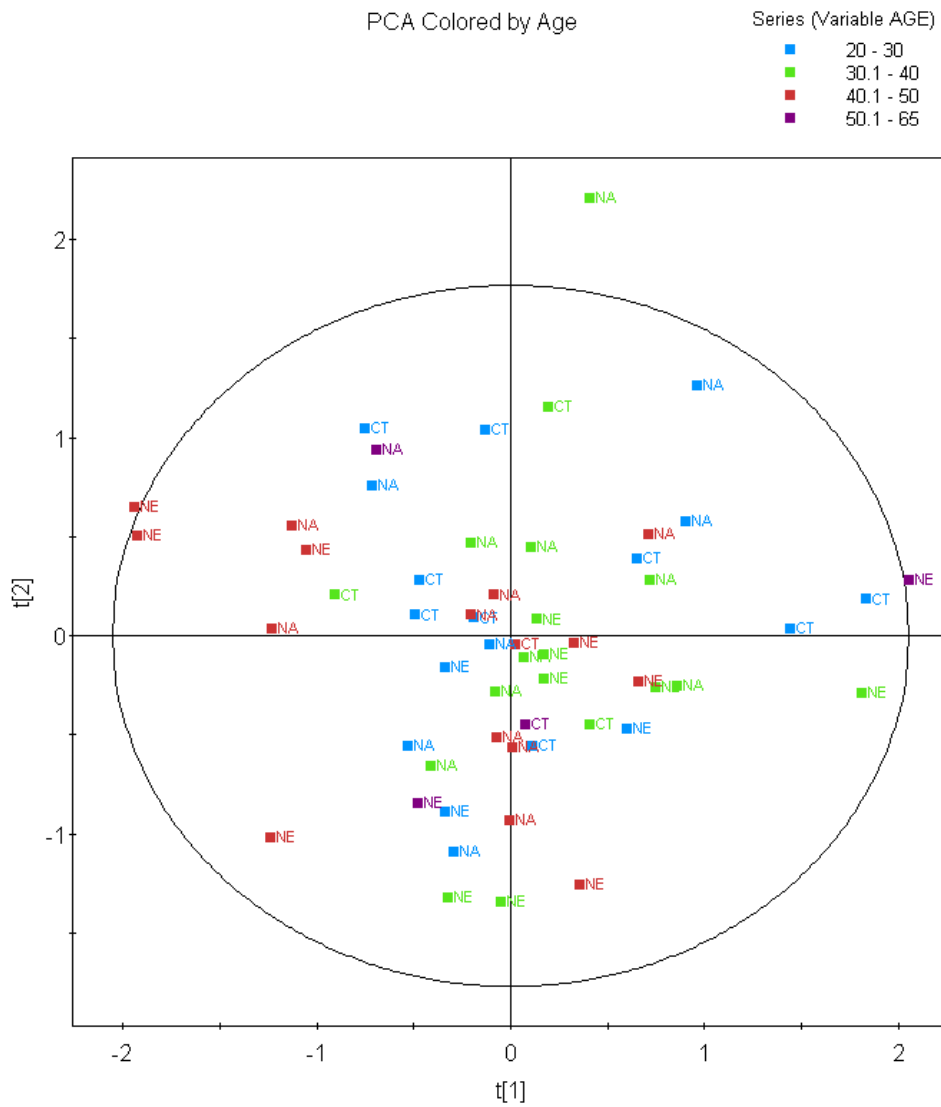
**Supplementary Information**

**Supplementary Table 1. Clinical Information of MS subjects.**

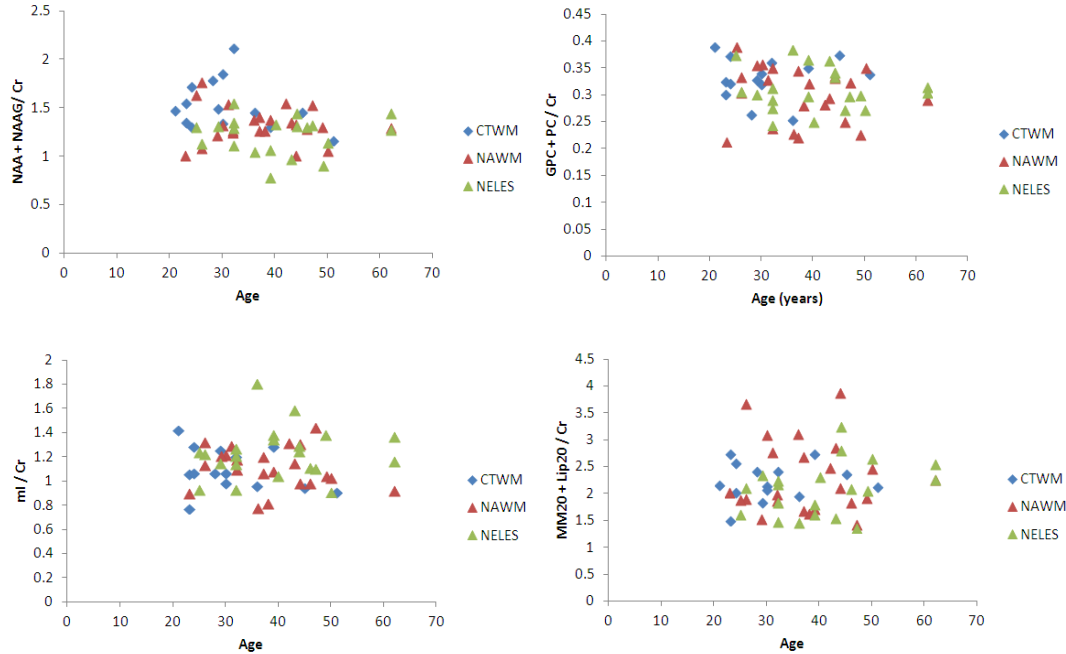
<b>Subject #</b>	<b>Age (years)</b>	<b>Gender</b>	<b>EDSS</b>	<b>Disease Duration (years)</b>	<b>Disease Modifying Treatment</b>	<b>Lesion Load (cm<sup>3</sup>)</b>
1	31	M	1.5	11	Yes	30.7
2	46	F	3.0	7	Yes	31.4
3	49	F	2.0	2	Yes	21.9
4	30	F	3.5	10	Yes	17.6
5	37	F	3.5	10	Yes	28.1
6	42	F	0	14	No	8.7
7	44	M	2.5	2	Yes	8.2
8	43	F	2.0	20	Yes	7.3
9	29	F	3.5	1	No	15.6
10	47	F	2.0	12	Yes	11.2
11	39	F	3.5	5	No	9.6
12	26	F	2.0	3.5	No	23.8
13	23	F	4.0	11	No	47.8
14	39	F	3.5	0.2	No	18.6
15	44	M	4.0	1	No	29.1
16	62	F	3.0	40	No	18.8
17	32	F	2.0	0.8	No	11.2
18	62	F	4.0	14	Yes	31.7
19	50	F	2.5	25	No	15.9
20	32	M	2.0	0.6	Yes	4.0
21	36	F	2.0	8	No	27.4
22	26	F	0	1	No	6.9
23	32	M	2.0	4	No	3.6
24	37	F	2.0	3	No	8.9
25	38	F	2.0	5	No	10.1
26	25	F	0	6	No	12.2
27	40	F	1.0	0.25	No	8.2
<b>Range</b>	23-62		0-4	0.2 - 40		3.6 – 47.8
<b>Mean</b>	38.6		2.3	8		17.3
<b>Median</b>	38		2.0	5		15.6

Adults diagnosed with RRMS and in remission at the time of imaging (N=27). The EDSS, a measure of neurological impairment ranging from 0 (none) to 10 (most severe), was performed

within 7 days of neuroimaging. Control subjects (N=14) had an age range of 23-52 years, mean of 31.1, and median of 30 years.



**Supplementary Figure 1. Unsupervised PCA of all MRS data annotated by age.** Age has no influence on the MRS data. CT = control white matter, NA = normal appearing white matter, NE = nonenhancing lesion. Colors represent different age groups.  $t[1]$  is the first principal component and  $t[2]$  is the second principal component.

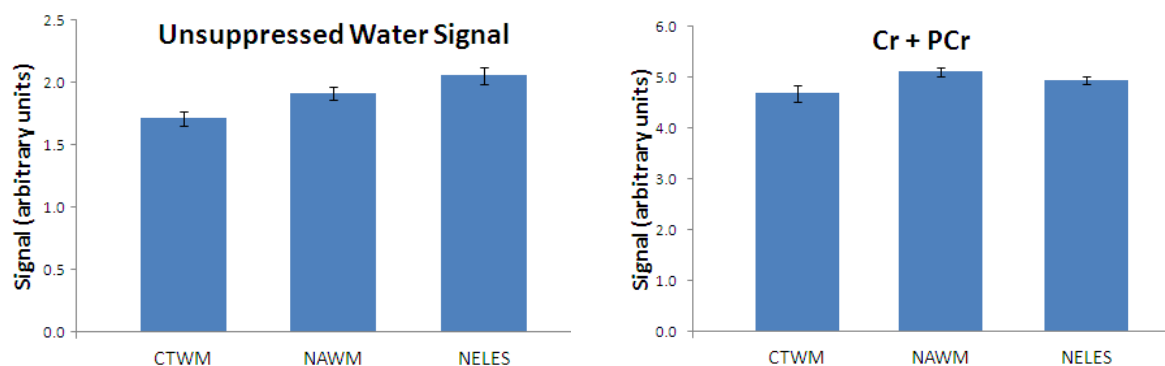


**Supplementary Figure 2. Scatter plots of key brain metabolites normalized to Creatine versus age.** The analysis of individual metabolites was done using LCModel. No age effect is observed for NAA (NAA+NAAG), Choline (GPC+PC), myoInositol (ml), and macromolecules/lipids (MM20+Lip20) in all experimental and control groups.

### Evaluation of Creatine and Water as References for Normalization.

The absolute quantity of creatine and phosphocreatine (Cr+PCr) from the LCModel and signal from the unsuppressed water spectra were both considered as normalization factors. A t-test was used to identify significant differences between groups (CTWM, NAWM, NELES) and Pearson's correlation was used to determine if there was a relationship with age. The water signal was found to significantly differ between all three groups (mean  $\pm$  SEM: CTWM:  $1.71 \pm 0.06$ , NAWM:  $1.91 \pm 0.05$ , NELES:  $2.06 \pm 0.07$ ; p-value CTWM vs. NAWM, 0.014; CTWM vs NELES, 0.001; NAWM vs NELES, 0.098), in addition to having a strong relationship with age ( $p = 0.009$ ). The absolute concentration of Cr+PCr was significantly decreased in CTWM in comparison to NAWM (CTWM:  $4.68 \pm 0.158$ ; NAWM  $5.12 \pm 0.086$ ;  $4.95 \pm 0.081$ ; p-values CTWM vs. NAWM, 0.025; CTWM vs NELES, 0.148; NAWM vs NELES, 0.167), and did not

have a significant relationship with age ( $p = 0.083$ ). This indicates that in our dataset, the levels of Cr+PCr are more consistent across groups than the water. Therefore, Cr+PCr was used as the normalization factor for the metabolomic analysis.



**Supplementary Figure 3. Absolute concentration of water and Cr+PCr signals in different tissue types.** Bar graphs represent mean  $\pm$  SEM of the unsuppressed water signal and Cr+PCr based on LCModel analysis. N=14 for CTWM, N=23 for NAWM, and N=19 for NELES.

### LCModel Analysis

**Supplementary Table 2. LCModel Quantification of Metabolites (mean  $\pm$  SEM)**

Metabolite/ Cr + PCr	CTWM	NAWM	NELES	ANOVA	CTWM vs NAWM	NAWM vs NELES	CTWM vs NELES
				p-value	Adjusted p-value		
NAA+NAAG	1.53 $\pm$ 0.26	1.33 $\pm$ 0.19	1.22 $\pm$ 0.19	>0.001	0.02	0.22	<0.001
Cho	0.33 $\pm$ 0.04	0.30 $\pm$ 0.05	0.31 $\pm$ 0.04	0.16	ns	ns	ns
Gua	0.43 $\pm$ 0.15	0.36 $\pm$ 0.10	0.34 $\pm$ 0.15	0.11	ns	ns	ns
ml	1.09 $\pm$ 0.18	1.11 $\pm$ 0.17	1.25 $\pm$ 0.21	0.03	0.97	0.05	0.06
Glu	1.31 $\pm$ 0.24	1.28 $\pm$ 0.23	1.11 $\pm$ 0.24	0.03	0.94	0.06	0.06
Asp	0.21 $\pm$ 0.07	0.23 $\pm$ 0.10	0.15 $\pm$ 0.10	0.02	0.77	0.02	0.16
GABA	0.09 $\pm$ 0.03	0.11 $\pm$ 0.05	0.12 $\pm$ 0.04	0.08	0.53	0.34	0.06
Gln	0.29 $\pm$ 0.16	0.43 $\pm$ 0.22	0.34 $\pm$ 0.24	0.13	ns	ns	ns
MM14+Lip13 a+Lip13b+M M12	1.09 $\pm$ 0.41	1.41 $\pm$ 0.76	1.27 $\pm$ 0.56	0.34	ns	ns	ns
MM17	0.32 $\pm$ 0.29	0.60 $\pm$ 0.35	0.51 $\pm$ 0.43	0.10	ns	ns	ns

## Metabolomics Statistical Methods.

Metabolomics typically uses multivariate regression to model nuclear magnetic resonance (NMR) or mass spectrometry data. The benefit of multivariate over univariate analysis is that one biomarker may be insufficient by itself to provide biologically meaningful knowledge. There are a variety of multivariate methods used for metabolomic analysis, for example unsupervised methods such as Principal Component Analysis (PCA); supervised linear methods such as Partial Least Squares (PLS) and various extensions of PLS; and supervised non-linear methods such as neural networks and support vector machines (Lindon and Nicholson, 2008; Smolinska et al., 2012). Linear methods are typically preferred over the non-linear methods since the signals responsible are identifiable from the loading matrices, leading to easy interpretation.

PCA is an unsupervised method as it does not use additional information about the data and models the innate variation of the dataset, hence providing an overview of the data. PLS and related methods are regression extensions of PCA, and model the variation in the data matrix  $\mathbf{X}$  in response to additional variables in a response matrix  $\mathbf{Y}$ . PLS can be described by the following equations where  $\mathbf{X}$  is the descriptor matrix,  $\mathbf{Y}$  is the response matrix,  $\mathbf{T}$  the scores matrix of  $\mathbf{X}$ ,  $\mathbf{U}$  the scores matrix of  $\mathbf{Y}$ ,  $\mathbf{P}$  the PLS loadings matrix,  $\mathbf{C}$  the weights expressing correlation between  $\mathbf{Y}$  and  $\mathbf{T}(\mathbf{X})$ , and  $\mathbf{E}$  and  $\mathbf{F}$  are matrices of residuals (Wold et al., 2001). There are also additional loadings called weights,  $\mathbf{W}$ , that express the correlation between  $\mathbf{U}$  and  $\mathbf{X}$  and are used to calculate  $\mathbf{T}$ :

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} = t_1\mathbf{p}_1^T + t_2\mathbf{p}_2^T + \dots + t_a\mathbf{p}_a^T + \mathbf{E} \quad (\text{S.1})$$

$$\mathbf{Y} = \mathbf{UC}^T + \mathbf{F} = u_1\mathbf{c}_1^T + u_2\mathbf{c}_2^T + \dots + u_a\mathbf{c}_a^T + \mathbf{F} \quad (\text{S.2})$$

$$\mathbf{T} = \mathbf{XW}(\mathbf{P}^t \mathbf{W})^{-1} \quad (\text{S.3})$$

Predictions,  $\hat{\mathbf{Y}}$ , are used to predict the response of new observations with respect to the specified PLS model. The predicted values of  $y$ ,  $\hat{y}$ , are calculated from the PLS model:

$$\hat{\mathbf{Y}} = \mathbf{UC}^T \quad (\text{S.4})$$

Partial Least Squares-Discriminant Analysis (PLS-DA) and Orthogonal Partial Least Squares (O-PLS) are both extensions of PLS. In PLS-DA, the response matrix  $\mathbf{Y}$  is an  $n \times k$  matrix (with  $n$  observations and  $k$  number of classes), where  $\mathbf{Y}$  is composed of dummy variables indicating class membership (Barker and Rayens, 2003). O-PLS models the variation in data matrix  $\mathbf{X}$  into 2 components, one linearly related to the  $\mathbf{Y}$  matrix and another that is orthogonal and associated with uncorrelated variation in the data, as described by equations S.5 and S.6 (Trygg and Wold, 2002). This allows for easy interpretation of the metabolic contributions related to the response matrix since they all lie along one component.

$$\mathbf{X} = \mathbf{TW}^T + \mathbf{T}_{\text{Yosc}} \mathbf{P}_{\text{Yosc}}^T + \mathbf{E} \quad (\text{S.5})$$

$$\mathbf{Y} = \mathbf{TC}^T + \mathbf{F}, \quad (\text{S.6})$$

where  $\mathbf{T}$  represents the score matrices for  $\mathbf{X}$  and  $\mathbf{Y}$ ,  $\mathbf{W}$  and  $\mathbf{C}$  are the orthonormal loading matrices,  $\mathbf{E}$  and  $\mathbf{F}$  are the respective residual matrices for  $\mathbf{X}$  and  $\mathbf{Y}$ .  $\mathbf{T}_{\text{Yosc}}$  and  $\mathbf{P}_{\text{Yosc}}$  are the score and loadings matrices orthogonal to  $\mathbf{Y}$ , respectively.

To describe the model validation procedures, we will consider the general case of predicting  $\mathbf{Y}$  data from  $\mathbf{X}$  data. As part of the modelling procedure, a component-wise cross validation is applied where 1/7<sup>th</sup> of the samples are randomly excluded from model building and used for predictions. This procedure is repeated until each sample is excluded from model building once. By this method, the  $y$  value for each sample is predicted using a model from which that sample has been excluded during the model building. All the predictions are pulled

together to form the prediction error sum of squares (*PRESS*), which is indicative of the predictive ability of the model:

$$PRESS = \sum (y_{ik} - \hat{y}_{ik})^2 \quad (\text{S.7})$$

where  $y_{ik}$  is the observed value and  $\hat{y}_{ik}$  is the predicted value.

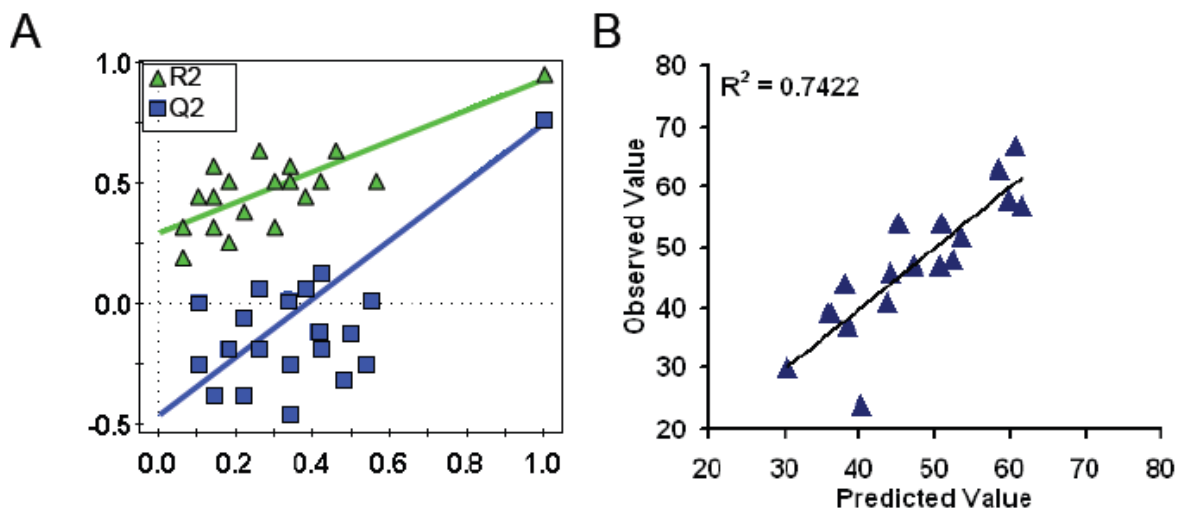
From *PRESS*, the cross validated  $R^2$ , called  $Q^2$ , can be calculated to provide an estimate of how well the model predicts the Y data:

$$Q^2 = 1 - \frac{PRESS}{SS} \quad (\text{S.8})$$

Where SS is the residual sum of squared deviations from the dataset means.

Additionally, ‘internal’ model validation can be performed by randomizing the positions of the Y data in relation to their corresponding rows in the X matrix and observing the effect of the randomization on  $R^2$  and  $Q^2$  values, as illustrated in **Supplementary Figure 4**. If the original model was valid, randomization of the Y data would be expected to considerably reduce  $Q^2$ .

Further, a plot of observed vs. predicted (**Supplementary Figure 4.**) can be used to assess the model. If all predictions  $\hat{y}_{ik}$  are perfect, this plot would be a straight line with a slope of 1 going through the origin and the correlation coefficient would be 1. In addition to assessing the predictability of the model, this plot can be used to identify “outliers” which are predicted incorrectly and may be skewing the model.



**Supplementary Figure 4.** Examples of validation tools for PLS models. A) The result of the permutation test performed to determine if the model is spurious by comparing the goodness of fit ( $R^2$  and  $Q^2$ ) of the original model with the goodness of fit of several models where the order of the Y-observations have been randomly permuted, while the **X**-matrix has been kept intact. B) The observed vs. predicted values of a selected y variable, with a regression line indicating the goodness of fit. In addition to assessing the predictability of the model, this plot can be used to identify “outliers” which are predicted incorrectly, for example the sample whose observed value is 20 and predicted value 40 in the above plot.

‘External’ validation of PLS-DA models can also be performed by taking a test set of samples that did not form part of the model-building population. While this is the ultimate test of validation for statistical modelling, since it assesses how well the model can predict new data, it is not often performed due to a low number of available samples. Thus, it becomes a trade-off between using all of the data to build a better model and using less data to be able to validate the model, which may result in a poorer quality model. The problem with the former, even with the use of internal validation, is that the model may only describe the data it was built on and not be able to describe new samples.

In PLS-DA, the dummy matrix containing class information is a matrix of zeros and ones, zero being defined as not belonging to the specified class, and a value of 1 being defined as belonging to the class. The cut-off value for accepting the sample as correctly being predicted is 0.5; therefore a predicted value  $>0.5$  indicates the spectrum belongs to that class, and a value  $<0.5$  indicates that the spectrum does not belong to that class. Further, the predictive scores can be overlaid onto the scores plot of the PLS-DA model.

### **Model summary statistics**

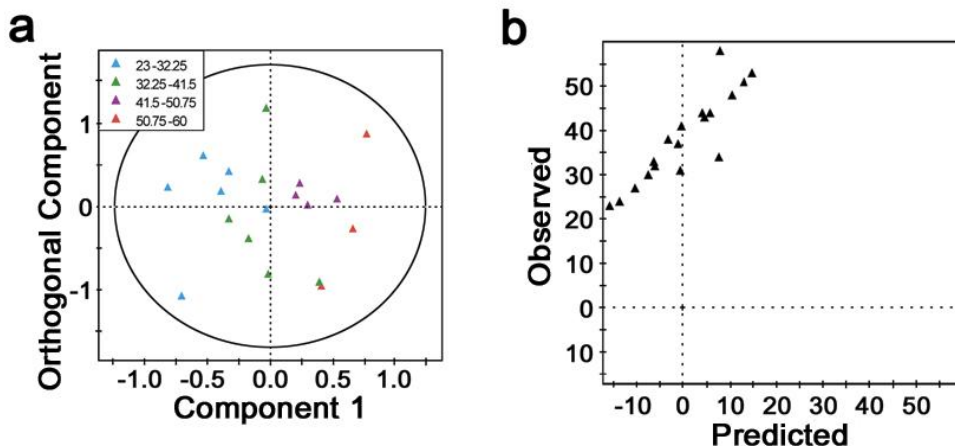
The PLS-DA model of metabolic differences between tissue types (CTWM, NAWM, NELES) had a  $R^2X$  of 0.11 and  $R^2Y$  of 0.63, indicating that 11% of the spectral data accounted for 63% of the variation in tissue type. The internal cross-validation resulted in 86% of the spectra (38 out of 44) being correctly predicted when left out of model building. Of the six misclassified spectra, two were classified as the wrong tissue type and four were not classified



as any tissue type with the *a priori* cutoff of a value of 0.5 or greater indicating class membership.

The O-PLS for neurologic impairment measured by EDSS had a R<sup>2</sup>X of 0.21 and R<sup>2</sup>Y of 0.86, indicating that 21% of the variation in the spectra explained 86% of the variation in EDSS score. The O-PLS model of verbal memory measured by RAVLT had a R<sup>2</sup>X of 0.22 and R<sup>2</sup>Y of 0.91, representing 22% of the spectral variation accounting for the 91% of the variation in RAVLT score. Additionally, the O-PLS model of stress measured by SRRS had a R<sup>2</sup>X of 0.29 and R<sup>2</sup>Y of 0.99, indicating 29% of the spectra variation explains 99% of the variation in SRRS.

It is always worrisome that multivariate modeling techniques may overfit the data, even with the use of internal cross-validation tools. Therefore, we built an O-PLS model for the PASAT test, one of the cognitive measures for which we could not obtain an internally validated O-PLS model. The O-PLS model is summarized in **Supplementary Figure 5**. The scores plot does not show a distinctive trend, and the observed versus predicted plot is extremely poor, with a RMSE of 42.5 for a variable that has a range of 23 – 60. Further, the model summary statistics are R<sup>2</sup>X: 0.068, R<sup>2</sup>Y 0.038, indicating a poor model in which 6.8% of the spectral variation explains 3.8% of the variation in PASAT score. This indicates that the metabolic variation in the NAWM is not predictive of variation in working memory, as measured by the PASAT.



**Supplemental Figure 5.** O-PLS does not produce an accurate model for working memory, which was unsuccessful in internal cross-validation, indicating that the metabolic patterns in the spectra do not describe the PASAT scores. The scores plot does not show a distinctive pattern of PASAT score, though it does show a slight trend from low to high. The observed vs. predicted plot, however, demonstrates that the model is not good, with a RMSE of 42.5.

### Supplementary Methods

Comparisons of group means and standard errors were calculated using R version 2.15.3, using two-sample, two-sided t-tests.

### References

- Barker M and Rayens W. 2003. Partial least squares for discrimination. *Journal of Chemometrics* 17(3):166-173.
- Lindon JC and Nicholson JK. 2008. Spectroscopic and statistical techniques for information recovery in metabonomics and metabolomics. *Annual Review of Analytical Chemistry* 1(1):45-69.
- Smolinska A, Blanchet L, Buydens LMC, Wijmenga SS. 2012. NMR and pattern recognition methods in metabolomics: From data acquisition to biomarker discovery: A review. *Analytica Chimica Acta* 750(0):82-97.
- Trygg J and Wold S. 2002. Orthogonal projections to latent structures (O-PLS). *Journal of Chemometrics* 16(3):119-128.
- Wold S, Sjöström M, Eriksson L. 2001. PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* 58(2):109-130.