

Supplement to: Aberrant levels of cortical myelin distinguish individuals with unipolar depression from healthy controls

David A.A. Baranger, Ph.D.<sup>1\*</sup>, Yaroslav O. Halchenko, Ph.D.<sup>2</sup>, Skye Satz, B.S.<sup>1</sup>, Rachel Ragozzino, M.A.<sup>1</sup>, Satish Iyengar, Ph.D.<sup>3</sup>, Holly A. Swartz, M.D.<sup>1</sup>, Mary L. Phillips, M.D., M.D. (Cantab)<sup>1</sup>, Anna Manelis, Ph.D.<sup>1\*</sup>

1. Department of Psychiatry, Western Psychiatric Institute and Clinic, University of Pittsburgh Medical Center, University of Pittsburgh, Pittsburgh, PA, USA

2. Department of Psychological and Brain Sciences, Dartmouth College, NH, USA

3. Department of Statistics, University of Pittsburgh, Pittsburgh, PA, USA

\* corresponding author

**Supplementary Information Table of Contents:**

Supplemental Methods: Pages 3-4

Supplemental Table 1: Page 5

Supplemental Figures 1-4: Pages 6-8

Supplemental Tables 2-4: See Supplemental data file

## Supplemental Methods:

Below is a step-by-step description of the nested-cross validation analysis.

1. We identified all possible combinations of UD and HC participants to be held out in the nested analysis:  $39 \text{ UD} * 47 \text{ HC} = 1833$  combinations. This determined the number of loops of nested cross-validation.
2. In each loop of nested cross-validation, one UD/HC participant pair was set aside. The remaining 84 participants were used in the elastic net analysis and were then used to run the linear discriminant analysis (LDA). The UD/HC pair that was set aside was then used to test the LDA model as the final step in each loop. As there were 1833 unique UD/HC participant pairs, the nested analysis was run 1833 times, with 84 participants in the training data and a different UD and HC pair set aside as testing data. Each of the 1833 nested analyses was conducted in the following way:
  - a. First, logistic elastic net was used to predict UD/HC status using cortical myelin values from  $n=349$  parcels, age, IQ, and sex. Leave-one out cross validation (i.e., train the model on 83 participants, test on 1 participant, repeat 84 times) was used to identify the optimal  $\lambda$  that corresponded to the minimal mean cross-validation error + 1 SE (Friedman et al., 2010).
  - b. Second, the optimal  $\lambda$  was used to fit an elastic net model and select variables important for UD vs. HC classification.
  - c. Third, this set of variables was used to train the LDA model on the same 84 participants.
  - d. Fourth, the LDA model was then tested on the two participants (1 UD and 1 HC) that were held out in each nested cross-validation loop. If elastic net did not identify any variables beyond the model intercept, then both participants were recorded as misclassified.
3. As each participant was tested using LDA 84 times, participant-wise accuracy was computed as the mean of 84 LDA accuracies. Total model accuracy was computed as the average of the participant-wise accuracies. Model sensitivity and specificity were computed as the average of UD-only and HC-only participant accuracies, respectively.
4. The list of all variables selected by at least one elastic net model was generated. For each variable on this list, the proportion of the  $n=1883$  models in which it was selected by the elastic-net regression was computed (i.e., variables selected in every model would be 100%, while variables that were selected in one model would be 0.0545%).
5. In order to identify the noise level for the frequency of variable selection as well as model fit, we repeated the same procedures described above while permuting UD/HC labels. In the permuted-labels analysis, each of 1833 loops was repeated 100 times for each set of 84 retained (training set) + 2 held-out (testing set) participants. To ensure unique randomization of UD/HC labels in each training set, a different random seed was used for each of  $1833*100=183300$  loops. True labels were kept for participants in the testing set. As UD/HC labels in the training sets were randomized, these classification results reflect model performance when myelin values do not carry useful information for

distinguishing UD from HC (i.e., they reflect false-positive results and over-fitting). Accuracy and variable selection frequency were computed as described in the procedures for true-label analysis.

6. To identify the variables most strongly predictive of case/control status, and which are less likely to reflect noise, the variable selection frequency with true case labels was compared to the variable selection frequency with permuted case labels. The variable in the true-label model was retained if it was above selection frequency for this variable in the permuted-label model plus 3.5 IQR. For example, if a variable was selected in 75% of models using the true case labels, but only in 2% of models with the permuted labels, then that variable was retained.

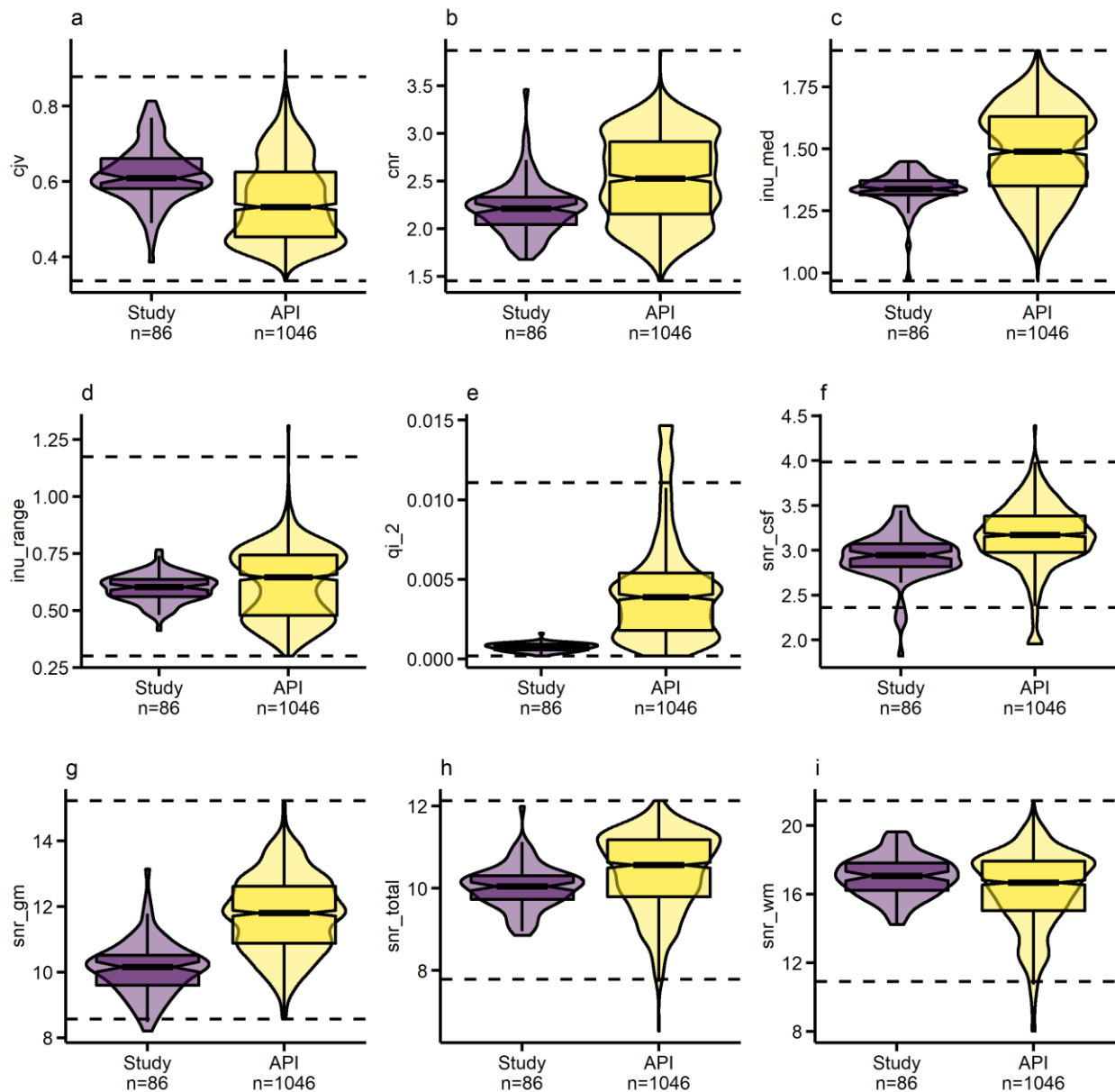
## Supplemental Results:

Supplemental Table 1: Rosner's test for outliers

Region	iteration (i)	CV	Mean.i	SD.i	R.i+1	lambda.i+1	Outlier?
R_Pir_ROI	0	0.247	0.021	0.016	13.7	3.8	TRUE
R_pOFC_ROI	1	0.139	0.021	0.011	10.4	3.8	TRUE
L_PreS_ROI	2	0.077	0.020	0.009	6.0	3.8	TRUE
L_Pir_ROI	3	0.073	0.020	0.009	5.9	3.8	TRUE
R_PreS_ROI	4	0.069	0.020	0.009	5.7	3.8	TRUE
L_pOFC_ROI	5	0.060	0.020	0.008	4.9	3.8	TRUE
R_EC_ROI	6	0.053	0.020	0.008	4.2	3.8	TRUE
R_25_ROI	7	0.052	0.020	0.008	4.2	3.8	TRUE
L_H_ROI	8	0.052	0.020	0.008	4.3	3.8	TRUE
R_H_ROI	9	0.050	0.020	0.007	4.1	3.8	TRUE
L_EC_ROI	10	0.049	0.020	0.007	4.1	3.8	TRUE
R_AAIC_ROI	11	0.043	0.019	0.007	3.4	3.8	FALSE
L_33pr_ROI	12	0.043	0.019	0.007	3.4	3.8	FALSE
R_52_ROI	13	0.042	0.019	0.007	3.4	3.8	FALSE
L_25_ROI	14	0.041	0.019	0.007	3.3	3.8	FALSE
L_PHA1_ROI	15	0.040	0.019	0.007	3.2	3.8	FALSE
L_47m_ROI	16	0.040	0.019	0.006	3.2	3.8	FALSE
R_MBelt_ROI	17	0.039	0.019	0.006	3.1	3.8	FALSE
R_V6A_ROI	18	0.037	0.019	0.006	2.8	3.8	FALSE
R_ProS_ROI	19	0.037	0.019	0.006	2.8	3.8	FALSE

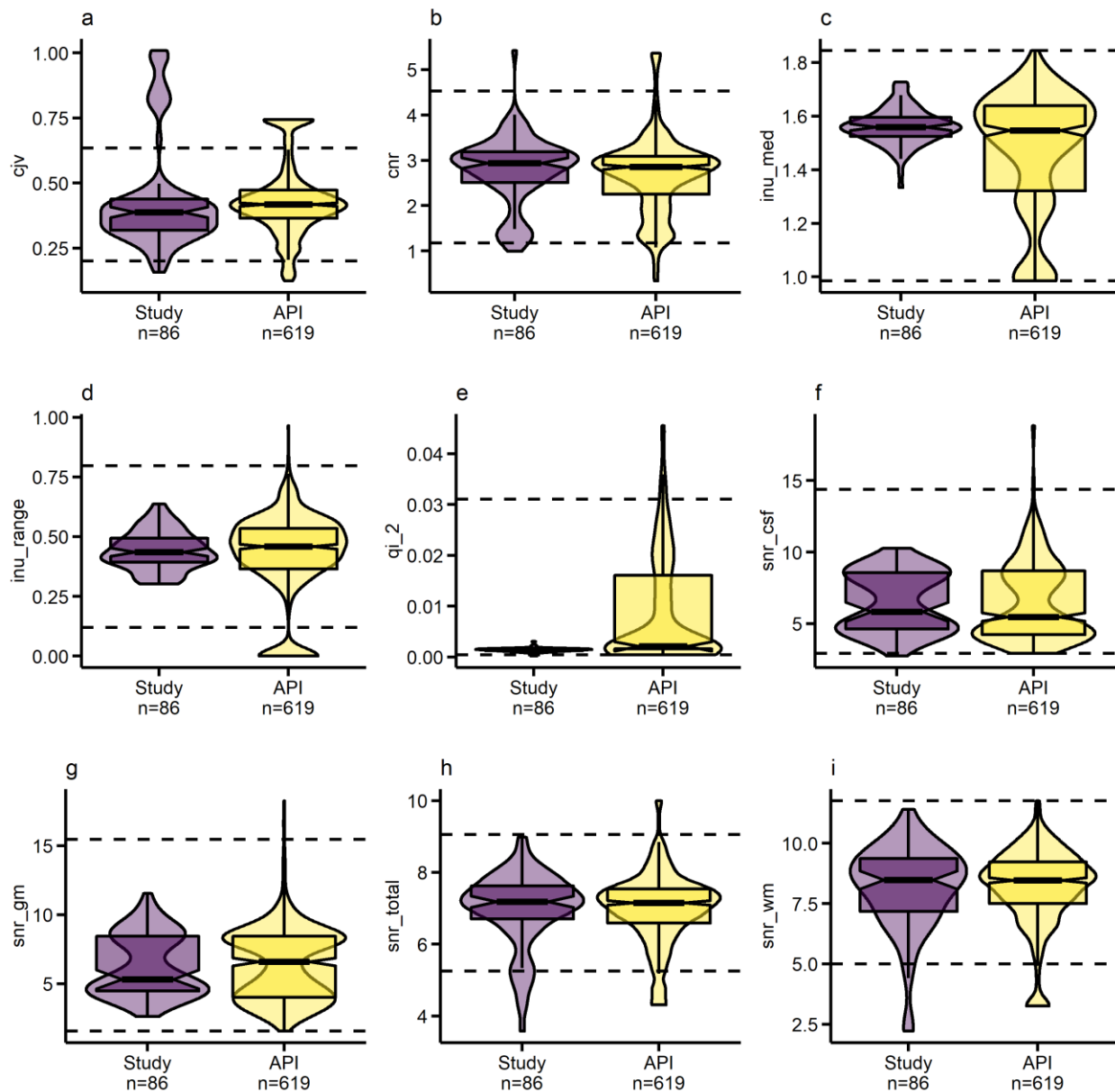
Results from testing for outliers in cortical myelin parcels. The coefficient of variation was used to summarize variance within each parcel. Rosner's test was used to identify regions with excessively high between-person variance, likely attributable to susceptibility artifacts. The 11 regions with the highest CV were identified as outliers ( $R_{i+1} > \lambda_{i+1}$ ).

Supplemental Figure 1. Comparison of T1 mriqc values between study data and API data



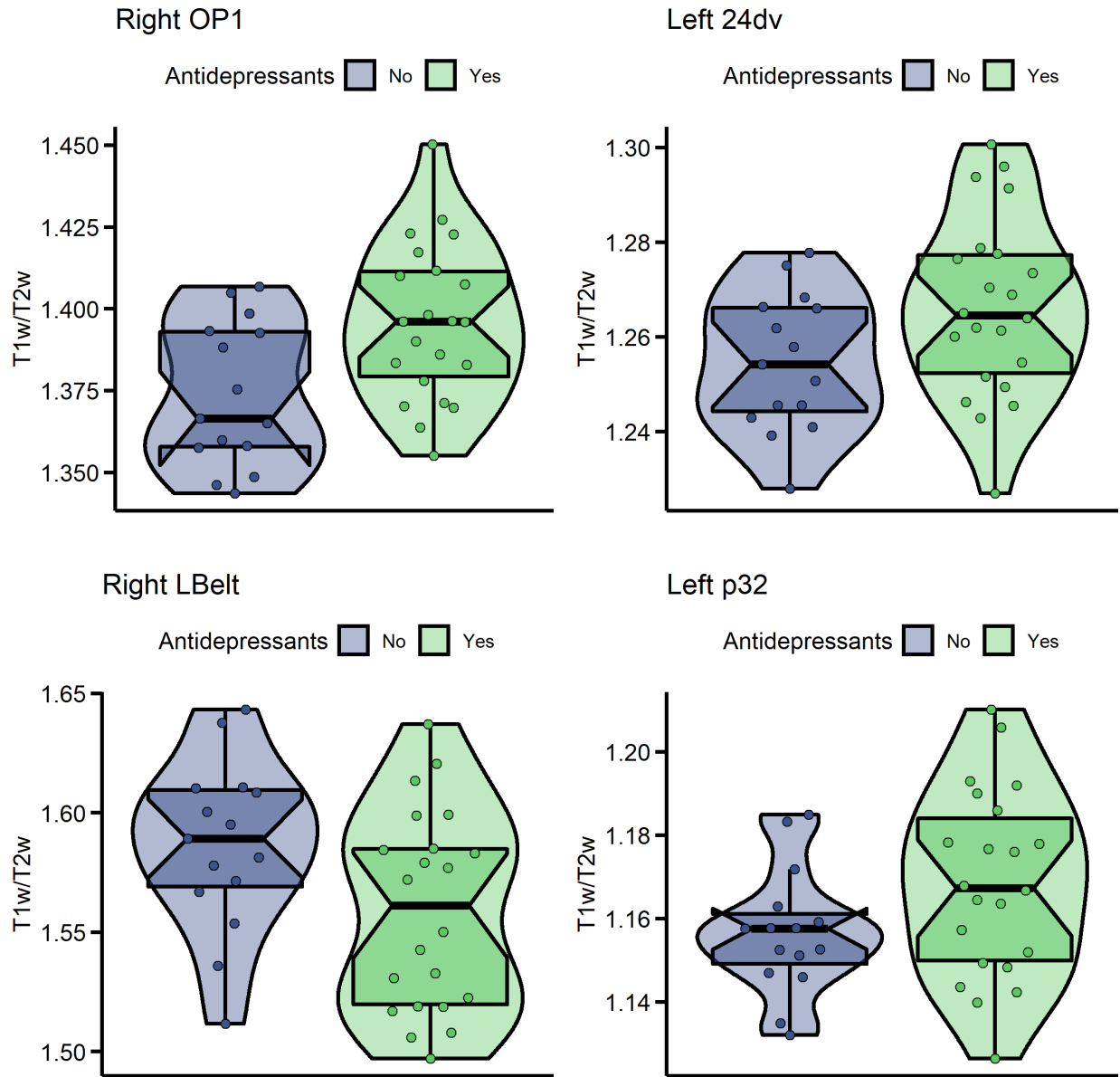
**a)** Coefficient of joint variation between white matter and gray matter. Higher values indicate more head motion and/or intensity non-uniformity artifacts. **b)** Contrast-to-noise ratio, reflecting separation between GM & WM. Higher values indicate higher quality. **c)** Intensity non-uniformity (bias field) median. Values closer to 1 indicate higher quality; further from zero indicate greater RF field inhomogeneity. **d)** Intensity non-uniformity (bias field) range. Values closer to 1 indicate higher quality; further from zero indicate greater RF field inhomogeneity. **e)** Mortamet's quality index 2. A quality index accounting for effects of both clustered and subtle artifacts in the air background. Higher values indicate lower quality. **f)** Signal-to-noise ratio within the CSF mask. Higher values indicate higher quality. **g)** Signal-to-noise ratio within the grey matter mask. Higher values indicate higher quality. **h)** Signal-to-noise ratio within the total mask. Higher values indicate higher quality. **i)** Signal-to-noise ratio within the white matter mask. Higher values indicate higher quality.

Supplemental Figure 2. Comparison of T2 mriqc values between study data and API data



**a)** Coefficient of joint variation between white matter and gray matter. Higher values indicate more head motion and/or intensity non-uniformity artifacts. **b)** Contrast-to-noise ratio, reflecting separation between GM & WM. Higher values indicate higher quality. **c)** Intensity non-uniformity (bias field) median. Values closer to 1 indicate higher quality; further from zero indicate greater RF field inhomogeneity. **d)** Intensity non-uniformity (bias field) range. Values closer to 1 indicate higher quality; further from zero indicate greater RF field inhomogeneity. **e)** Mortamet's quality index 2. A quality index accounting for effects of both clustered and subtle artifacts in the air background. Higher values indicate lower quality. **f)** Signal-to-noise ratio within the CSF mask. Higher values indicate higher quality. **g)** Signal-to-noise ratio within the grey matter mask. Higher values indicate higher quality. **h)** Signal-to-noise ratio within the total mask. Higher values indicate higher quality. **i)** Signal-to-noise ratio within the white matter mask. Higher values indicate higher quality.

Supplemental Figure 3. Association of antidepressant medication with cortical myelin



In participants with UD (n=39), antidepressant medication (n=22 were taking antidepressants, n=17 were not) was nominally associated ( $p < 0.05$  uncorrected) with cortical myelin in 4 of the 33 selected parcels.