Supplementary Material

Performance of Validated MicroRNA Biomarkers for Alzheimer's Disease in Mild Cognitive Impairment

Supplementary Table 1. MiRNAs in the Custom TaqMan Array and Amplification Summary for NTC and No RT. Details of each miRNA assay included in the Custom TaqMan Array: miRBase ID, mature miRNA sequence, role in the study, TFS assay ID. We also report the mean Cq of each miRNA following quality control filtering (AmpScore >1.0; CqConf < 0.8) in the negative control assays (CSF RNA with no RT enzyme [CSF-RT]; water only [NTC]; and water with no RT enzyme [NTC-RT]).

				Mean Cq		
miRBase ID	**Mature MiRNA Sequence**	**Role**	**TFS ID**	**CSF -RT**	**NTC**	**NTC -RT**
hsa-miR-15b-5p	UAGCAGCACAUCAUGGUUUACA	Biomarker	000390	UD	UD	UD
hsa-miR-19b-3p	UGUGCAAAUCCAUGCAAAACUGA	Biomarker	000396	UD	UD	UD
hsa-miR-30a-3p	CUUUCAGUCGGAUGUUUGCAGC	Biomarker	000416	UD	UD	UD
hsa-miR-140-5p	CAGUGGUUUUACCCUAUGGUAG	Biomarker	001187	UD	UD	UD
hsa-miR-142-3p	UGUAGUGUUUCCUACUUUAUGGA	Biomarker	000464	UD	UD	UD
hsa-miR-146a-5p	UGAGAACUGAAUUCCAUGGGUU	Biomarker	000468	UD	UD	UD
hsa-miR-146b-5p	UGAGAACUGAAUUCCAUAGGCUG	Biomarker	001097	UD	29.88	UD
hsa-miR-193a-5p	UGGGUCUUUGCGGGCGAGAUGA	Biomarker	002281	UD	UD	UD
hsa-miR-223-3p	UGUCAGUUUGUCAAAUACCCCA	Biomarker	002295	UD	29.88	UD
hsa-miR-331-3p	GCCCCUGGGCCUAUCCUAGAA	Biomarker	000545	UD	UD	UD
hsa-miR-365a-3p	UAAUGCCCCUAAAAAUCCUUAU	Biomarker	001020	UD	UD	UD
hsa-miR-378a-3p	ACUGGACUUGGAGUCAGAAGGC	Biomarker	002243	UD	UD	UD
hsa-miR-484	UCAGGCUCAGUCCCCUCCCGAU	Biomarker	001821	UD	UD	UD
hsa-miR-519b-3p	AAAGUGCAUCCUUUUAGAGGUU	Biomarker	002384	UD	UD	UD
hsa-miR-584-5p	UUAUGGUUUGCCUGGGACUGAG	Biomarker	001624	UD	34.87	UD
hsa-miR-597-5p	UGUGUCACUCGAUGACCACUGU	Biomarker	001551	UD	34.01	UD
hsa-miR-1291	UGGCCCUGACUGAAGACCAGCAGU	Biomarker	002838	UD	UD	UD
hsa-miR-30e-3p	CUUUCAGUCGGAUGUUUACAGC	Pos. Control	000422	UD	UD	UD
hsa-miR-92a-3p	UAUUGCACUUGUCCCGGCCUGU	Pos. Control	000431	UD	UD	UD
hsa-miR-574-3p	CACGCUCAUGCACACACCCACA	Pos. Control	002349	UD	29.69	UD
hsa-miR-638	AGGGAUCGCGGGCGGGUGGCGGCCU	Pos. Control	001582	UD	35.08	UD
hsa-miR-217-5p	UACUGCAUCAGGAACUGAUUGGA	Non-Expressor	002337	UD	UD	UD
hsa-miR-647	GUGGCUGCACUCACUUCCUUC	Non-Expressor	001600	UD	UD	UD
U6 snRNA	GTGCTCGCTTCGGCAGCACATATACTA AAATTGGAACGATACAGAGAAGATTA GCATGGCCCCTGCGCAAGGATGACACG CAAATTCGTGAAGCGTTCCATATTTT	TFS Control	001973	17.55	32.06	31.24

UD, undetermined in the qPCR; NTC, no template control; -RT, no-RT; TFS, ThermoFisher Scientific

Supplementary Table 2. Description of Cq values for 4 positive control miRNAs, by diagnostic group and overall. The means, variances, and ranges of the positive control miRNAs were very stable across diagnostic groups, and the average of the non-U6 controls had negligible variance, which made it highly suitable for use as an endogenous positive control value calculable from the measured Cq values from each probe. (See Materials and Methods - *Array card batch correction, normalization, and transformation of expression scale* for details.)

	Mean ± SD [min, max]			
Positive control miRNAs	**NC (n=65)**	**MCI (n=31)**	**AD (n=37)**	**All (n=133)**
U6 snRNA (card means)	24.63 ± 1.46 [19.42, 27.44]	24.48 ± 1.35 [20.66, 27.08]	24.41 ± 1.27 [20.66, 27.08]	24.53 ± 1.39 [19.42, 27.44]
miR-30e-3p	28.20 ± 1.68 [25.09, *34.00*]	28.36 ± 1.46 [24.99, 32.41]	28.71 ± 1.71 [26.46, *34.00*]	28.38 ± 1.64 [24.99, *34.00*]
miR-574-3p	25.73 ± 1.51 [21.99, 31.04]	26.09 ± 1.63 [20.54, 28.83]	26.59 ± 1.88 [23.67, *34.00*]	26.05 ± 1.67 [20.54, *34.00*]
miR-638	31.62 ± 1.88 [27.82, 34.52]	31.60 ± 1.82 [28.37, 34.77]	31.62 ± 1.95 [28.51, 35.64]	31.62 ± 1.87 [27.82, 35.64]
miR-92a	27.23 ± 1.32 [24.21, 30.09]	27.67 ± 1.90 [23.43, *34.00*]	27.80 ± 1.12 [26.21, 29.86]	27.49 ± 1.44 [23.43, *34.00*]
Average {30e-3p, 574-3p, 638, 92a} (adjusted for U6 card means and centered at min Cq=21)	20.98 ± 0.17 [20.67, 21.40]	20.97 ± 0.21 [20.48, 21.42]	21.06 ± 0.18 [20.63, 21.34]	21.00 ± 0.19 [20.48, 21.42]

Note: values of "34.00" (in italics) were not measured but represent censoring of levels below limits of detection

SUPPLEMENTARY RESULTS

Verification of biomarker relevance in the cohort

As a global check on the relevance of the panel of miRNAs for the cohort in the present study (which included both new NC, MCI, and AD samples, as well as repeated NC and AD samples), we verified that the miRNA profiles (taken across the entire panel) for MCI and AD samples tended to differ consistently from those of NC samples. Briefly, we calculated Mahalanobis distances [1] for all samples with respect to the center of the NC group in the miRNA expression response space spanning the entire analytic panel of 15 markers, scaled by the miRNA covariance within the NC group. Before calculating Mahalanobis distances, for censored observations we imputed values >34 Cq using predictions from a Tobit regression model [2] of the index miRNA using all the other biomarker miRNAs as predictors; this imputation was not used in expression or trend analyses, but only for enabling the visualization. Mahalanobis distances provided a way to measure how similar a set of multivariate observations (in this case, the miRNA profiles for MCI and AD) was to a reference set of observations (here, NC), while accounting for the covariance among the variables. Plotting the distances as a scatter across the observations allowed us to see whether or not the MCI and AD samples tended to be dissimilar to the NC samples, and whether the AD samples were more different from NC than the MCI samples are. Supplementary Figure 1 summarizes the analysis, showing the distribution of Mahalanobis distances for all samples, color-coded by diagnosis. We clearly saw that MCI and AD samples were atypical compared to the NC group, indicating that the miRNAs were informative for MCI and AD.

Robustness of findings to exclusion of repeated samples

As a sensitivity analysis, we temporarily excluded the repeated 21 NC and 23 AD samples (retaining only the 44 NC, 31 MCI, and 14 AD new samples), and on the reduced cohort performed the same trend and classification analyses as described in the manuscript (Materials and Methods: *Statistical analysis*).

The trend analysis on the reduced cohort is summarized in Supplementary Figure 2. Despite the drastically reduced power (just n=89 total, with only 14 in the AD group), we found that the patterns were for the most part qualitatively similar. The trend profiles of miR-19b-3p and miR-378a-3p were less clearly a match to their trend category, as was that of miR-140-5p (which looked like a "TREND" example), but among the trending miRNAs only miR-146a-5p was in question, as it appeared more like "NO TREND". (The slope for miR-142-3p was quite attenuated, but still negative.) We believe the distortion in the patterns was due to the small number of AD examples in the cohort of new samples. Overall, this sensitivity analysis appears to adequately mirror the full-cohort findings.

The sensitivity analysis of classification performance for the 5 trending miRNAs yielded similar concordance with full-cohort findings. The AUC for predicting AD in the reduced cohort was 0.833 (compare to 0.770 in the full cohort), and for MCI was 0.754 (compare to 0.705). By way of comparison, the AUC of $A\beta_{42}$:T-Tau in the reduced cohort was 0.810 (compare to 0.867) for AD and 0.780 (compare to 0.758) for MCI. This was quite good agreement considering the reduced power and variety in the examples (particularly for AD). Looking at the patterns of discrepancy, we inferred that the new NC samples were more tightly clustered in miRNA profile (i.e., contained fewer examples of AD-leaning profiles) than the previous NC samples, and the new AD samples were less extreme in $A\beta_{42}$:T-Tau ratio (i.e., tended to have fewer very small

values of the ratio) than the previous AD samples. (In fact, the mean value of $A\beta_{42}$:T-Tau in the MCI group was smaller than the mean in the 14 new AD samples.) These two differences caused the miRNAs to appear more informative than they actually were, and $A\beta_{42}$:T-Tau to appear less informative for AD (and somewhat more informative for MCI). Clearly the full cohort provided a more robust picture of the value of the miRNA biomarkers for discriminating MCI from NC and AD, but the classification performance still showed the same rank ordering of models as in the full cohort, and good agreement in AUC values over the scenarios considered.

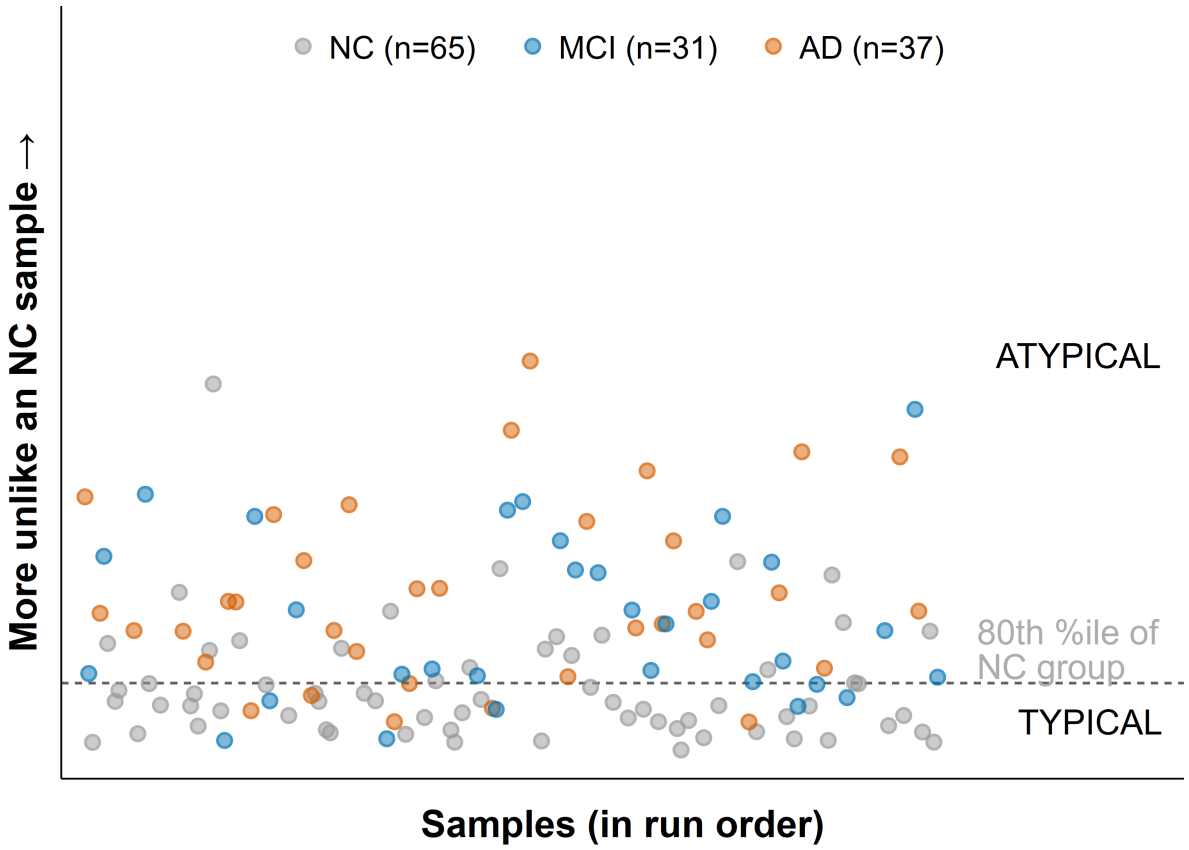
In these sensitivity analyses we found no indication that the results or conclusions of the full-cohort analysis would be greatly altered had we performed the analysis on just the reduced cohort of new samples instead. We believe the findings are robust to the choice of whether to include repeated samples or not.

REFERENCES

[1] Mahalanobis PC (1936) On the Generalised Distance in Statistics. *Proc Natl Acad Sci India* **2**, 49-55.

[2] Tobin J (1958) Estimation of relationships for limited dependent variables. *Econometrica* **26**, 24-36.

Supplementary Figure 1. Mahalanobis distances from the center of the NC group, calculated from multivariate miRNA expression profiles, revealed that MCI and AD samples tended to be atypical compared to NC samples.



Supplementary Figure 2. Sensitivity analysis of trends, performed on a reduced cohort excluding the repeated samples. Compare to Figure 2 in the main results. The trend profiles for individual miRNAs were broadly similar to what was observed in the full cohort, with some attenuation or distortion in a few cases (e.g., miR-146a-5p) due to the drastically reduced NC and especially AD sample sizes. From this data we conclude that 5 miRNAs were trending (swapping miR-140-5p for miR-146a-5p), and found similar classification performance characteristics when those miRNAs were used to discriminate MCI and AD.